

Министерство культуры Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ КИНО И ТЕЛЕВИДЕНИЯ»

А.И. ЧЕРЕМИСИН

Ф У Р Ъ Е – О П Т И К А

Часть 1

Световые волны и сигналы

Учебное пособие для студентов технических
факультетов, обучающихся по специальностям
210400 и 210100

Санкт-Петербург
2012

УДК 535

ББК 22.34

Черемисин А.И. Фурье-оптика, часть 1. Световые волны и сигналы. Учебное пособие. – СПб.: Изд. СПбГУКиТ, 2012. – 340 с.

Предлагаемое пособие состоит из двух частей и содержит достаточно полное изложение методов фурье-оптики и их приложений. В первой части рассматриваются вопросы применения методов фурье-анализа в волновой оптике (интерференция, когерентность, дифракция),

Учебное пособие может быть рекомендовано студентам, обучающимся по специальностям 210400, 210100 и изучающим физическую и прикладную оптику, радиотехнику, звукотехнику, «Основы записи и воспроизведения изображения», «Линейные интегральные преобразования в оптике», а также аспирантам и всем специалистам, использующим методы фурье-оптики.

Рецензент д.ф.-м.н., профессор С.В. Ульянов (СПбТЭИ)

Рекомендовано к изданию в качестве учебного пособия Советом факультета мультимедиа аппаратуры

Протокол № 12 от 21 марта 2012 г.

© Санкт-Петербургский государственный университет кино и телевидения, 2012

ОГЛАВЛЕНИЕ

| | |
|------------------|---|
| Предисловие..... | 7 |
|------------------|---|

Глава 1. Световые волны

| | |
|--|----|
| 1.1. Свет как электромагнитная волна..... | 10 |
| 1.2. Плоские световые волны..... | 16 |
| 1.2.1. Формула плоской монохроматической световой волны...17 | |
| 1.2.2. Поляризация плоской волны..... | 25 |
| 1.3. Сферические световые волны..... | 28 |
| 1.4. Цилиндрические световые волны..... | 34 |
| 1.5. Волновое скалярное уравнение Гельмгольца..... | 38 |
| 1.6. Пространственные частоты..... | 42 |
| 1.7. Интенсивность светового излучения..... | 48 |

Глава 2. Частотный спектр сигналов

| | |
|---|-----|
| 2.1. Общие представления о частотном спектре..... | 52 |
| 2.2. Частотный спектр некоторых простых сигналов..... | 62 |
| 2.2.1. Спектр биполярного П-образного сигнала..... | 62 |
| 2.2.2. Спектр незатухающего волнового цуга..... | 65 |
| 2.2.3. Спектр периодической последовательности цугов..... | 70 |
| 2.2.4. Спектр монохроматического излучения..... | 72 |
| 2.2.5. Спектр модулированного колебания..... | 76 |
| 2.2.6. Спектр затухающего цуга..... | 85 |
| 2.3. Пространственный спектр излучения..... | 89 |
| 2.3.1. Спектр одномерного сигнала..... | 89 |
| 2.3.2. Спектр двумерного сигнала..... | 94 |
| 2.4. Квазимонохроматический свет..... | 100 |
| 2.5. Пучок света конечной ширины..... | 110 |

Глава 3. Интерференция света

| | |
|--|-----|
| 3.1. Интерференция световых монохроматических волн..... | 116 |
| 3.2. Интерференция волн различных видов..... | 121 |
| 3.2.1. Интерференция плоских волн..... | 121 |
| 3.2.2. Интерференция сферической и цилиндрической волн.... | 134 |
| 3.2.3. Интерференция плоской и сферической волн..... | 142 |
| 3.2.4. Интерференция плоской и цилиндрической волн..... | 146 |
| 3.3. Методы получения когерентных световых пучков..... | 148 |
| 3.4. Интерференция по методу деления волнового фронта. Опыт Юнга..... | 152 |
| 3.5. Интерференция по методу деления амплитуды..... | 157 |
| 3.5.1. Кольца Ньютона | 157 |
| 3.5.2. Интерферометр Майкельсона..... | 160 |

Глава 4. Когерентность

| | |
|---|-----|
| 4.1. Частичная когерентность. Корреляция световых волн..... | 163 |
| 4.2. Временная когерентность | 171 |
| 4.3. Спектральный подход к рассмотрению временной когерентности..... | 176 |
| 4.4. Теорема Винера – Хинчина..... | 183 |
| 4.5. Пространственная когерентность..... | 187 |
| 4.6. Теорема Винера – Ван-Циттера – Цернике..... | 197 |

Глава 5. Дифракция света. Дифракция Френеля

| | |
|--|-----|
| 5.1. Определение дифракции. Виды дифракции..... | 204 |
| 5.2. Принцип Гюйгенса – Френеля..... | 207 |
| 5.3. Дифракционный интеграл Кирхгофа – Френеля – Зоммерфельда..... | 211 |
| 5.4. Представление поля излучения в виде совокупности плоских волн..... | 213 |
| 5.5. Приближение Френеля..... | 220 |

| | |
|---|-----|
| 5.6. Дифракция Френеля на круглом отверстии..... | 226 |
| 5.7. Дифракция Френеля на круглом диске..... | 231 |
| 5.8. Дифракция Френеля на крае полуплоскости..... | 233 |
| 5.9. Дифракция Френеля на щели..... | 238 |

Глава 6. Дифракция Фраунгофера

| | |
|---|-----|
| 6.1. Приближение Фраунгофера..... | 243 |
| 6.2. Дифракции Фраунгофера как пространственное преобразование Фурье..... | 246 |
| 6.3. Дифракция Фраунгофера на двух точечных отверстиях.... | 252 |
| 6.4. Дифракция Фраунгофера на щели..... | 255 |
| 6.5. Дифракция Фраунгофера на двух параллельных щелях.... | 260 |
| 6.6. Дифракция Фраунгофера на прямолинейном крае протяженного экран..... | 265 |
| 6.7. Дифракция Фраунгофера на прямоугольном отверстии.... | 269 |
| 6.8. Дифракция Фраунгофера на круглом отверстии..... | 274 |
| 6.9. Применение метода подобия к изучению дифракции Фраунгофера..... | 280 |
| 6.10. Пространственная когерентность и дифракция Фраунгофера..... | 282 |

Глава 7. Дифракция света на периодических структурах

| | |
|---|-----|
| 7.1. Амплитудная гармоническая решетка..... | 287 |
| 7.2. Фазовая гармоническая решетка..... | 295 |
| 7.3. Плоская одномерная щелевая решетка..... | 298 |
| 7.3.1. Дифракция света на щелевой решетке..... | 298 |
| 7.3.2. Разложение апертурной функции щелевой решетки в ряд Фурье..... | 308 |

| | |
|--|------------|
| 7.3.3. Распределение амплитуды поля в дифракционной картине щелевой решетки как произведение двух фурье-преобразований | 312 |
| 7.3.4. Разрешающая способность щелевой дифракционной решетки..... | 314 |
| 7.4. Решетки, периодические относительно квадратичных переменных..... | 318 |
| 7.4.1. Решетка, периодическая относительно одной квадратичной переменной..... | 318 |
| 7.4.2. Зонная решетка Габора..... | 319 |
| 7.4.3. Зонная пластинка Френеля..... | 323 |
| 7.5. Двумерные решетки..... | 329 |
| 7.5.1. Решетка, полученная наложением двух скрещенных щелевых решеток..... | 329 |
| 7.5.2. Решетка, состоящая из точечных отверстий..... | 331 |
| 7.6. Трехмерные решетки..... | 334 |
| Использованная литература..... | 339 |

Часть вторая

Глава 8. Линейные интегральные преобразования в оптике

| | |
|--|----|
| 8.1. Общие представления о линейных интегральных преобразованиях | 8 |
| 8.2. Преобразование Фурье | 10 |
| 8.3. Преобразование Фурье – Бесселя | 13 |
| 8.4. Преобразование свертки | 15 |
| 8.5. Корреляция | 22 |
| 8.6. Преобразование Френеля | 26 |
| 8.7. Преобразование Абеля | 30 |
| 8.8. Преобразование Гильберта | 31 |
| 8.9. Преобразование Мелина | 34 |

| | |
|---|----|
| 8.10. Преобразование Дирака и преобразование отсчетов | 36 |
| 8.11. Преобразование Лапласа | 40 |

Глава 9. Линейные оптические системы

| | |
|---|----|
| 9.1. Понятие линейной оптической системы | 43 |
| 9.2. Импульсная характеристика оптической системы | 49 |
| 9.3. Передаточная функция оптической системы | 56 |
| 9.4. Слой свободного пространства | 61 |
| 9.4.1. Импульсная характеристика и передаточная функция слоя свободного пространства | 61 |
| 9.4.2. Решение интегрального уравнения распространения световой волны в свободном пространстве | 69 |
| 9.6. Линейность оптической системы при различных видах освещения | 71 |
| 9.7. Оптическая передаточная функция оптической системы | 77 |
| 9.8. Функция рассеяния оптической системы | 80 |

Глава 10. Центрированные оптические системы формирования изображения

| | |
|---|-----|
| 10.1. Основные понятия геометрической теории изобра- жений | 88 |
| 10.2. Центрированные оптические системы | 94 |
| 10.3. Тонкая линза | 99 |
| 10.3.1. Функция пропускания тонкой линзы | 100 |
| 10.3.2. Фокусирующие свойства линзы | 105 |
| 10.3.3. Тонкая линза как элемент, выполняющий преобразование Фурье | 117 |
| 10.3.4. Двойное преобразование Фурье | 132 |
| 10.3.5. Частотная характеристика тонкой линзы | 135 |

Глава 11. Формирование оптического изображения при когерентном освещении

| | |
|---|-----|
| 11.1. Искажение изображения при распространении света в свободном пространстве | 139 |
|---|-----|

| | |
|--|-----|
| 11.2. Изображение линзой точечного объекта | 144 |
| 11.3. Изображение линзой протяженного предмета | 147 |
| 11.4. Образование изображения при когерентном освещении как двойное преобразование Фурье | 159 |
| 11.5. Влияние на изображение конечных размеров линзы | 163 |
| 11.6. Влияние конечных размеров линзы на спектр пространственных частот | 170 |
| 11.7. Формирование оптического изображения по Аббе – Портеру | 176 |

Глава 12. Формирование оптического изображения при некогерентном и частично когерентном освещении

| | |
|--|-----|
| 12.1. Разрешающая способность оптических систем | 190 |
| 12.2. Оптические передаточные функции дифракционно-ограниченных систем | 198 |
| 12.3. Функции рассеяния точки, линии и края дифракционно-ограниченных систем | 204 |
| 12.4. Формирование изображения при некогерентном освещении | 211 |
| 12.3.1. Изображение простейших объектов | 216 |
| 12.3.2. Изображение гармонической дифракционной решетки | 218 |
| 12.5. Сравнение формирования изображения при когерентном и некогерентном освещении | 224 |
| 12.6. Изображение при частично когерентном квазимонохроматическом освещении | 229 |
| 12.7. Фотографическое отображение сигналов | 232 |
| 12.8. Оптическая обработка пространственной информации | 241 |
| 12.8.1. Пространственная фильтрация | 242 |
| 12.8.2. Восстановление искаженных изображений | 251 |

Глава 13. Голография

| | |
|--|-----|
| 13.1. Физические принципы голографии | 255 |
| 13.2. Голография простейших объектов | 261 |

| | |
|---|------------|
| 13.2.1. Плоская предметная и плоская опорная волны | 261 |
| 13.2.2. Точечный объект, плоская нормально падающая опорная волна | 266 |
| 13.2.3. Точечный объект, Плоская наклонная волна | 270 |
| 13.3. Голография произвольного предмета | 275 |
| 13.4. Голографическая схема Денисюка | 281 |
| 13.5. Геометрия формирования голограмм и восстановления изображений | 287 |
| 13.5.1. Точечный объект | 287 |
| 13.5.2. Плоский предмет | 291 |
| 13.6. Голография Фурье | 301 |
| 13.6.1. Голография Фурье с использованием линз | 301 |
| 13.6.2. Безлинзовая голография Фурье | 306 |
| 13.7. Формирование и реконструкция голограмм изобра- жений, создаваемых оптическими элементами | 311 |
| 13.8. Применения голографии | 317 |
| Использованная литература | 327 |

ПРЕДИСЛОВИЕ

Фурье-оптика изучает вопросы физической и прикладной оптики с использованием математического аппарата Фурье. Сюда входят вопросы когерентности, дифракции, голографии, формирования оптического изображения и др. Методы фурье-оптики, основанные на применении анализа Фурье, находят широкое применение в физике, радиофизике, прикладной оптике и других областях. Преобразования Фурье пришли в оптику из теории связи (поэтому фурье-оптику иногда называют радиооптикой). Они позволяют применять методы теории электрических цепей и электрических сигналов при рассмотрении оптических явлений. В оптике сигналом называется двумерное распределение светового (электрического) поля, поэтому различие между системой связи и оптической системой, создающей изображение, заключается лишь в том, что в первом случае информация носит временную природу, т.е. имеет место временная модуляция сигнала, во втором – пространственную (модулируются амплитуды и фазы светового поля как функции координат). С точки зрения использования математического аппарата Фурье это различие несущественно, поэтому методы теории связи могут быть использованы и при анализе оптических систем. Они применимы и к теории формирования оптического изображения.

Применение математического аппарата Фурье в оптике придало теории оптических изображений математическую строгость и открыло возможности для плодотворных аналогий между оптикой и теорией связи, например, между различными операциями, проводимыми над электрическими и оптическими сигналами (спектральный анализ, преобразование частот и др.). Вместе с преобразованием Фурье в оптику пришли такие понятия и термины, как линейная оптическая система, импульсное воздействие (точечный источник света), импульсный отклик (изображение точечного источника), пространственная частота,

частотная характеристика, импульсная характеристика, передаточная функция, фурье-спектр, фурье-анализ и др.

В предлагаемом учебном пособии дано достаточно полное изложение вопросов фурье-оптики. Пособие состоит из двух частей, связанных между собой. Нумерация глав сквозная. В первой части пособия рассматриваются в основном вопросы физической оптики, во второй – вопросы прикладной оптики.

Первая часть пособия состоит из 7 глав.

Первую главу можно рассматривать как вводную. В ней рассматриваются общие вопросы световых электромагнитных волн. Подробно рассматриваются плоские, сферические и цилиндрические волны как в обычной форме, так и в параксиальном приближении, которое широко используется в фурье-оптике, рассматривается скалярное уравнение Гельмгольца. Вводятся также понятия пространственной частоты, оптического или пространственного сигнала, интенсивности света – понятий, которыми оперирует фурье-оптика..

Вторая глава посвящена вопросам спектрального разложения излучения. Рассматривается как временной, так и пространственный спектры сигналов, периодических и непериодических. Вводятся математические понятия sinc-функции и δ -функции Дирака, широко используемые в фурье-оптике.

В третьей и четвертой главах рассматриваются вопросы интерференции света. В третьей главе описываются все возможные эффекты сложения монохроматических волн: плоских, сферических, цилиндрических. Полученные здесь результаты используются при изложении вопросов голографии. Рассмотрены методы получения когерентных пучков света – метод деления волнового фронта и метод деления амплитуды. Четвертая глава посвящена вопросам когерентности света. Здесь подробно рассматриваются временная и пространственная когерентности. Вводится понятие функции корреляции световых полей. Рассматриваются теоремы Винера – Хинчина и Винера – Ван Циттера – Цернике.

В пятой, шестой и седьмой главах излагаются вопросы дифракции света. Подробно рассмотрена дифракция Френеля и дифракция Фраунгофера. Вводится понятие амплитудной функции пропускания объектного экрана. Показано, что дифракция Фраунгофера представляет собой фурье-преобразование функции пропускания. В пятой и шестой главах рассматривается дифракция на одномерных и двумерных структурах. Седьмая глава посвящена дифракции на периодических структурах (гармонических, щелевых, одномерных, двумерных и трехмерных решетках)

Пособие предназначено для студентов, изучающих физическую и прикладную оптику, светотехнику, радиотехнику, звукотехнику, а так же для магистрантов и аспирантов, всех, кто использует методы фурье-оптики. Материал, изложенный в пособии, может служить базой для изучаемых в СПбГУКиТ дисциплин «Линейные интегральные преобразования в оптике» (на факультете мультимедиа аппаратуры) и «Основы записи и воспроизведения изображения (ОЗВИ)».

Автор выражает благодарность студентам СПбГУКиТ Куксову В.М., Михайлову Л.М., Морозову А.В. Сморжевскому (Кириленко) А.А. за выполненные рисунки, доценту кафедры светотехники Тульевой Н.Н. за проявленный интерес к работе и ряд полезных советов, а также доценту кафедры физики НИУ ИТМО Зинчику А.А. за внимательное ознакомление с рукописью первой части пособия и положительную оценку работы.

Глава 1

СВЕТОВЫЕ ВОЛНЫ

1.1. Свет как электромагнитная волна

В основе волновой теории света и ее приложений лежит система уравнений Максвелла. Для однородной и изотропной среды без источников электрического и магнитного полей (зарядов и токов) эта система имеет вид

$$\nabla \times \mathbf{E} = -\partial \mathbf{B} / \partial t, \quad \nabla \times \mathbf{H} = \partial \mathbf{D} / \partial t,$$

$$\nabla \cdot \mathbf{D} = 0, \quad \nabla \cdot \mathbf{B} = 0.$$

Уравнения Максвелла устанавливают связь между производными по координатам и времени от векторных величин – напряженности электрического поля \mathbf{E} и индукции магнитного поля \mathbf{B} , характеризующих единое электромагнитное поле. Они показывают, что изменение во времени одного поля приводит к изменению в пространстве другого поля, и наоборот, изменение в пространстве одного поля приводит к изменению во времени другого поля; при этом $\mathbf{E} = (\varepsilon_0 \varepsilon)^{-1} \mathbf{D}$, $\mathbf{B} = \mu_0 \mu \mathbf{H}$, где \mathbf{D} – вектор индукции электрического поля, \mathbf{H} – вектор напряженности магнитного поля, ε и μ – соответственно диэлектрическая и

магнитная проницаемости среды, ε_0 и $\mu_0 = (\varepsilon_0 c^2)^{-1}$ – соответственно электрическая и магнитная постоянные.

Из этих уравнений вытекает существование электромагнитных волн – распространяющихся в пространстве колебаний электрического и магнитного полей. Действительно, из второго и первого уравнений видно, что переменное электрическое поле порождает переменное магнитное поле, а переменное магнитное поле – переменное электрическое поле. Поэтому если в какой-либо точке пространства возбудить переменное электрическое поле (например, с помощью колеблющегося электрического заряда), то это поле породит в соседних точках магнитное поле, которое также оказывается переменным. Переменное магнитное поле в свою очередь породит в других точках пространства переменное электрическое поле и т.д. Возникает последовательность взаимных превращений электрических и магнитных полей, распространяющихся с конечной скоростью от одной точки пространства к другой. Эти взаимные превращения полей и приводят к их сохранению в отсутствие источников. Такой процесс оказывается периодическим во времени и в пространстве, т.е. имеет волновой характер.

Для электромагнитных волн, распространяющихся в произвольном направлении в однородной электронейтральной непроводящей среде, из уравнений Максвелла можно получить линейное дифференциальное уравнение в частных производных второго порядка, называемое волновым уравнением электромагнитных волн:

$$\Delta E(x, y, z, t) - \frac{1}{v^2} \frac{\partial^2 E(x, y, z, t)}{\partial t^2} = 0, \quad (1.1)$$

где $E(x, y, z, t)$ – любая компонента вектора напряженности электрического поля $\mathbf{E}(x, y, z, t)$, Δ – оператор Лапласа, представленный в декартовой системе координат (сумма вторых частных производных по координатам x, y, z),

$$v = \frac{1}{\sqrt{\varepsilon_0 \mu_0 \varepsilon \mu}} \quad (1.2)$$

– величина, имеющая размерность скорости и представляющая собой скорость распространения электромагнитных волн. Такое же уравнение имеет место и для самого вектора $\mathbf{E}(x, y, z, t)$. Общим решением волнового уравнения (1.1) является сумма волновых функций

$$E(x, y, z, t) = f\left(t \pm \frac{\mathbf{n} \cdot \mathbf{r}}{v}\right), \quad (1.3)$$

соответствующих распространению электромагнитного процесса со скоростью v , где $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ – радиус-вектор произвольной точки P (точки наблюдения), в которой определяется поле; x, y, z – координаты точки наблюдения на волновой поверхности; $\mathbf{i}, \mathbf{j}, \mathbf{k}$ – единичные векторы (орты) вдоль координатных осей X, Y, Z соответственно;

$$\mathbf{n} = \cos\alpha \mathbf{i} + \cos\beta \mathbf{j} + \cos\gamma \mathbf{k}$$

– единичный вектор, проведенный в направлении распространения волны; $\cos\alpha, \cos\beta, \cos\gamma$ – компоненты вектора \mathbf{n} (направляющие косинусы) вдоль координатных осей X, Y, Z соответственно; α, β, γ – углы между вектором \mathbf{n} и осями координат X, Y, Z соответственно (направляющие углы вектора \mathbf{n});

$$\mathbf{n} \cdot \mathbf{r} = x \cos\alpha + y \cos\beta + z \cos\gamma.$$

Убедиться в этом можно путем непосредственной подстановки функции (1.3) в волновое уравнение (1.1). При этом следует учесть, что сумма квадратов направляющих косинусов равна единице.

Функции вида (1.3) описывают электромагнитную волну, распространяющуюся со скоростью v в направлении,

определяемом вектором \mathbf{n} . Других решений волновое уравнение (1.1) не имеет. Вид функции f может быть любым; он не определяется только волновым уравнением, а зависит от начальных и граничных условий. Поэтому для определения конкретного решения волнового уравнения оно должно быть дополнено соответствующими начальными и граничными условиями.

Волновое уравнение, подобное уравнению (1.1), можно записать и для любой компоненты $B(x, y, z, t)$ вектора индукции магнитного поля $\mathbf{B}(x, y, z, t)$, и для самого вектора $\mathbf{B}(x, y, z, t)$ (и вообще для любой полевой величины $f(\mathbf{r}, t)$). При этом модули векторов \mathbf{E} и \mathbf{B} в электромагнитной волне связаны между собой соотношением

$$B(x, y, z, t) = \frac{1}{v} E(x, y, z, t). \quad (1.4)$$

Таким образом, из уравнений Максвелла вытекает, что изменяющийся во времени электромагнитный процесс, возникающий в данном месте, вызывает его изменение в другом месте с некоторым запаздыванием. Тем самым устанавливается конечность скорости распространения электромагнитных возмущений, а следовательно, и существование электромагнитных волн, распространяющихся в пространстве со скоростью v .

Следует отметить, что волновое уравнение (1.1) и подобное ему уравнение для любой проекции вектора индукции магнитного поля \mathbf{B} описывают распространение электромагнитных волн в пространстве при отсутствии электрических токов и зарядов. Однако, как показывается в электродинамике, электромагнитные волны возбуждаются именно токами (изменяющимися во времени, например, переменными токами) и зарядами (ускоренно движущимися, например, колеблющимися электрическими диполями).

Поскольку для вакуума $\varepsilon = \mu = 1$, то для скорости электромагнитных волн в вакууме из (1.2) будем иметь

$$v_0 = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} = c = 3 \cdot 10^8 \text{ м/с.} \quad (1.5)$$

Эта скорость совпадает с измеренной астрономическими методами скоростью света c . На основании этого совпадения Максвелл сделал физический вывод о том, что свет является электромагнитной волной. Этот вывод лежит в основе всей электромагнитной теории света.

С учетом соотношения (1.5) формулу скорости электромагнитных волн можно записать в виде

$$v = \frac{c}{\sqrt{\varepsilon\mu}}.$$

Эту формулу называют формулой Максвелла. Все среды, в которых может распространяться свет (такие среды называют прозрачными), являются либо диамагнетиками, либо парамагнетиками. Для таких сред $\mu \approx 1$, поэтому фазовая скорость света в прозрачных средах $v = c / n$, где

$$n = \sqrt{\varepsilon}$$

– показатель преломления среды. Эта формула связывает оптические и электрические свойства прозрачной среды и также называется формулой Максвелла. Физический смысл показателя преломления связан с колебаниями электронов и ионов вещества под действием распространяющейся в веществе световой электромагнитной волны.

Световая электромагнитная волна переносит в пространстве высокочастотные колебания электрического и связанного с ним магнитного поля. Как электромагнитная волна свет характеризуется амплитудой, длиной волны, частотой, фазой, поляризацией, скоростью и направлением распространения.

Световые электромагнитные волны с длиной волны λ_0 (в вакууме), заключенные в интервале от 0,40 мкм до 0,76 мкм, и соответственно частоты ν – в интервале от $0,39 \cdot 10^{15}$ Гц до $0,75 \cdot 10^{15}$ Гц, оказывают непосредственное воздействие на человеческий глаз, вызывая зрительные ощущения. Поэтому их называют видимым светом, а диапазон указанных длин волн и частот – видимым диапазоном. Благодаря волнам этого диапазона мы видим предметы окружающего нас мира. Однако и более коротковолновое ($\lambda_0 < 0,40$ мкм) ультрафиолетовое излучение и более длинноволновое ($\lambda_0 > 0,76$ мкм) инфракрасное излучение не отличаются по своей физической природе от видимого света и вместе с ним составляют так называемое оптическое излучение. Оптический диапазон простирается по длине волны от 10 нм до 2 мм и по частоте от $1,5 \cdot 10^{11}$ Гц до $3 \cdot 10^{16}$ Гц.

С помощью уравнений Максвелла можно показать, что колебания вектора напряженности электрического поля **E** и вектора индукции магнитного поля **B** в световой волне происходят перпендикулярно друг другу и перпендикулярно направлению скорости ее распространения. В этом заключается свойство поперечности световой электромагнитной волны. Векторы **v**, **E** и **B** в световой волне образуют правовинтовую ортогональную тройку. Следует отметить, что в природе существуют только правовинтовые электромагнитные волны; в этом состоит одно из проявлений законов взаимного создания переменных электрического и магнитного полей. В свободной световой волне векторы **E** и **B** колеблются синфазно (в одинаковой фазе, т.е. одновременно, и в одних и тех же точках пространства достигают максимального и минимального значения). Поскольку значения обоих векторов строго связаны между собой и их колебания синфазные, то совершенно безразлично, какой из векторов исследуется при изучении явлений, обусловленных характером волн. Обычно рассматривается электрический вектор, так как этот вектор в большинстве случаев определяет воздействие световой

электромагнитной волны на вещество; глаз человека и фотоматериалы тоже реагируют на эту составляющую электромагнитного поля. Поэтому вектор **E** в оптике называют световым вектором, а электрическое поле – световым полем.

Важную роль в оптике играют световые монохроматические волны. Световая электромагнитная волна называется монохроматической (одноцветной), если проекции векторов **E** и **B** на оси прямоугольной системы координат (перпендикулярные направлению распространения волны) совершают гармонические колебания с какой-либо фиксированной частотой ($\nu = \text{const}$), называемой частотой волны. Такие волны описываются гармоническими функциями и воспринимаются человеческим глазом как определенный цвет. Монохроматические волны возбуждаются гармоническими колебаниями источника волн.

Свет излучается атомами и молекулами вещества. Согласно классическим представлениям, излучение световых волн происходит в результате колебаний внутриатомных электронов, вызванных возбуждением атома (увеличением его внутренней энергии за счет энергии, получаемой извне). Все оптические явления обусловлены внешними (наиболее удаленными от ядра, а значит, наиболее слабо связанными с ядром) электронами. По этой причине их называют оптическими электронами. Считается, что оптический электрон связан в атоме квазиупругой силой, поэтому при возбуждении атома, приводящем к смещению электрона из положения равновесия, электрон под действием этой силы начнет совершать колебания с некоторой частотой и испускать электромагнитную волну той же частоты. Выведенный из положения равновесия и предоставленный затем самому себе, электрон будет совершать не гармонические, а затухающие колебания, так как он будет непрерывно терять энергию на излучение. Затухание, обусловленное излучением атома, называется радиационным затуханием. Поэтому излучение атома можно считать подобным излучению колеблющегося электрического диполя (электрического осциллятора), совершающего затухающие колебания. Время затухания (по

квантовой теории излучения – время жизни атома в возбужденном состоянии) $\tau \approx 10^{-8}$ с. За это время электрон совершает $N = \tau/T_0 = \omega_0\tau/2\pi \approx 10^7$ полных колебаний (здесь T_0 и ω_0 – период и собственная частота колебаний оптического электрона). Образуется последовательность $N \approx 10^7$ горбов и впадин, т.е. цуг волн длительностью $\tau \approx 10^{-8}$ с и протяженностью $l = c\tau \approx 3$ м. Таким образом, световая волна, излучаемая атомом, не является строго монохроматической (бесконечной синусоидой), а представляет собой последовательность затухающих цугов, испущенных атомом в его последовательных возбуждениях.

Характерной особенностью светового излучения является наличие у него не только волновых, но и корпускулярных свойств. С корпускулярной точки зрения световое излучение представляет собой поток особых частиц, называемых фотонами.

1.2. Плоские световые волны

В качестве исходных объектов для теоретического исследования явлений, связанных с распространением и взаимодействием света с различными объектами, обычно используют волны наиболее простого вида, а именно плоские монохроматические волны. Волна называется плоской, если ее амплитуда и фаза в любой момент времени постоянны вдоль плоской поверхности, перпендикулярной направлению распространения волны (эта поверхность называется волновой или фазовой поверхностью). Характерным признаком плоской волны является также постоянство направления ее распространения.

Понятие плоской монохроматической волны значительно облегчает описание и изучение волновых процессов. Выбор плоских монохроматических волн в качестве исходных объектов обусловлен не только их простотой, но и тем, что в целом ряде физических задач реальные волны можно приближенно

рассматривать как плоские монохроматические. Такой выбор обусловлен и тем, что, в соответствии с принципом суперпозиции, общее решение волнового уравнения (1.1) можно представить в виде линейной комбинации плоских монохроматических волн. Последнее утверждение следует из линейности волнового уравнения. Кроме того, по теореме Фурье любая физически реализуемая функция может быть выражена через совокупность гармонических функций с определенными частотами, амплитудами и начальными фазами.

1.2.1. Формула плоской монохроматической световой волны

Как уже отмечалось, решением волнового уравнения (1.1) для поля E является сумма волновых функций (1.3) с разными знаками в аргументе. Убедиться в этом можно непосредственной подстановкой функции (1.3) в волновое уравнение (1.1). При этом следует учесть, что сумма квадратов направляющих косинусов равна единице.

Как и всякое дифференциальное уравнение, уравнение (1.1) не определяет явного вида функции $E(\mathbf{r}, t)$. Для определения конкретного вида этой функции (частного решения) волновое уравнение дополняется начальными и граничными условиями, соответствующими конкретной задаче. Неизвестная функция $E(\mathbf{r}, t)$ может быть определена либо с помощью граничного условия, т.е. заданием закона колебаний электрического поля в плоскости $\mathbf{r} = 0$ ($E(0, t) = f(t)$), либо с помощью начального условия по известной зависимости напряженности электрического поля E от координат в какой-либо момент времени, например в момент времени $t = 0$ ($E(\mathbf{r}, 0) = g(\mathbf{r})$).

Среди различных типов волн, описываемых соотношением $E(z, t) = f(t - \mathbf{n} \cdot \mathbf{r} / v)$ (взяв для краткости записи верхний знак в выражении $t \pm \mathbf{n} \cdot \mathbf{r} / v$), особое значение имеет плоская гармоническая волна.

Пусть колебания электрического поля в плоскости $\mathbf{r} = 0$, где расположен источник волн, происходят по гармоническому закону (граничное условие):

$$E(0, t) = E_{00} \cos(\omega t + \alpha),$$

где E_{00} , ω , α – соответственно амплитуда, циклическая частота и начальная фаза колебаний поля. Заменяв в этой формуле t на $t - \mathbf{n} \cdot \mathbf{r}/v$ (с учетом запаздывания колебаний на время $\mathbf{n} \cdot \mathbf{r} / v$ в плоскости $\mathbf{r} = \text{const}$), приходим к выражению, описывающему колебания поля E в произвольной плоскости $\mathbf{r} = \text{const} \neq 0$:

$$E(z, t) = E_{00} \cos[\omega(t - \mathbf{n} \cdot \mathbf{r}/v) + \alpha].$$

Учитывая, что $\omega / v = k$, будем иметь

$$E(\mathbf{r}, t) = E_{00} \cos(\omega t - k \mathbf{n} \cdot \mathbf{r} + \alpha)$$

Введем вектор \mathbf{k} , направленный перпендикулярно фронту волны в направлении ее распространения и равный по модулю волновому числу k : $\mathbf{k} = k\mathbf{n}$. Определяемый таким образом вектор \mathbf{k} называется волновым вектором. Этот вектор определяет направление распространения плоской световой волны. Модуль этого вектора равен $k = \omega / v = 2\pi / (vT) = 2\pi / \lambda$, где $\lambda = vT$ – длина световой волны, расстояние, на которое распространяется волна за один период колебаний светового поля. Как видим, волновое число имеет размерность обратной длины.

С учетом определения волнового вектора формулу волны, распространяющейся вдоль произвольного направления \mathbf{k} , можно записать в виде

$$E(\mathbf{r}, t) = E_{00} \cos(\omega t - \mathbf{k} \cdot \mathbf{r} + \alpha). \quad (1.6)$$

Это и есть формула плоской монохроматической световой волны. Как видим, такая волна создается гармоническими колебаниями источника.

Формула (1.6) показывает, что в плоской монохроматической волне зависимость компоненты вектора \mathbf{E} от радиус-вектора \mathbf{r} и времени t выражается гармонической функцией. Здесь E_{00} – амплитуда волны, аргумент косинуса, т.е. функция $\varphi(\mathbf{r}, t) = \omega t - \mathbf{k} \cdot \mathbf{r} + \alpha$ – фаза волны, $\alpha = \varphi(0, 0)$ – начальная фаза волны. Для какой-либо конкретной плоскости $\mathbf{r} = \mathbf{r}_1$ начальной фазой называют величину $\varphi(\mathbf{r}_1) = -\mathbf{k} \cdot \mathbf{r}_1 + \alpha$. Отметим, что фаза световой волны является одной из важнейших ее характеристик, определяющей результат взаимодействия когерентных световых волн друг с другом. Скалярное произведение $\mathbf{k} \cdot \mathbf{r}$, как и любое число, является величиной, не зависящей от ориентации координатных осей. А это и означает, что формула (1.6) описывает плоскую монохроматическую волну, распространяющуюся в произвольном относительно данной системы отсчета направлении (произвольна ориентация вектора \mathbf{n} по отношению к координатным осям). Функция (1.6) является простейшим решением волнового уравнения (1.1).

Зафиксируем какое-либо значение фазы

$$\omega t - \mathbf{k} \cdot \mathbf{r} + \alpha = C = \text{const.} \quad (1.7)$$

Отсюда видно, что каждому значению времени t соответствует определенное значение радиус-вектора \mathbf{r} , т.е. плоскость $\mathbf{k} \cdot \mathbf{r} = \omega t + \alpha - C$, перпендикулярная вектору \mathbf{k} . Это соотношение показывает, что фазовую (или волновую) поверхность световой волны можно определить как геометрическое место точек, в которых в рассматриваемый момент времени фаза волны имеет одно и то же значение, или как геометрическое место точек, в которых колебания светового вектора \mathbf{E} происходят в одной фазе. Из соотношения (1.7) видно также, что фазовая плоскость перемещается в пространстве вдоль направления вектора \mathbf{k} со скоростью $v = \omega/k$. Скорость перемещения поверхности постоянной фазы в пространстве называется фазовой скоростью волны. Расстояние, на которое перемещается волновая

поверхность за время, равное одному периоду колебаний светового поля, равно длине световой волны λ .

Монохроматическая световая волна вида (1.6) является периодической как во времени, так и в пространстве. Временным периодом световой волны является период колебаний светового поля в волне, равный периоду колебаний источника T . Пространственным периодом монохроматической волны является длина волны λ . Поэтому длину световой волны можно определить как расстояние между двумя ближайшими точками в волне, в которых колебания светового поля происходят в одинаковой фазе (с разностью фаз, равной или кратной 2π). Длина волны, в отличие от ее периода и частоты, задаваемых колебанием источника волн, различна в разных средах, так как в разных средах различна скорость света v . Поэтому длина волны в среде с показателем преломления $n = c/v$ определится как

$$\lambda = 2\pi v / \omega = (2\pi c / \omega) / n.$$

Здесь $2\pi c / \omega = \lambda_0$ – длина волны света в вакууме. Следовательно,

$$\lambda = \lambda_0 / n.$$

Так как $n > 1$, то $\lambda < \lambda_0$, причем, чем больше показатель преломления среды, тем более коротковолновым в этой среде является свет.

При математическом описании волновых явлений удобно использовать комплексную форму записи волны. В соответствии с формулой Эйлера

$$\cos\varphi \pm i \sin\varphi = \exp(\pm i\varphi)$$

и, следовательно,

$$\cos\varphi = \operatorname{Re} \{ \exp(\pm i\varphi) \},$$

где $i = \sqrt{-1}$ – мнимая единица. С учетом этого функцию (1.6) можно представить в виде действительной части комплексной функции

$$E(\mathbf{r}, t) = E_{00} \exp[-i(\omega t \pm \mathbf{k} \cdot \mathbf{r} + \alpha)],$$

где, исходя из соображений удобства, перед показателем степени экспоненты взят знак минус. Отметим, возможность использования комплексной формы записи обусловлена линейностью волнового уравнения. Представим это выражение в виде произведения

$$E(\mathbf{r}, t) = E_0 \exp(\pm i \mathbf{k} \cdot \mathbf{r}) \cdot \exp(-i\omega t). \quad (1.8)$$

Формула (1.8) записана для волны, распространяющейся вдоль вектора $\mathbf{k} > 0$ (при верхнем знаке в показателе первой экспоненты) и $\mathbf{k} < 0$ (при нижнем знаке). Величину

$$E(\mathbf{r}) = E_0 e^{\pm i \mathbf{k} \cdot \mathbf{r}} \quad (1.9)$$

называют комплексной амплитудой плоской волны, распространяющейся в направлении $\mathbf{k} > 0$ при знаке плюс в показателе экспоненты (перед произведением $i \mathbf{k} \cdot \mathbf{r}$) и в направлении $\mathbf{k} < 0$ – при знаке минус. Заметим, что в комплексном представлении волновые функции волн, распространяющихся в противоположных направлениях, являются комплексно сопряженными друг другу (отличаются знаком перед i). Множитель $E_0 = E_{00} \exp(-i\alpha)$, определяющий как амплитуду E_{00} , так и начальную фазу волны α , называют комплексной амплитудой в плоскости $\mathbf{r} = 0$, т.е. $E_0 = E(0)$. Формула (1.9) позволяет определить поле плоской волны в любой плоскости $\mathbf{r} = \text{const} > 0$ или $\mathbf{r} = \text{const} < 0$ по известному полю в плоскости $\mathbf{r} = 0$, и наоборот. Фазовый множитель $e^{\pm i \mathbf{k} \cdot \mathbf{r}}$ определяет набег фазы $k r$ плоской волны при ее распространении в свободном пространстве между двумя плоскостями, разделенными расстоянием r .

Если k_x, k_y, k_z – компоненты волнового вектора, то

$$\mathbf{k} \cdot \mathbf{r} = k_x x + k_y y + k_z z,$$

а комплексная амплитуда

$$E(x, y, z) = E_0 \exp [i (k_x x + k_y y + k_z z)]. \quad (1.10)$$

Отметим, что фаза колебаний светового поля является линейной функцией координат x, y, z . В плоскости $z = 0$ плоская волна с такой амплитудой создает распределение поля

$$E(x, y, 0) = E_0 \exp [i (k_x x + k_y y)].$$

В случае волны, распространяющейся вдоль оси Z , $E(x, y, 0) = E_0 = \text{const}$.

Комплексная амплитуда плоской волны имеет тот же физический смысл, что и комплексная амплитуда колебаний, но в отличие от последней содержит также множитель $\exp [i (k_x x + k_y y + k_z z)]$, определяющий зависимость фазы волны от пространственных координат. Физический смысл комплексности амплитуды связан с эллиптической поляризацией волны. Поверхность постоянной комплексной амплитуды представляет собой плоскость

$$\mathbf{k} \cdot \mathbf{r} = k_x x + k_y y + k_z z = C = \text{const},$$

перпендикулярную вектору \mathbf{k} и отсекающую на осях координат X, Y, Z отрезки длиной $C / k_x, C / k_y, C / k_z$ соответственно. Комплексная амплитуда плоской волны, распространяющейся вдоль оси Z , получается из амплитуды (1.10) как частный случай при $k_x = k_y = 0$ и $k_z = k$. Пусть $\cos \alpha, \cos \beta, \cos \gamma$ – направляющие косинусы вектора \mathbf{n} , а значит, и вектора \mathbf{k} . Тогда

$$k_x = k \cos \alpha = (2\pi / \lambda) \cos \alpha, \quad k_y = k \cos \beta = (2\pi / \lambda) \cos \beta,$$

$$k_z = k \cos \gamma = (2\pi / \lambda) \cos \gamma,$$

а комплексная амплитуда

$$E(x, y, z) = E_0 \exp(i \mathbf{k} \cdot \mathbf{r}) = E_0 \exp \left[i \frac{2\pi}{\lambda} (x \cos \alpha + y \cos \beta + z \cos \gamma) \right]. \quad (1.11)$$

Выражение (1.11) содержит в явном виде все параметры волны (кроме частоты ω): амплитуду E_0 , фазу и длину волны $\lambda = 2\pi / k$. Направление распространения волны задается направляющими косинусами $\cos \alpha$, $\cos \beta$, $\cos \gamma$. Ориентация поля может быть задана, если рассматривать амплитуду волны как векторную величину.

Если ориентировать систему координат так, чтобы, например, волновой вектор волны лежал в плоскости XZ и составлял угол φ с осью Z (рис. 1.1), то комплексная амплитуда такой волны будет иметь вид

$$E(x, z) = E_0 \exp [i(x \sin \varphi + z \cos \varphi)]. \quad (1.12)$$

Если волна распространяется вдоль оси Z , то $k_x = k_y = 0$, $k_z = k$, и тогда плоская гармоническая световая электромагнитная волна будет иметь вид

$$E(z, t) = E_{00} \cos(\omega t \pm kz + \alpha), \quad (1.13)$$

где нижний знак относится к волне, распространяющейся в положительном направлении оси Z , а верхний – в отрицательном направлении. Эта волна является решением одномерного волнового уравнения

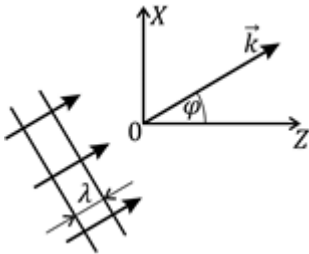


Рис. 1.1

запишется как

$$\frac{\partial^2 E}{\partial z^2} - \frac{1}{v^2} \frac{\partial^2 E}{\partial t^2} = 0. \quad (1.14)$$

В комплексной форме волна (1.13)

$$E(z) = E_0 e^{\pm ikz}, \quad (1.15)$$

где $E_0 = E_{00} e^{ik\alpha}$ – начальная комплексная амплитуда.

В теоретических построениях часто используются параксиальные плоские волны. Параксиальными плоскими волнами называют такие плоские волны, для которых направление распространения близко к направлению оси Z , т.е. когда угол γ мал, а углы α и β близки к $\pi/2$. Введя углы $\varphi = \pi/2 - \alpha$ и $\theta = \pi/2 - \beta$, дополняющие углы α и β до $\pi/2$, и учтя, что эти углы малы, будем иметь

$$\cos\gamma \approx 1, \quad \cos\alpha = \sin\varphi \approx \varphi, \quad \cos\beta = \sin\theta \approx \theta.$$

В этом случае комплексная амплитуда плоской волны примет вид

$$\begin{aligned} E(x, y, z) &= E_0 \exp \left[i \frac{2\pi}{\lambda} (x \cos \alpha + y \cos \beta + z \cos \gamma) \right] = \\ &= E_0 \exp \left[i \frac{2\pi}{\lambda} (\varphi x + \theta y + z) \right]. \end{aligned}$$

Это выражение можно представить в виде двух сомножителей:

$$E(x, y, z) = E_0 \exp \left(i \frac{2\pi}{\lambda} z \right) \exp \left[i \frac{2\pi}{\lambda} (\varphi x + \theta y) \right].$$

Здесь первый фазовый множитель $\exp \left(i \frac{2\pi}{\lambda} z \right)$ указывает на то, что волна, распространяясь почти вдоль оси Z , приобретает вдоль этой оси набег фазы, равный $\frac{2\pi}{\lambda} z$ (запаздывает на время

$$t = \frac{k}{\omega} z), \text{ а второй фазовый множитель } \exp \left[i \frac{2\pi}{\lambda} (\varphi x + \theta y) \right]$$

определяет распределение комплексной амплитуды поля в

плоскости, перпендикулярной оси Z . В частном случае, когда волновой вектор \mathbf{k} лежит в плоскости XZ под углом φ к оси Z (и значит, $\theta = 0$), для параксиальной плоской волны получаем

$$E(x, z) = E_0 \exp\left(i \frac{2\pi}{\lambda} z\right) \exp\left(i \frac{2\pi}{\lambda} \varphi x\right).$$

Плоская монохроматическая волна представляет собой синусоиду, имеющую бесконечную протяженность и длительность, т.е. не ограничена в пространстве и времени ($-\infty < t, z < \infty$ для волны, распространяющейся вдоль оси Z). Свойства ее в любой точке и в любой момент времени одинаковы: постоянны ее амплитуда, частота и начальная фаза. Плоская монохроматическая волна могла бы возбуждаться электрическими зарядами, совершающими бесконечно длительные гармонические колебания. В действительности же никакие реальные колебания не продолжаются бесконечно долго, все они имеют начало и конец. Следовательно, ограничены во времени и в пространстве и все реальные волны, а такие волны монохроматическими, а значит, и плоскими монохроматическими волнами не являются. Поэтому на практике под монохроматическими и, в частности, под плоскими монохроматическими волнами подразумевают волны, выражающиеся гармоническими функциями в конечном интервале времени и координат. Эти временные и пространственные интервалы должны быть много больше соответственно периода и длины волны. Чем лучше выполнены эти условия, тем с большим основанием можно пользоваться приведенными выше формулами плоских волн.

1.2.2. Поляризация плоской волны

В электромагнитной волне как волне поперечной электрические и магнитные поля колеблются в плоскости, перпендикулярной направлению распространения волны, т.е. перпендикулярно волновому вектору \mathbf{k} . Однако направление

колебаний векторов \mathbf{E} и \mathbf{B} в этой плоскости может быть произвольным. Поэтому для полного описания плоской поперечной монохроматической волны, кроме ее амплитуды, частоты и фазы, необходимо учитывать еще состояние ее поляризации, т.е. направление колебаний электрического \mathbf{E} и связанного с ним магнитного \mathbf{B} векторов. В этом проявляется векторный характер электромагнитной волны. Если вектор \mathbf{E} волны все время колеблется только в одной плоскости, содержащей вектор \mathbf{k} , ориентация которой фиксирована в пространстве, то такую волну называют плоско поляризованной. Ее называют также линейно поляризованной, так как если смотреть навстречу волне (навстречу вектору \mathbf{k}), колебания вектора \mathbf{E} в ней будем видеть происходящими вдоль одной линии, перпендикулярной направлению распространения волны. Линейная поляризация показана на рис. 1.2, а. Плоскость, в которой колеблется вектор \mathbf{E} , называют плоскостью поляризации (иногда – плоскостью колебаний). В векторном виде колебания векторов \mathbf{E} и \mathbf{B} в плоской монохроматической волне можно записать как

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{e} E_{00} \cos (\omega t - \mathbf{k} \cdot \mathbf{r} + \alpha),$$

$$\mathbf{B}(\mathbf{r}, t) = \underline{\mathbf{n}} \times \mathbf{e} B_{00} \cos (\omega t - \mathbf{k} \cdot \mathbf{r} + \alpha).$$

Здесь \mathbf{e} – вещественный единичный вектор, направленный вдоль направления колебаний электрического вектора \mathbf{E} волны и называемый вектором поляризации.

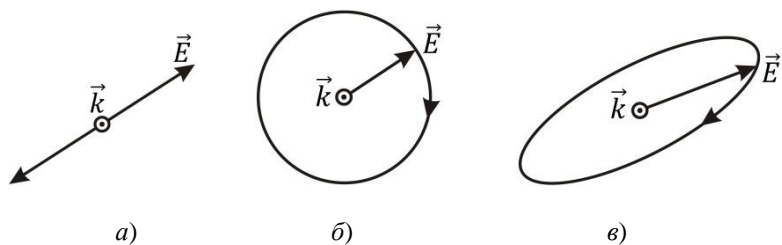


Рис. 1.2

Линейная поляризация не является единственным видом поляризации плоской монохроматической волны. Для определения других возможных поляризационных состояний учтем, что из уравнений Максвелла вытекает независимость волн с компонентами E_x, B_y и E_y, B_x . Это означает, что в общем случае плоская волна может быть представлена в виде суперпозиции двух линейно поляризованных волн с взаимно перпендикулярной ориентацией одноименных полей.

Рассмотрим суперпозицию двух плоских монохроматических волн одинаковой частоты, линейно поляризованных в двух взаимно перпендикулярных направлениях (например, в направлении осей X и Y) и распространяющихся в направлении оси Z . Тогда в каждой точке этой оси будет происходить сложение двух взаимно перпендикулярных колебаний одинаковой частоты:

$$E_x = E_{01} \cos(\omega t + \varphi_1), \quad E_y = E_{02} \cos(\omega t + \varphi_2).$$

В зависимости от разности фаз $\delta = \varphi_2 - \varphi_1$ и соотношения амплитуд E_{01} и E_{02} получаются следующие виды поляризации: линейная поляризация (при $\delta = 0$ или π), эллиптическая правая поляризация (при $\delta = \pi / 2, E_{01} \neq E_{02}$), эллиптическая левая поляризация (при $\delta = -\pi / 2, E_{01} \neq E_{02}$), круговая правая поляризация (при $\delta = \pi / 2, E_{01} = E_{02}$), круговая левая поляризация (при $\delta = -\pi / 2, E_{01} = E_{02}$). При эллиптической или круговой поляризации конец результирующего вектора \mathbf{E} движется в плоскости XY по эллипсу или по кругу по часовой стрелке (если смотреть навстречу волне) при правой поляризации и против часовой стрелки – при левой. В пространстве в световой волне, поляризованной по кругу или эллипсу, конец вектора \mathbf{E} описывает круговую или эллиптическую спираль, накручивающуюся на направление луча. Конец вектора магнитного поля волны \mathbf{B} при этом также описывает эллипс или окружность в той же плоскости XY , в которой движется вектор \mathbf{E} .

При таких поляризациях векторы \mathbf{E} и \mathbf{B} в каждой точке z вращаются с угловой скоростью ω , оставаясь неизменными по модулю. Уравнения прямой, эллипса или окружности (в координатах E_x и E_y), как легко убедиться, получаются из приведенных выше формул для E_x и E_y при исключении из них параметра t .

Эллиптическая поляризация является наиболее общим случаем поляризации электромагнитной волны. Она возникает при любой постоянной разности фаз δ . При разности фаз $\delta = \pm \pi / 2$ получается эллипс, приведенный к осям координат. Все другие виды поляризации получаются при соответствующей разности фаз δ и соотношении между амплитудами E_{01} и E_{02} .

Круговая и эллиптическая поляризации представлены соответственно на рис. 1.2, б и 1.2, в. Показаны правые круговая и эллиптическая поляризации.

Поляризация, как и амплитуда, и фаза, является важной информационной характеристикой электромагнитной волны. Устройства для управления поляризацией волны называются поляризаторами.

Векторный характер составляющих электрического и магнитного поля световой электромагнитной волны имеет значение только при рассмотрении явлений, связанных с поперечностью электромагнитных волн таких как поляризация, двойное лучепреломление и т.п. При прохождении света через оптические системы векторный характер световой волны практически не играет роли. А поскольку нас будут интересовать именно эти вопросы, то в дальнейшем мы будем рассматривать только скалярные величины: составляющие векторов электрического и магнитного полей.

1.3. Сферические световые волны

Другим простейшим решением волнового уравнения (1.1), но выраженного в сферических координатах, является сферическая

волна. Как и плоская, сферическая волна часто используется при описании волновых явлений. Но, как и плоская, сферическая монохроматическая волна является математической абстракцией. Никакая реальная волна строго монохроматической плоской или сферической не является. Все эти волны вводятся в рассмотрение лишь в качестве предельных случаев, удобных для описания физических явлений в теоретических модельных расчетах.

В сферической волне проекция вектора \mathbf{E} на направление, вдоль которого он колеблется, описывается функциями

$$E(r, t) = \frac{f(t \pm r/v)}{r}, \quad (1.16)$$

где r – расстояние от источника (расположенного в начале координат) до произвольной точки наблюдения $P(x, y, z)$, отсчитываемое вдоль направления распространения волны, т.е. радиус сферического фронта волны, v – скорость распространения волны. Сферический фронт волна может иметь только в изотропной среде, т.е. в такой среде, величина скорости волны v в которой по всем направлениям одинакова, и на достаточно большом расстоянии от источника. Волновые функции (1.16) являются общим решением волнового уравнения

$$\frac{\partial^2(rE)}{\partial r^2} - \frac{1}{v^2} \frac{\partial^2(rE)}{\partial t^2} = 0,$$

которое получается из уравнения (1.1), если в нем перейти к сферическим координатам с учетом того, что в сферической волне поле E зависит только от одной сферической координаты $r = \sqrt{x^2 + y^2 + z^2}$, а от углов θ и φ не зависит. Это уравнение имеет такой же вид, как и уравнение (1.14), но вместо E здесь стоит произведение rE , поэтому общим решением этого уравнения будет

$$rE = f_1(t - r/v) + f_2(t + r/v),$$

откуда и получается выражение (1.16). Монохроматическая сферическая волна является частным решением указанного волнового уравнения и описывается волновой функцией

$$E(r,t) = \frac{E_{00}}{r} \cos(\omega t \pm kr + \alpha),$$

где k – волновое число ($k = \omega / v$), r – радиус сферического волнового фронта, E_{00} – постоянная, имеющая смысл амплитуды волны на единичном расстоянии от источника, ω – циклическая частота, α – начальная фаза волны. Знак минус в фазе волны относится к расходящейся из начала координат сферической волне, а плюс – к сферической волне, сходящейся в начале координат.

Поверхность постоянной фазы сферической волны

$$\omega t \pm kr + \alpha = \text{const}$$

представляет собой сферу, радиус которой с течением времени увеличивается при знаке минус перед множителем kr в фазе волны ($dr / dt = \omega / k > 0$) и убывает при знаке плюс ($dr / dt = -\omega / k < 0$). Скорость перемещения сферической поверхности в пространстве $v = dr / dt = \omega / k$. В сферической волне сферами являются и поверхности постоянной амплитуды.

Сферические световые волны возникают, если во всех точках некоторой сферической поверхности возбудить одинаковые синфазные колебания светового поля, передающиеся в однородной и изотропной среде. Расходящуюся сферическую волну излучает точечный источник (расположенный на бесконечности, он будет давать плоскую волну). Монохроматическая сферическая световая волна получается, если источник совершает гармонические колебания.

Сферической электромагнитной волной является, например, волна, излучаемая колеблющимся электрическим диполем на расстояниях $r \gg \lambda$ (в волновой зоне диполя):

$$E(r, t) = -\frac{1}{4\pi\epsilon_0} \frac{\omega^2 p_0 \sin \theta}{c^2 r} \cos(\omega t - k r + \alpha),$$

где ω и p_0 – соответственно частота и амплитуда колебаний диполя, θ – угол между осью диполя и направлением на точку наблюдения.

В комплексной форме зависимость $E(r, t)$ записывается в виде

$$\begin{aligned} E(r, t) &= \frac{E_0}{r} \exp[-i(\omega t \pm k r)] = \\ &= \frac{E_0}{r} \exp(\pm i k r) \cdot \exp(-i \omega t) = E(r) \cdot \exp(-i \omega t). \end{aligned}$$

Здесь

$$E(r) = \frac{E_0}{r} \exp(\pm i k r)$$

– комплексная амплитуда сферической волны, а $E_0 = E_{00} \exp(-i \alpha)$ – комплексная постоянная, модуль которой определяет амплитуду волны E_{00} на единичном расстоянии от источника, а аргумент – начальную фазу α . Знак плюс в показателе экспоненты комплексной амплитуды соответствует расходящейся, а минус – сходящейся сферическим волнам. Заметим, что комплексное сопряжение переводит формулу расходящейся сферической волны в формулу сходящейся, или наоборот.

Если точечный источник находится в начале координат, то расстояние $r = \sqrt{x^2 + y^2 + z^2}$, а комплексная амплитуда сферической волны будет иметь вид

$$E(x, y, z) = E_0 \frac{\exp(\pm i k \sqrt{x^2 + y^2 + z^2})}{\sqrt{x^2 + y^2 + z^2}}. \quad (1.17)$$

Положив в этой формуле $z = 0$, найдем распределение поля, создаваемое сферической волной в плоскости $z = 0$:

$$E(x, y, 0) = E_0 \frac{\exp(\pm ik \sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}},$$

Положив в той же формуле (1.17) $x = y = 0$, найдем распределение поля вдоль оси Z :

$$E(0, 0, z) = \frac{E_0}{z} \exp(\pm ikz). \quad (1.18)$$

Мы получили плоскую волну с амплитудой, убывающей обратно пропорционально расстоянию z от источника.

Рассмотрим приближение, при котором $x^2 + y^2 \ll z^2$, т.е. когда координаты x и y малы по сравнению с координатой z , т.е. мал размер интересующей нас области по сравнению с расстоянием z . В этом приближении ограничиваются рассмотрением участков волнового фронта, лежащих вблизи оси Z (рис. 1.3). В случае сферической волны эта область представляет собой конус с вершиной в точке расположения источника и основанием в плоскости наблюдения в виде круга, радиус которого $R = \sqrt{x^2 + y^2} \ll z$, т.е. с малым телесным углом при вершине. Такое приближение называют параксиальным приближением. Запишем выражение $r = \sqrt{x^2 + y^2 + z^2}$ в виде

$$r = z \sqrt{1 + \frac{x^2 + y^2}{z^2}}. \quad \text{В параксиальном приближении радикал}$$

можно разложить в ряд по степеням величины $\alpha = \frac{x^2 + y^2}{z^2}$ (

$$\sqrt{1 + \alpha} = 1 + \frac{1}{2}\alpha - \frac{1}{8}\alpha^2 + \dots, \quad |\alpha| < 1)$$

и ограничиться первыми двумя членами. Тогда получим

$$r = z \left(1 + \frac{1}{2} \frac{x^2 + y^2}{z^2} - \dots \right) \approx z + \frac{x^2 + y^2}{2z}.$$

С учетом этого комплексная амплитуда (1.18) в параксиальном приближении примет вид

$$E(x, y, z) \approx \frac{E_0}{z} e^{\pm ikz} \exp \left(\pm ik \frac{x^2 + y^2}{2z} \right).$$

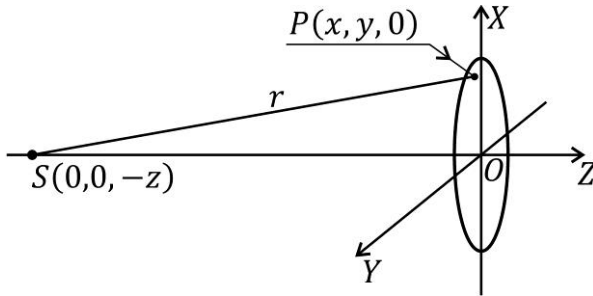


Рис. 1.3

В этом выражении коэффициент E_0 / z показывает, что амплитуда сферической волны обратно пропорциональна расстоянию z ; множитель $\exp(\pm ikz)$ характеризует изменение (набег) фазы волны при прохождении волной расстояния z (запаздывание волны на время $t = (k / \omega) z$). Распределение поля в плоскости XY описывается квадратичным фазовым множителем $\exp \left(\pm ik \frac{x^2 + y^2}{2z} \right)$. Так как x и y малы, то радиус сферической

волны r приближенно можно принять равным z .

Уравнение изофазной поверхности в параксиальном приближении

$$z + \frac{x^2 + y^2}{2z} = \text{const}$$

определяет в пространстве параболоид вращения. Следовательно, в рассматриваемом приближении сферические волновые поверхности превращаются в параболические с кривизной при вершине, совпадающей с кривизной сферической волны. Поэтому параксиальное приближение часто называют также параболическим приближением.

Если источник сферической волны находится не в начале координат, а в точке $S(x_0, y_0, z_0)$, то

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}.$$

В частном случае, когда источник располагается на оси Z в точке $(0, 0, z_0)$, $r = \sqrt{x^2 + y^2 + (z - z_0)^2}$, а комплексная амплитуда поля, создаваемого такой сферической волной в плоскости $z = \text{const}$, определится выражением

$$E(x, y, z) = E_0 \frac{\exp(\pm ik \sqrt{x^2 + y^2 + (z - z_0)^2})}{\sqrt{x^2 + y^2 + (z - z_0)^2}}.$$

Положив в этом выражении $z = 0$, найдем поле в плоскости $z = 0$. В этом случае $r = \sqrt{x^2 + y^2 + z_0^2}$. В параксиальном приближении, когда $x^2 + y^2 \ll z_0^2$, будем иметь

$$E(x, y, 0) \cong \text{const} \cdot \exp\left(\pm ik \frac{x^2 + y^2}{2z_0}\right), \quad (1.19)$$

где $\text{const} = (E_0 / z_0) e^{\pm ikz_0}$, знак плюс относится к расходящейся, а минус – к сходящейся волне; радиус волны приближенно равен z_0 . Отметим, что в параксиальном приближении радиус волны

одновременно является и z -координатой центра ее расхождения или схождения.

1.4. Цилиндрические световые волны

Еще одним предельным случаем волн являются цилиндрические волны. Они являются решением волнового уравнения (1.1), записанного в цилиндрических координатах. При описании волновых явлений цилиндрические волны используются реже, чем рассмотренные выше плоские и сферические волны. Как и две предыдущие, эти волны являются физической идеализацией, моделью, удобной в некоторых теоретических построениях. Строго цилиндрической волной, как и строго плоской или строго сферической, реальная волна не является. Поэтому все эти идеализации имеют ограниченную применимость и во многих случаях волновой оптики оказываются недостаточными. Плоские, сферические и цилиндрические волны часто называют эталонными волнами.

Цилиндрические волны обладают цилиндрической симметрией; их волновые поверхности имеют цилиндрическую форму: $R = \text{const}$, где R – расстояние от оси симметрии волны до произвольной точки волновой поверхности. Цилиндрический фронт волна может иметь только в изотропной среде. Цилиндрические световые волны возникают, если во всех точках некоторой цилиндрической поверхности возбудить одинаковые синфазные колебания светового поля, передающиеся в изотропной среде. В цилиндрической световой волне проекция вектора E на направление его колебания описывается формулой

$$E(R, t) = \frac{f(t \pm R/v)}{\sqrt{R}}.$$

Если источник совершает гармонические колебания, то излучаемые им цилиндрические волны будут

монохроматическими. Монохроматическая цилиндрическая волна описывается функцией

$$E(R, t) = \frac{E_{00}}{\sqrt{R}} \cos(\omega t \pm k R + \alpha),$$

где E_{00} – постоянная, имеющая тот же смысл, что и аналогичная величина для сферической волны (амплитуда волны на единичном расстоянии от источника), знак минус в фазе волны соответствует расходящейся, а плюс – сходящейся цилиндрическим волнам. Расходящуюся цилиндрическую волну излучает линейный источник света, например, светящаяся нить или щель (прямоугольное отверстие один из размеров которого много больше другого). Волновые поверхности в расходящейся цилиндрической волне удаляются от оси симметрии, а в сходящейся – приближаются к оси симметрии с постоянной скоростью $v = \omega / k$. По мере увеличения расстояния R кривизна волновой поверхности цилиндрической волны уменьшается, и поэтому ее, как и волну сферическую, на достаточно большом расстоянии от источника можно рассматривать как плоскую волну.

В комплексной форме волновая функция цилиндрической волны имеет такой же вид, что и волновая функция сферической волны, с той лишь разницей, что вместо амплитудного множителя $1/r$ здесь стоит множитель $1/\sqrt{R}$, а в фазовом множителе вместо r стоит R :

$$\begin{aligned} E(R, t) &= \frac{E_{00}}{\sqrt{R}} \exp[-(\omega t \pm k R + \alpha)] = \\ &= \frac{E_{00}}{\sqrt{R}} \exp(\pm ikR) \cdot \exp(-i\alpha) \cdot \exp(-i\omega t). \end{aligned}$$

Комплексная амплитуда цилиндрической волны

$$E(R) = \frac{E_0}{\sqrt{R}} \exp(\pm ikR), \quad (1.20)$$

где $E_0 = E_{00} \exp(-i\varphi_0)$ – комплексная постоянная, имеющая такой же смысл, что и аналогичная постоянная в формуле сферической волны. Знак плюс в фазовом множителе в выражении (1.20) соответствует расходящейся волне, а знак минус – сходящейся волне. И в случае цилиндрической волны комплексное сопряжение переводит формулу расходящейся волны в формулу сходящейся, или наоборот.

Совместим координатную ось Y с осью симметрии цилиндрической волны, тогда $R = \sqrt{x^2 + z^2}$. Комплексную амплитуду цилиндрической волны в этом случае можно записать в координатной форме:

$$E(x, z) = E_0 \frac{\exp(\pm ik\sqrt{x^2 + z^2})}{\sqrt{\sqrt{x^2 + z^2}}}.$$

Заметим, что от фазы сферических волн можно перейти к фазе цилиндрических волн, если произвести формальную замену $\sqrt{x^2 + y^2 + z^2}$ на $\sqrt{x^2 + z^2}$, т.е. если в фазе сферической волны положить $y = 0$. В плоскости $z = 0$ волна с указанной амплитудой создаст распределение поля

$$E(x, 0) = E_0 \frac{\exp(\pm ikx)}{\sqrt{x}}.$$

Заметим, что фаза этого распределения линейно зависит от поперечной координаты x , а амплитуда обратно пропорциональна \sqrt{x} . Вдоль оси Z цилиндрическая волна создает распределение амплитуды поля

$$E(0, z) = E_0 \frac{\exp(\pm ikz)}{\sqrt{z}}.$$

Это – плоская волна с амплитудой, убывающей по закону $1/\sqrt{z}$.

В параксиальном приближении, когда $x \ll z$, радиус волнового фронта R в амплитудном множителе можно заменить на z , а в фазовом – на $R \approx z + \frac{x^2}{2z}$. Комплексная амплитуда цилиндрической волны примет вид

$$E(x, z) \cong E_0 \frac{\exp(\pm ikz)}{\sqrt{z}} \exp\left(\pm ik \frac{x^2}{2z}\right).$$

Область, соответствующая параксиальному приближению, в плоскости XY' будет представлять собой прямоугольную полосу, ширина которой будет много меньше расстояния от оси симметрии волны до указанной плоскости.

Если ось симметрии волны направлена вдоль прямой $z = z_0$, то $R = \sqrt{x^2 + (z - z_0)^2}$. Для определения комплексной амплитуды поля в плоскости $z = 0$ положим в этой формуле $z = 0$. Тогда $R = \sqrt{x^2 + z_0^2}$. Комплексная амплитуда в этом случае будет определяться формулой

$$E(x, 0) = E_0 \frac{\exp(\pm ik \sqrt{x^2 + z_0^2})}{\sqrt{\sqrt{x^2 + z_0^2}}}.$$

В параксиальном приближении, когда $x^2 \ll z_0^2$, распределение поля в плоскости $z = 0$ будет иметь вид

$$E(x, 0) \approx \text{const} \cdot \exp\left(\pm ik \frac{x^2}{2z_0}\right),$$

где $\text{const} = \frac{E_0}{\sqrt{z_0}} \exp(\pm ikz_0)$.

Уравнение изофазной поверхности цилиндрической волны

$$z + \frac{x^2}{2z} = \text{const}$$

в параксиальном приближении представляет собой параболу с кривизной при вершине, равной кривизне цилиндрической волны.

1.5. Волновое скалярное уравнение Гельмгольца

В плоской монохроматической волне амплитуда колебаний поля не зависит от положения точки наблюдения. В других монохроматических волнах (например, в сферических и цилиндрических волнах) такая зависимость существует. Поэтому в общем случае монохроматическую волну можно представить гармонической функцией

$$E(\mathbf{r}, t) = E_0(\mathbf{r}) \cos [\omega t + \varphi(\mathbf{r})], \quad (1.21)$$

где $\mathbf{r}\{x, y, z\}$ – радиус-вектор точки наблюдения, ω – круговая, или циклическая частота волны, $E_0(\mathbf{r})$ и $\varphi(\mathbf{r})$ – соответственно амплитуда и фаза волны в точке $\mathbf{r}\{x, y, z\}$.

Плоскую монохроматическую волну можно рассматривать как частный случай монохроматической волны этого общего вида, когда область применения последней становится бесконечно малой. Действительно, в этом случае можно считать E_0 постоянной, а $\varphi(\mathbf{r}) = \varphi(x, y, z)$ разложить в ряд в окрестности какой-либо точки (x_0, y_0, z_0) , принадлежащей этой области:

$$\varphi(x, y, z) = \varphi(x_0, y_0, z_0) + a(x - x_0) + b(y - y_0) + c(z - z_0)$$

с отбрасыванием членов высшего порядка малости, где a , b и c – частные производные φ по координатам x , y и z соответственно.

Фаза $\varphi(x, y, z)$ в этом разложении имеет вид фазы плоской волны отличается только постоянной величиной $\varphi(x_0, y_0, z_0) + ax_0 + by_0 + cz_0$, не имеющей принципиального значения, так как она соответствует общему изменению фазы для всех точек области. Следует отметить, что у плоской волны, описываемой формулой (1.21), поверхности постоянной фазы не совпадают с поверхностями постоянной амплитуды. Такие волны называют неоднородными волнами.

В любой фиксированной точке наблюдения $\mathbf{r}\{x, y, z\}$ волна (1.21) создает гармонические колебания поля E с частотой ω , амплитудой $E(\mathbf{r})$ и начальной фазой $\varphi(\mathbf{r})$. Используя формулу Эйлера, функцию (1.21) можно представить в виде действительной части комплексной функции

$$E(\mathbf{r}, t) = E(\mathbf{r}) \exp(-i\omega t),$$

где $E(\mathbf{r}) = E_0(\mathbf{r}) \exp[-i\varphi(\mathbf{r})]$, или, выразив поле через декартовы координаты,

$$E(x, y, z, t) = E(x, y, z) \exp(-i\omega t). \quad (1.22)$$

Здесь

$$E(x, y, z) = E_0(x, y, z) \exp[-i\varphi(x, y, z)] \quad (1.23)$$

– комплексная функция, описывающая пространственную структуру данного волнового поля, т.е. распределение в пространстве амплитуд и фаз колебаний (а также и форму фазовой поверхности волны). Эту функцию называют комплексной амплитудой. Модуль комплексной амплитуды световой волны определяет амплитуду волны $E_0(x, y, z)$, а аргумент – ее фазу $\varphi(x, y, z)$. Таким образом, комплексная амплитуда (при заданной частоте) содержит всю информацию о волне. Экспоненциальный множитель в (1.23) называют фазовым множителем.

При записи монохроматической волны в виде (1.22) гармоническая функция заменяется действительной частью комплексной экспоненты:

$$\cos(\omega t + \alpha) = \operatorname{Re}\{\exp[-i(\omega t + \alpha)]\}.$$

Возможность комплексного представления электрического поля волны обусловлена, как уже отмечалось, линейностью волнового уравнения (1.1). Представляя волновую функцию $E(x, y, z, t)$ в комплексной форме, следует, однако, помнить, что в действительности физическая величина напряженность электрического поля есть величина вещественная, поэтому после проведения всех математических преобразований из результата произведенных операций должна выделяться действительная часть.

Свет, падающий на объект, может поглощаться, отражаться, рассеиваться или проходить через объект, что сопровождается изменением амплитуды и фазы световой волны. Амплитуда и фаза светового возмущения, прошедшего через объект (или отраженного объектом), изменяется от точки к точке в плоскости, параллельной плоскости объекта. Это распределение амплитуды и фазы можно описать некоторой комплексной функцией $E(x, y, z)$ – комплексной амплитудой. Комплексная амплитуда светового поля $E(x, y, z)$ несет информацию об объекте, поэтому в прикладной оптике ее называют оптическим сигналом. Ее называют также пространственным сигналом, так как она зависит только от пространственных координат x, y, z .

Напомним, что под сигналом вообще понимают всякий физический процесс, несущий информацию. Информация – это совокупность сведений и данных, являющихся объектом хранения, передачи и преобразования. Информацию передают в виде сообщений. Сообщение – это форма представления информации (текст, речь, музыка, изображение и т.п.), предназначенная для передачи от источника к пользователю (потребителю информации). Сигналы могут иметь различную физическую природу. Они, например, могут быть акустическими,

электрическими, электромагнитными, в том числе световыми. Электрический сигнал – это переменный ток или напряжение, параметры (амплитуда, начальная фаза, частота) которого изменяются во времени по определенному закону (в соответствии с передаваемым сообщением). Электромагнитным сигналом называют модулированную электромагнитную волну, несущую информацию. Если эта волна принадлежит радиодиапазону, электромагнитный сигнал называют радиосигналом, если видимому – световым сигналом.

Обычно рассматривают распределение поля в плоскости $z = 0$ (или в любой фиксированной плоскости $z = \text{const}$). Тогда комплексная амплитуда в этой плоскости

$$E(x, y) = |E(x, y)| \exp[-i\varphi(x, y)].$$

Эта комплексная функция описывает двумерный оптический сигнал. При этом действительная функция $|E(x, y)|$ описывает информацию, заложенную в амплитуде световой волны (амплитудную информацию), а действительная функция $\varphi(x, y)$ – информацию, заложенную в фазе световой волны (фазовую информацию).

Получим теперь уравнение, которому удовлетворяет комплексная амплитуда волны. Для этого продифференцируем дважды комплексную функцию (1.22) сначала по координатам x , y , z , а затем по времени t . Тогда получим

$$\Delta E(x, y, z, t) = e^{-i\omega t} \Delta E(x, y, z),$$

$$\frac{\partial^2 E(x, y, z, t)}{\partial t^2} = (-i\omega)^2 e^{-i\omega t} E(x, y, z) = \omega^2 e^{-i\omega t} E(x, y, z).$$

Подставляя полученные выражения производных в волновое уравнение (1.1), приходим к искомому уравнению для комплексной амплитуды светового поля:

$$\Delta E(x, y, z) + k^2 E(x, y, z) = 0, \quad (1.24)$$

где $k = \omega/v = 2\pi/\lambda$ – волновое число. Уравнение (1.24) описывает распространение комплексной амплитуды и называется волновым скалярным уравнением Гельмгольца для свободного пространства. Иногда уравнение Гельмгольца записывают в виде

$$(\Delta + k^2) E(x, y, z) = 0.$$

В одномерном случае, когда комплексная амплитуда поля зависит от одной какой-либо переменной ξ , где через ξ обозначена переменная x , y или z , уравнение Гельмгольца имеет вид

$$\frac{d^2 E(\xi)}{d\xi^2} + k^2 E(\xi) = 0. \quad (1.25)$$

Непосредственной подстановкой легко убедиться, что комплексная амплитуда плоской одномерной волны (1.15) удовлетворяет одномерному уравнению Гельмгольца (1.25), когда $\xi = z$, а волны (1.23) – общему уравнению Гельмгольца (1.24) при условии, что

$$k_x^2 + k_y^2 + k_z^2 = k^2 = (\omega/c)^2 = (2\pi/\lambda)^2.$$

Комплексная амплитуда сферической волны подчиняется волновому уравнению Гельмгольца, выраженному через сферические координаты, а цилиндрической – через цилиндрические координаты.

Для определения конкретного вида функции $E = f(\mathbf{r})$ уравнение Гельмгольца должно дополняться двумя граничными условиями: значением поля $E(\mathbf{r})$ в какой-либо точке $\mathbf{r}_0(x_0, y_0, z_0)$ и значениями первой производной функции $E(\mathbf{r})$ по координатам в этой точке.

Вследствие линейности уравнения Гельмгольца для комплексных амплитуд справедлив принцип суперпозиции: комплексная амплитуда поля, созданного несколькими

источниками, равна сумме комплексных амплитуд полей от каждого источника в отдельности. Граничные условия при этом также суммируются.

1.6. Пространственные частоты

Понятие пространственной частоты является одним из важных понятий фурье-оптики. Чтобы ввести это понятие, запишем сначала комплексную амплитуду плоской волны, распространяющейся вдоль координатной оси X :

$$E(x) = E_0 \exp\left(i \frac{2\pi}{\lambda} x\right) = E_0 \exp(i2\pi ux), \quad (1.26)$$

где $u = 1 / \lambda$. Заметим, что выражение (1.26) аналогично формуле гармонического колебания

$$E(t) = E_0 \exp\left(i \frac{2\pi}{T} t\right) = E_0 \exp(i2\pi vt),$$

где $\nu = 1 / T$ – частота колебаний, причем роль временной координаты t играет пространственная координата x , а роль частоты ν – величина, обратная длине волны, $u = 1 / \lambda$. Так как длина волны λ – это пространственный период волны, то, в соответствии с указанным замечанием, величину u следует назвать пространственной частотой волны. Перепишем теперь выражение (1.10) в виде

$$E(x, y, z) = E_0 \exp [i2\pi (ux + vy + wz)], \quad (1.27)$$

где

$$u = \cos \alpha / \lambda, \quad v = \cos \beta / \lambda, \quad w = \cos \gamma / \lambda. \quad (1.28)$$

На комплексную амплитуду (1.27) можно смотреть как на трехмерное гармоническое колебание с пространственными периодами вдоль координатных осей X, Y, Z соответственно

$$d_x = 1 / u, d_y = 1 / v, d_z = 1 / w.$$

Поэтому величины u, v, w , определяемые равенствами (1.28), можно называть пространственными частотами вдоль координатных осей X, Y, Z соответственно. Таким образом, пространственные частоты связаны с соответствующими пространственными периодами волны, измеренными вдоль осей X, Y и Z , подобно то-

му, как связана линейная частота колебания ν с его периодом T .

Введенная на основании соотношения (1.28) пространственная частота $u = 1 / \lambda$ получается из общего выражения $u = \cos \alpha / \lambda$ при угле $\alpha = 0$ (при этом $\beta = \gamma = \pi / 2$). Измеряются пространственные частоты либо в периодах на миллиметр (пер. / мм), либо в линиях на миллиметр (лин. / мм), либо просто в $1 / \text{мм}$.

Пространственные частоты часто выражают через углы

$$\varphi = \pi / 2 - \alpha, \theta = \pi / 2 - \beta, \phi = \pi / 2 - \gamma,$$

дополняющие направляющие углы α, β и γ до $\pi / 2$. Тогда они записываются в виде

$$u = \sin \varphi / \lambda, v = \sin \theta / \lambda, w = \sin \phi / \lambda.$$

В качестве примера на рис. 1.4 показаны пространственные частоты u и w волны, распространяющейся в плоскости XZ . На этом рисунке параллельные прямые изображают перпендикулярные плоскости XZ волновые поверхности плоской волны, волновой вектор которой равен \mathbf{k} . Они характеризуют мгновенную картину световой волны. Расстояние между соседними волновыми поверхностями равно длине волны λ . Расстояния между этими поверхностями (максимумами или минимумами амплитуды), отсчитываемые вдоль координатных

осей X и Z , т.е. между точками пересечения соседних волновых плоскостей с осями X и Y , – пространственные периоды волны, как видно из рис. 1.4, равны соответственно

$$d_x = \lambda / \cos \alpha = 1 / u, d_z = \lambda / \cos \varphi = 1 / w.$$

Углы φ и α представляют собой углы, образованные направлением распространения волны \vec{k} с плоскостями YZ и XZ соответственно.

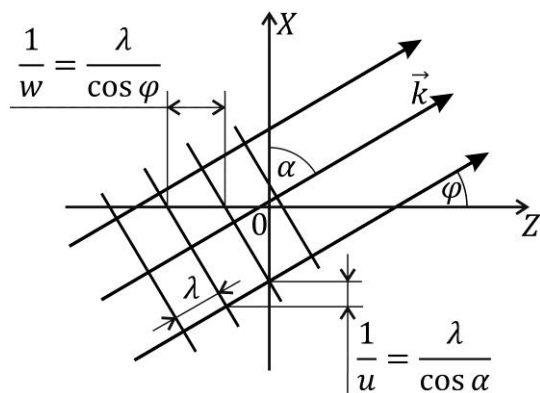


Рис. 1.4

Из рис. 1.4 видно, что, изменяя угол падения плоской волны, можно получить в плоскости $z = 0$ поля с различными пространственными частотами. Максимальные значения пространственных частот, получаемых при $\alpha = \beta = \gamma = 0$, равны $1 / \lambda$. Так как направляющие косинусы могут принимать как положительные, так и отрицательные значения (в зависимости от того меньше или больше $\pi / 2$ направляющие углы α, β, γ вектора \vec{k} волны), то и пространственные частоты могут быть как положительными, так и отрицательными. Этим случаям соответствуют различные физические состояния.

Легко видеть, что

$$u = \frac{1}{2\pi} \frac{\partial \varphi}{\partial x}, \quad v = \frac{1}{2\pi} \frac{\partial \varphi}{\partial y}, \quad w = \frac{1}{2\pi} \frac{\partial \varphi}{\partial z}, \quad (1.29)$$

где $\varphi = 2\pi(ux + vy + wz)$ – фаза плоской волны. Следовательно, пространственные частоты u , v и w определяют скорость изменения фазы волны в направлениях координатных осей X , Y и Z соответственно. Подобно тому, как частота ν определяет скорость изменения фазы $\varphi = 2\pi\nu t + \alpha$ во времени:

$\nu = \frac{1}{2\pi} \frac{\partial \varphi}{\partial t}$. В случае плоской волны, где v в зависимости от

переменных x , y и z фаза изменяется по линейному закону, пространственные частоты u , v и w не зависят от этих переменных. Дифференцируя по координатам x и y фазу сферической волны в параксиальном приближении

$\varphi = \pm k \frac{x^2 + y^2}{2a}$, в соответствии с формулой (1.28) получим

пространственные частоты u и v сферической волны:

$$u = \frac{1}{2\pi} \frac{\partial \varphi}{\partial x} = \pm \frac{x}{\lambda a}, \quad v = \frac{1}{2\pi} \frac{\partial \varphi}{\partial y} = \pm \frac{y}{\lambda a}.$$

Таким образом, у сферической волны в параксиальном приближении в зависимости от переменных x и y фаза изменяется по квадратичному закону, а пространственные частоты u и v – по линейному закону. Такую волну можно рассматривать как двумерное колебание с постоянной амплитудой и линейной частотой модуляции. Пространственная частота цилиндрической волны

$$u = \frac{1}{2\pi} \frac{\partial \varphi}{\partial x} = \pm \frac{x}{\lambda z}.$$

Видим, что в зависимости от координаты x пространственная частота цилиндрической волны изменяется по линейному закону.

Пространственные частоты u , v и w связаны с компонентами k_x, k_y и k_z волнового вектора \mathbf{k} соотношениями

$$k_x = 2\pi u, \quad k_y = 2\pi v, \quad k_z = 2\pi w.$$

Заметим, что эта связь такая же, как между круговой, или циклической, частотой ω и линейной частотой ν : $\omega = 2\pi\nu$.

Поэтому величины k_x, k_y и k_z иногда называют круговыми, или циклическими, пространственными частотами. Отсюда видно также, что пространственные частоты u , v и w являются компонентами вектора $\mathbf{q} = \mathbf{k} / 2\pi$.

Величины u , v и w не являются независимыми. Соотношение, связывающее эти величины, можно получить из условия, что сумма квадратов направляющих косинусов равна единице:

$$u^2 + v^2 + w^2 = 1/\lambda^2.$$

Это соотношение позволяет выразить одну пространственную частоту через две другие. Например,

$$w = \pm \sqrt{1/\lambda^2 - u^2 - v^2}.$$

Знак плюс или минус перед корнем в этом выражении определяется направлением распространения волны. Так как пространственная частота – величина действительная, то должно выполняться соотношение

$$u^2 + v^2 \leq 1/\lambda^2.$$

В предельном случае, когда $u^2 + v^2 = 1/\lambda^2$, частота $w = 0$. Это соответствует плоским волнам, которые распространяются перпендикулярно оси Z и не переносят энергию в направлении этой оси.

С учетом соотношения, выражающего пространственную частоту w через пространственные частоты u и v , комплексную амплитуду плоской волны можно записать в виде

$$E(x, y, z) = E_0 e^{i2\pi (ux + vy)} e^{i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}}. \quad (1.30)$$

В аргументе второй экспоненты можно было бы взять и знак минус. Выражение (1.30) будет использовано в последующих главах при рассмотрении спектрального метода решения дифракционных задач.

Заметим, что произведение первых двух сомножителей в правой части формулы (1.30) представляет собой комплексную амплитуду плоской волны в плоскости $z = 0$. Следовательно, величина комплексной амплитуды плоской волны на произвольном расстоянии z (в плоскости $z = \text{const}$) равна произведению комплексной амплитуды волны при $z = 0$ и экспоненты, возрастающей при увеличении z :

$$E(x, y, z) = E(x, y, 0) \exp(i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}).$$

Понятие пространственной частоты применимо к синусоидальной составляющей пространственного распределения любой физической величины, а не только к распределению комплексных амплитуд света. Пространственная частота синусоидальной составляющей в каком-либо направлении в 2π раз меньше скорости пространственного изменения фазы данной составляющей в этом направлении и равна обратной величине пространственного периода этой синусоидальной составляющей, измеренного в выбранном направлении. Разлагая пространственный период по ортогональным направлениям X , Y и Z , получим соответствующие пространственные частоты u , v и w .

1.7. Интенсивность светового излучения

Перенос энергии световой электромагнитной волной определяется вектором плотности потока энергии электромагнитной волны – вектором Умова – Пойнтинга, определяемым как $\mathbf{S} = w\mathbf{v}$, где $w = \varepsilon_0 \varepsilon E^2$ – объемная плотность энергии электромагнитной волны, \mathbf{v} – вектор скорости волны. Модуль вектора Умова – Пойнтинга с учетом формулы Максвелла $n = \sqrt{\varepsilon}$ определится как

$$S = wv = \varepsilon_0 \varepsilon E^2 \frac{c}{\sqrt{\varepsilon}} = \varepsilon_0 \sqrt{\varepsilon} c E^2 = \varepsilon_0 n c E^2.$$

Для вакуума (где $n = 1$) модуль вектора плотности потока энергии

$$S = \varepsilon_0 c E^2.$$

Физически модуль вектора плотности потока энергии представляет собой энергию, переносимую волной за единицу времени через единичную площадку, расположенную перпендикулярно направлению распространения волны, или величину мощности оптического излучения, проходящего через единичную площадку, перпендикулярную направлению распространения света.

Все приемники оптического излучения (фотопленка, фотодиоды, сетчатка глаза и др.) в отличие от приемников волн радиоизлучения реагируют не на мгновенные значения напряженности электрического поля световой электромагнитной волны, а на величину, пропорциональную квадрату напряженности поля. При этом все приемники света в той или иной степени обладают инерционностью, характеризуемой временем $t_{\text{пр}}$. Это время называется временем разрешения («срабатывания») приемника, или просто временем прибора. Для фотоматериалов, например, время разрешения равно времени экспозиции, $t_{\text{пр}} \sim 10^{-2} - 10^{-4}$ с; для сетчатки глаза $t_{\text{пр}} \sim 0,1$ с. В

течение этого времени глаз сохраняет зрительное восприятие, т.е. продолжает «видеть свет», хотя свет перестал в него попадать. Глаз не замечает также миганий света, если они следуют друг за другом через промежутки времени, малые по сравнению с $t_{\text{пр}}$ (по этой же причине глаз не замечает и мельканий кадров на экране кино или телевизора). Даже у самых малоинерционных приемников, основанных на фотоэффекте (фотоэлементы), $t_{\text{пр}} \sim 10^{-10}$ с, что на четыре-пять порядков больше периода изменения электрического поля световой волны. Поэтому любой приемник оптического излучения регистрирует не просто величину, пропорциональную квадрату напряженности поля (плотность потока энергии), а эту величину, но усредненную по времени срабатывания прибора. Такие величины называются величинами, квадратичными по полю. К ним относится важнейшая физическая величина – интенсивность света, определяемая как среднее по времени наблюдения значение модуля вектора плотности потока энергии:

$$I = \langle S(t) \rangle = \frac{1}{t_{\text{пр}}} \int_0^{t_{\text{пр}}} S(t) dt = \varepsilon_0 c \langle |E(t)|^2 \rangle \quad (1.31)$$

(для вакуума). Угловые скобки в этом выражении означают операцию усреднения по времени. Заметим, что усреднение здесь производится не по периоду T световой волны, как для электромагнитной волны радиодиапазона, а по времени разрешения приемника $t_{\text{пр}}$.

Кроме интенсивности, величинами, квадратичными по полю, являются также все энергетические и фотометрические величины: лучистый и световой поток, яркость, освещенность и др.

Для монохроматической волны

$$\langle E^2(t) \rangle = \frac{1}{2} E_{00}^2,$$

так что

$$I = \frac{1}{2} \varepsilon_0 c E_{00}^2.$$

Если волновая функция представляется в комплексном виде, то и величина $\langle E^2(t) \rangle$ будет комплексной, но интенсивность (как и энергия) должна быть величиной действительной. Поэтому в этом случае в определении интенсивности величину $\langle E^2(t) \rangle$ следует заменить на среднее значение квадрата модуля

$$\langle |E(t)|^2 \rangle = \langle E(t) \cdot E^*(t) \rangle.$$

Кроме того, перед выражением $\langle |E(t)|^2 \rangle$ следует поставить множитель $1/2$, так как $\langle E^2(t) \rangle = (1/2) E_{00}^2$, а $\langle |E(t)|^2 \rangle = E_{00}^2$. Учитывая это, для интенсивности света будем иметь:

$$I = \frac{1}{2} \varepsilon_0 c \langle |E(t)|^2 \rangle = \frac{1}{2} \varepsilon_0 c |E_0|^2.$$

(Для получения интенсивности света в среде с показателем преломления n в произведение $\varepsilon_0 c$ следует включить также множитель n .)

В дальнейшем под интенсивностью света будем понимать среднюю по времени наблюдения мощность излучения, переносимую волной через единичную площадку, и при записи постоянный множитель $(1/2)\varepsilon_0 c$ будем опускать, принимая, таким образом, за интенсивность света среднее по времени наблюдения значение квадрата модуля комплексной амплитуды поля, т.е. величину

$$I = \left\langle |E(t)|^2 \right\rangle = \frac{1}{t_{\text{пр}}} \int_0^{t_{\text{пр}}} |E(t)|^2 dt.$$

Строго говоря, это равенство имеет место при соответствующем выборе размера элементарной площадки, через которую проходит световая энергия (эта площадка должна быть равна не единице, а величине $(1/2) \varepsilon_0 c$).

Заметим, что фазовый множитель $\exp(-i\omega t)$ при определении интенсивности исчезает. Поэтому в расчетах с помощью волновой функции (1.22) указанный фазовый множитель обычно опускают и пользуются комплексной амплитудой (1.23), удовлетворяющей волновому уравнению Гельмгольца (1.24). Его опускают и по той причине, что на распределение поля в пространстве он не влияет. При определении интенсивности с помощью формулы (1.31) фазовый множитель $\exp[-i\varphi(x, y, z)]$ также не будет иметь значения, так как при определении квадрата модуля функции $E(x, y, z)$ он исчезает. Поэтому для монохроматической волны формулу интенсивности можно записать в виде $I = |E(x, y, z)|^2$.

В не поглощающей среде интенсивность плоской монохроматической волны не зависит от расстояния r , проходимого волной. В сферической волне интенсивность обратно пропорциональна квадрату этого расстояния ($I \sim 1 / r^2$), а в цилиндрической – обратно пропорциональна расстоянию R ($I \sim 1 / R$). Это связано с трехмерностью физического пространства. Если излучение не поглощается средой, то поток энергии через поверхность сферы любого радиуса r , окружающей источник излучения, одинаков, т.е. $\Phi_w = S\sigma = \text{const}$, где S – плотность потока энергии, σ – площадь сферы, окружающей источник. А поскольку плотность потока энергии $S \sim E_{00}^2$, а площадь сферы в трехмерном пространстве $\sigma = 4\pi r^2$, то из условия $\Phi_w = \text{const}$ следует, что $E_{00} \sim 1 / r$, а $I \sim 1 / r^2$. Аналогично для цилиндрических волн из условия сохранения потока энергии через боковую поверхность цилиндра $\sigma = 2\pi Rh$, где h – высота цилиндра,

$$\Phi_w = S \cdot 2\pi Rh \sim E_{00}^2 \cdot R = \text{const},$$

получаем $E_{00} \sim 1/r$, $aI \sim 1/R$.

Тем самым мы показали, что для сохранения потока энергии волны интенсивность сферической волны должна убывать обратно пропорционально квадрату расстояния, а интенсивность цилиндрической волны – обратно пропорционально расстоянию от источника волн.

Глава 2

ЧАСТОТНЫЙ СПЕКТР СИГНАЛОВ

2.1. Общие представления о частотном спектре

Как уже отмечалось, реальные световые волны никогда не бывают строго монохроматическими. Это обусловлено многими причинами и, прежде всего, самим механизмом излучения света. Свет, излучаемый реальными источниками, как будет показано в следующей главе, состоит из множества быстро сменяющих друг друга цугов (отрезков синусоид), амплитуды и начальные фазы которых изменяются беспорядочным образом; циклические частоты и длительности этих цугов также могут быть различными. Такой свет не является монохроматическим. Даже если выделить (например, с помощью светофильтра) цуги одинаковых частот, испускаемые одинаково возбужденными атомами, то и тогда из-за хаотичности изменения амплитуд и фаз цугов результирующее излучение не будет монохроматическим. Излучение не будет монохроматическим, как будет показано ниже, и вследствие ограниченности длительности и протяженности волновых цугов.

Уже отмечалось также, что любое сложное немонахроматическое излучение можно представить в виде совокупности (в общем случае бесконечной) монохроматических

волн. Такое представление называется спектральным разложением, или спектральным анализом (иногда это называют гармоническим анализом). Совокупность частот гармонических составляющих, входящих в состав сложного излучения, называют частотным спектром этого излучения. Спектр может быть как дискретным, состоящим из отдельных частот

$$\omega_1, \omega_2, \omega_3, \dots, \omega_m, \dots, \quad (2.1)$$

так и непрерывным, когда частоты непрерывно заполняют некоторый интервал, называемый спектральным интервалом. Спектральный интервал может быть и бесконечным. Графически спектр изображают как распределение интенсивности излучения по частотам (или длинам волн). На рис. 2.1 показаны примеры дискретного (а) и непрерывного (б) спектров.

С математической точки зрения замена некогерентного излучения совокупностью когерентных волн означает разложение функции, описывающей излучение, либо в ряд Фурье (если эта функция периодическая), либо в интеграл Фурье (если – непериодическая). Смысл спектрального разложения состоит в том, что ему соответствует реальный физический процесс: любой спектральный прибор производит фактическое разложение сложного излучения источника света на когерентные составляющие, причем каждая такая составляющая дает в спектральном приборе отдельную линию, называемую спектральной линией.

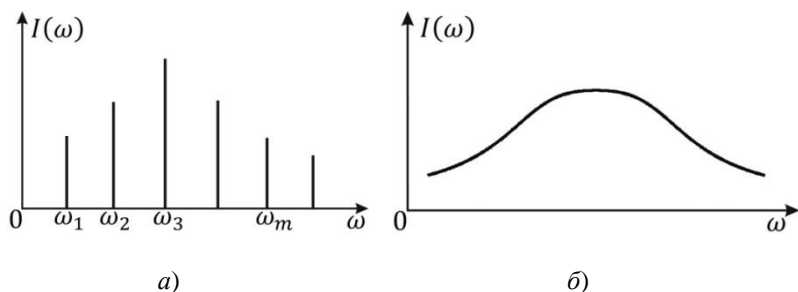


Рис. 2.1

Предположим, что напряженность электрического поля излучения имеет определенное направление и зависит только от одной пространственной переменной z и от времени t : $E = E(z, t)$. В соответствии с теоремой Фурье, ее можно представить в виде суперпозиции плоских монохроматических волн $\exp[i(kz - \omega t)]$. Положив в волновой функции $E(z, t)$ переменную $z = 0$, получим напряженность поля как функцию только времени t : $E(0, t) = E(t)$. Эта функция описывает колебание электрического поля в точке $z = 0$. В этом случае рассматривается спектральное разложение колебательного процесса, создаваемого волной в точке $z = 0$. Разложение будет производиться по гармоническим функциям $\exp(-i\omega t)$.

Если функция $E(t)$ периодическая и удовлетворяет некоторым условиям, называемым условиями Дирихле, то согласно теореме Фурье ее можно разложить в ряд Фурье

$$E(t) = \sum_{m=-\infty}^{\infty} E_m e^{-i\omega_m t}. \quad (2.2)$$

Здесь $\omega_m = m\omega_1$, $\omega_1 = 2\pi/T$, T – период функции $E(t)$; число m принимает целые положительные и отрицательные значения. Берется вещественная часть ряда (2.2). Члены этого ряда называются гармониками соответственно первой или основной (при $m = 1$), второй ($m = 2$), третьей ($m = 3$) и т.д. Частота ω_1 представляет собой частоту основной гармоники, поэтому ее называют основной частотой периодического процесса. Число гармоник бесконечно велико. Комплексные коэффициенты E_m ряда Фурье определяются интегралами

$$E_m = \frac{1}{T} \int_{-T/2}^{T/2} E(t) e^{i\omega_m t} dt. \quad (2.3)$$

Формально суммирование в (2.2) производится по всем частотам, как положительным (которые только и имеют физический смысл), так и отрицательным¹. При этом, поскольку функция $E(t)$ вещественная, то должно выполняться условие $E_{-m} = E_m^*$, где символ «*» обозначает комплексное сопряжение. Как комплексные величины коэффициенты E_m можно представить в виде

$$E_m = |E_m| e^{-i\alpha_m}. \quad (2.4)$$

Здесь $|E_m|$ – амплитуда, а α_m – начальная фаза m -ой гармоники. Две характеристики – амплитудная и фазовая, т.е. модули и аргументы комплексных коэффициентов ряда Фурье, полностью определяют структуру частотного спектра периодического колебания поля $E(t)$. Совокупность величин $|E_m|$ часто называют спектром амплитуд, а совокупность φ_m – спектром фаз. Спектр можно изобразить графически, откладывая по оси абсцисс частоты ω_m гармоник, а по оси ординат – их амплитуды $|E_m|$. Графическое изображение дает наглядное представление о «ширине» спектра и относительной величине отдельных ее гармонических составляющих (гармоник). Спектр можно изобразить и в виде зависимости интенсивности гармоник ($I_m \sim |E_m|^2$) от их частоты.

Спектр периодического процесса является дискретным, или линейчатым, так как состоит из отдельных линий – вертикальных отрезков, соответствующих дискретным частотам $\omega_1, 2\omega_1, 3\omega_1$ и т.д. Длины отрезков (в заданном масштабе) равны амплитудам

¹ Отрицательной частоте в геометрическом представлении гармонических колебаний формально можно поставить в соответствие вращение вектора E в направлении, противоположном соответствующем положительной частоте, т.е. по часовой стрелке.

соответствующих гармоник E_m или их интенсивностям I_m (рис. 2.1, а, если $\omega_2 = 2\omega_1$, $\omega_3 = 3\omega_1$ и т.д.).

Подставляя в (2.2) соотношение (2.4), будем иметь

$$E(t) = \sum_{m=-\infty}^{\infty} |E_m| e^{-i(\omega_m t + \alpha_m)}.$$

Используя формулу

$$2 \cos \varphi = \exp(i\varphi) + \exp(-i\varphi),$$

вытекающую из формулы Эйлера, можно от комплексной формы ряда Фурье перейти к его тригонометрической форме

$$E(t) = C_0 + \sum_{m=1}^{\infty} C_m \cos(m\omega t + \alpha_m), \quad (2.5)$$

где $C_m = 2|E_m|$, $C_0 = E_0$ – постоянная составляющая, представляющая собой среднее значение функции $E(t)$ за период. Запись ряда Фурье в таком виде удобна в тех случаях, когда нужно знать амплитуду и начальную фазу m -ой гармоники.

Для практических применений более удобной является следующая тригонометрическая форма ряда Фурье:

$$E(t) = \frac{A_0}{2} + \sum_{m=1}^{\infty} (A_m \cos m\omega_1 t + B_m \sin m\omega_1 t),$$

где A_0 , A_m , B_m – коэффициенты ряда Фурье, определяемые равенствами

$$A_0 = \frac{2}{T} \int_{-T/2}^{T/2} E(t) dt;$$

$$A_m = \frac{2}{T} \int_{-T/2}^{T/2} E(t) \cos m\omega_1 t dt; \quad B_m = \frac{2}{T} \int_{-T/2}^{T/2} E(t) \sin m\omega_1 t dt.$$

В такой форме сигнал $E(t)$ представляется в виде бесконечной суммы синусоидальных и косинусоидальных составляющих, причем A_m – амплитуда m -ой косинусоидальной составляющей сигнала, B_m – амплитуда m -ой синусоидальной составляющей. Амплитуды C_m и фазы φ_m гармоник выражаются через коэффициенты A_m и B_m следующим образом:

$$C_0 = A_0 / 2, \quad C_m = \sqrt{A_m^2 + B_m^2}, \quad \alpha_m = \arg C_m,$$

$$\alpha_m = \begin{cases} \arctg(B_m / A_m), & A_m \geq 0, \\ \pi + \arctg(B_m / A_m), & A_m < 0, B_m > 0, \\ -\pi + \arctg(B_m / A_m), & A_m < 0, B_m \leq 0. \end{cases}$$

Комплексные амплитуды $E_{\pm m}$ в свою очередь связаны с коэффициентами A_m и B_m следующими соотношениями:

$$E_{\pm m} = \frac{1}{2}(A_m \pm iB_m).$$

Комплексные амплитуды E_m и E_{-m} являются взаимно сопряженными комплексными величинами, так что равенство $E_{-m} = E_m^*$ выполняется.

Если функция $E(t)$ – четная, то функции $E(t) \sin m\omega_1 t$ будут нечетными, и тогда все коэффициенты $B_m = 0$; если же $E(t)$ – нечетная, то функции $E(t) \cos m\omega_1 t$ будут нечетными, и тогда все $A_m = 0$. В первом случае ряд Фурье будет состоять из одних только косинусов, а во втором – из одних только синусов.

Используя соотношения (2.2), (2.3) и учитывая свойство ортогональности гармонических функций с кратными частотами, для интенсивности сложного периодического излучения можно получить

$$I = \frac{1}{2} \langle |E(t)|^2 \rangle = \sum_{m=0}^{\infty} |E_m|^2.$$

Предполагается, что постоянная составляющая ряда равна нулю (в физических задачах это почти всегда имеет место). Отсюда следует, что интенсивность сложного периодического процесса равна сумме интенсивностей гармонических составляющих (гармоник), т.е. в энергетическом отношении отдельные спектральные составляющие сложного периодического процесса аддитивны. По виду огибающей величин $|E_m|^2$ можно судить о распределении интенсивности в спектре периодического сигнала.

Отметим важное свойство любого периодического процесса: его гармонические составляющие взаимно независимы. Поэтому при изменении или даже удалении из спектра любых гармоник амплитуды и фазы оставшихся гармоник не изменяются.

Непериодический сигнал является частным случаем периодического сигнала, у которого период бесконечно велик: $T \rightarrow \infty$. Для непериодического процесса разложение функции $E(t)$ по гармоническим составляющим представляется в виде интеграла Фурье

$$E(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} E(\omega) e^{-i\omega t} d\omega. \quad (2.6)$$

Это выражение представляет непериодическую функцию в виде суммы (интеграла) гармонических колебаний с бесконечно малыми амплитудами. Чтобы такое представление было возможно, функция $E(t)$ должна удовлетворять определенным требованиям (условиям Дирихле), в частности, функция $E(t)$ должна быть абсолютно интегрируема, т.е. должно выполняться неравенство

$$\int_{-\infty}^{\infty} |E(t)| dt < \infty.$$

В физических задачах условия Дирихле обычно выполняются. Согласно теореме Фурье, интеграл (2.6) можно обратить и найти

$$E(\omega) = \int_{-\infty}^{\infty} E(t) e^{i\omega t} dt. \quad (2.7)$$

Здесь $E(-\omega) = E^*(\omega)$.

Интегралы (2.7) и (2.6) называют соответственно преобразованием Фурье и обратным преобразованием Фурье. Пределы интегрирования в (2.6) и (2.7) определяются областями существования функций $E(t)$ и $E(\omega)$ соответственно. Преобразование Фурье будем обозначать символом $F\{\dots\}$, а обратное преобразование Фурье – символом $F^{-1}\{\dots\}$. В этих обозначениях

$$E(\omega) = F\{E(t)\}, \quad E(t) = F^{-1}\{E(\omega)\}.$$

Функцию $E(\omega)$ называют фурье-спектром, спектральной плотностью или фурье-образом функции $E(t)$. В свою очередь функцию $E(t)$ называют обратным (или инверсным) фурье-образом функции $E(\omega)$ или ее прообразом. Время t и круговую частоту ω называют сопряженными переменными преобразования Фурье.

При $\omega = 0$ выражение (2.7) принимает вид

$$E(0) = \int_{-\infty}^{\infty} E(t) dt.$$

Эта величина равна площади, ограниченной кривой $E(t)$. Аналогично из (2.6) при $t = 0$ получаем

$$E(0) = \int_{-\infty}^{\infty} E(\omega) d\omega.$$

Здесь правая часть представляет собой площадь, ограниченную кривой $E(\omega)$. Полученные результаты являются частными случаями преобразований Фурье.

Отметим здесь, что хотя функция $E(t)$ и ее фурье-образ $E(\omega)$ обозначаются одним и тем же символом E , это не означает, что функции $E(t)$ и $E(\omega)$ имеют одинаковый вид и получаются одна из другой простой заменой аргументов t на ω (или ω на t). Вид этих функций различен. При этом размерность величины $E(\omega)$ есть В·с / м, размерностью же величины $E(t)$ является В / м.

Как видно из (2.7), функцию $E(t)$ можно рассматривать как суперпозицию бесконечно большого числа гармонических составляющих всевозможных частот. Величина $E(\omega)d\omega/2\pi$ определяет вклад в $E(t)$ гармонических составляющих из интервала частот $(\omega, \omega + d\omega)$. Поэтому функция $E(\omega)$ представляет собой спектральную плотность функции $E(t)$. Как и в случае периодического поля $E(t)$, она определяет распределение комплексных амплитуд гармоник по частотам, поэтому в общем случае она является комплексной. Комплексную функцию $E(\omega)$ называют также комплексной амплитудой волны с частотой ω (хотя на самом деле это – амплитуда, приходящаяся на единичный спектральный интервал). Как комплексную функцию ее можно записать в виде

$$E(\omega) = |E(\omega)| \exp [-i\varphi(\omega)].$$

Модуль $|E(\omega)|$ этой функции определяет амплитудно-частотную характеристику (сокращенно АЧХ), а аргумент $\varphi(\omega)$ – фазочастотную характеристику (ФЧХ) спектра непериодического процесса. Как видим, в случае непериодического процесса его спектр амплитуд является непрерывным.

Если найти спектр поля $E(t)$ по формуле (2.7), то, подставляя его в (2.6), получим исходное поле $E(t)$. Таким образом, описания

полей временными функциями и спектрами равноправны, и поэтому можно использовать ту или иную форму, исходя из простоты математического анализа.

Полная энергия непериодического процесса пропорциональна интегралу $\int_{-\infty}^{\infty} E^2(t) dt$. Подставляя сюда выражение (2.6),

получим

$$\int_{-\infty}^{\infty} E^2(t) dt = \int_{-\infty}^{\infty} E(t) \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} E(\omega) e^{-i\omega t} d\omega \right) dt.$$

Изменив порядок интегрирования по ω и t и используя выражение (2.7), будем иметь

$$\begin{aligned} \int_{-\infty}^{\infty} E^2(t) dt &= \frac{1}{2\pi} \int_{-\infty}^{\infty} E(\omega) \left(\int_{-\infty}^{\infty} E(t) e^{-i\omega t} dt \right) d\omega = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} E(\omega) E(-\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} |E(\omega)|^2 d\omega, \end{aligned}$$

где учтено, что $E(-\omega) = E^*(\omega)$, или

$$\int_{-\infty}^{\infty} E^2(t) dt = \frac{1}{\pi} \int_0^{\infty} |E(\omega)|^2 d\omega. \quad (2.8)$$

Из этого важного соотношения, известного как равенство Парсеваля, следует, что энергия непериодического сигнала равна сумме энергий всех его спектральных составляющих. Таким образом, энергия сигнала может быть найдена интегрированием либо во временной области, либо в частотной. Сказанное имеет место и для интенсивности излучения, так как $I = |E|^2$.

Величина, пропорциональная $|E(\omega)|^2 d\omega$, представляет собой энергию (и интенсивность) волны в интервале частот от ω до $\omega + d\omega$. Множитель $|E(\omega)|^2$ в этом выражении имеет смысл энергии, приходящейся на единичный интервал частот, и может рассматриваться как спектральная плотность энергии. Эта величина характеризует энергетический спектр сигнала. Отметим, что энергия непериодического сигнала не зависит от фаз спектральных составляющих.

Равенство Парсевала позволяет определить практическую ширину спектра сигнала: $\Delta\omega = \omega_{\text{в}} - \omega_{\text{н}}$, где $\omega_{\text{в}}$ и $\omega_{\text{н}}$ – соответственно верхняя и нижняя граничные частоты спектра, при которых величина спектральной плоскости энергии ниже некоторого заданного значения.

Преобразования Фурье (обратное и прямое) часто бывает удобно выражать не через круговые частоты ω , а через линейные частоты ν . Выраженные через частоты ν , эти преобразования имеют вид

$$E(t) = \int_{-\infty}^{\infty} E(\nu) e^{-i2\pi\nu t} d\nu \quad (2.9)$$

(обратное преобразование Фурье),

$$E(\nu) = \int_{-\infty}^{\infty} E(t) e^{i2\pi\nu t} dt \quad (2.10)$$

(прямое преобразование Фурье). В таком виде прямое и обратное преобразования Фурье являются симметричными друг другу. Сопряженными переменными здесь являются время t и линейная частота ν .

Чтобы придать физическую конкретность полученным выше формулам, рассмотрим несколько примеров их применения.

2.2. Частотный спектр некоторых простых сигналов

2.2.1. Спектр биполярного П-образного сигнала

В качестве примера применения ряда Фурье рассмотрим частотный спектр биполярного П-образного сигнала (меандра), представленного на рис. 2.2. При показанном на рисунке выборе начала отсчета времени функция $E(t)$ в интервале $(-T/2, T/2)$ определяется следующим образом

$$E(t) = \begin{cases} E, & -T/4 \leq t < T/4; \\ -E, & -T/2 \leq t < -T/4, T/4 < t \leq T/2. \end{cases}$$

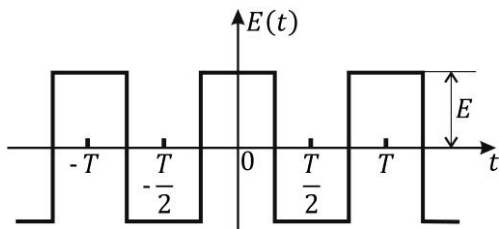


Рис. 2.2

При $t = 0$ эта функция имеет разрыв первого рода. Следовательно, условия Дирихле выполняются. Поскольку функция $E(t)$ является четной, то ряд Фурье будет содержать только косинусоиды с коэффициентами

$$A_0 = \frac{2}{T} \int_{-T/2}^{T/2} E(t) dt = 2 \cdot \frac{2}{T} \left(\int_0^{T/4} E dt + \int_{T/4}^{T/2} (-E) dt \right) = 0;$$

$$A_m = \frac{2}{T} \int_{-T/2}^{T/2} E(t) \cos m\omega_1 t dt = 2 \cdot \frac{2}{T} \left(E \int_0^{T/4} \cos m\omega_1 t dt + \right.$$

$$\left. + (-E) \int_{T/4}^{T/2} \cos m\omega_1 t dt \right) = \frac{4E}{m\pi} \sin \frac{m\pi}{2}.$$

Следовательно, $A_m = \frac{4E}{m\pi}$ при нечетном m и $A_m = 0$ при четном m . Как видим, при $B_m = 0$ коэффициенты A_m совпадают с амплитудами C_m соответствующих гармоник. Ряд Фурье в этом случае запишется в виде

$$E(t) = \frac{4E}{\pi} \left(\cos \omega_1 t + \frac{1}{3} \cos 3\omega_1 t + \frac{1}{5} \cos 5\omega_1 t + \dots \right) = \frac{4E}{\pi} \sum_{m=1}^{\infty} \frac{\cos(2m-1)\omega_1 t}{2m-1},$$

где ω_1 – основная частота, а $3\omega_1, 5\omega_1, 7\omega_1, \dots$ – частоты гармоник.

Амплитуды основной гармонической составляющей и гармоник более высокого порядка равны $\frac{4E}{\pi}, \frac{4E}{3\pi}, \frac{4E}{5\pi}, \dots, \frac{4E}{(2m-1)\pi}$. Как видим, амплитуды гармоник быстро убывают с ростом номера m .

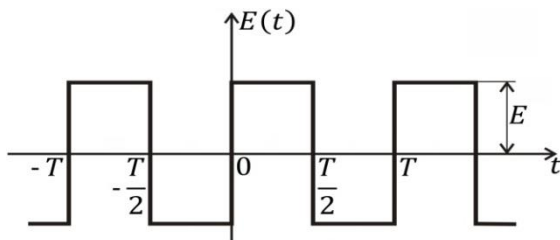


Рис. 2.3

Как и следовало ожидать, спектр рассматриваемого периодического процесса является линейчатый, состоящим из отдельных линий, соответствующих дискретным частотам $\omega_1, 3\omega_1, 5\omega_1, 7\omega_1$ и т.д.; длины линий равны амплитудам гармоник

$4E/\pi$, $4E/3\pi$, $4E/5\pi$ и т.д. Заметим, что амплитуды гармоник убывают пропорционально $1/(2m - 1)$. Спектр рассматриваемого сигнала показан на рис. 2.4.

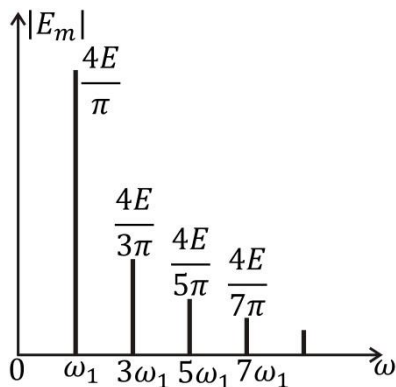


Рис. 2.4

Если начало отсчета времени сместить вправо (в положительном направлении по оси времени t) на величину $T/4$, как показано на рис.2.3, то функция $E(t)$ в интервале $(-T/2, T/2)$ определится в виде

$$E(t) = \begin{cases} E, & 0 < t \leq T/2; \\ -E, & -T/2 \leq t \leq 0. \end{cases}$$

Эта функция становится нечетной, и ряд Фурье в этом случае будет содержать только синусоиды, причем, как и выше, коэффициент $A_0 = 0$, а коэффициенты

$$\begin{aligned} B_m &= \frac{2}{T} \int_{-T/2}^{T/2} E(t) \sin m\omega_1 t dt = 2 \cdot \frac{2}{T} \int_0^{T/2} E \sin m\omega_1 t dt = \\ &= \frac{2E}{m\pi} (1 - \cos m\pi). \end{aligned}$$

Так как $\cos m\pi = (-1)^m$, то все $B_m = 0$ при четном m и равны $\frac{4E}{m\pi}$ при нечетном m . Ряд Фурье будет иметь вид

$$E(t) = \frac{4E}{\pi} \left(\sin \omega_1 t + \frac{1}{3} \sin 3\omega_1 t + \frac{1}{5} \sin 5\omega_1 t + \dots \right) = \\ = \frac{4E}{\pi} \sum_{m=1}^{\infty} \frac{\sin(2m-1)\omega t}{2m-1}.$$

Видим, что в обоих случаях частоты гармоник и соответствующие им амплитуды одинаковы, спектр не зависит от положения начала отсчета времени. Если в разложении ограничиться только основной и двумя следующим гармониками, то точность такого представления функции $E(t)$ составит примерно 91%.

2.2.2. Спектр незатухающего волнового цуга

В качестве первого примера определения спектра непериодического излучения рассмотрим спектральный состав волнового цуга – отрезка синусоиды длительностью τ и с амплитудой E_0 . Если начало отсчета времени поместить в середине цуга, то функцию $E(t)$, описывающую цуг, в комплексной форме можно представить в виде

$$E(t) = \begin{cases} E_0 \exp(-i\omega_0 t), & |t| \leq \tau/2; \\ 0, & |t| > \tau/2, \end{cases}$$

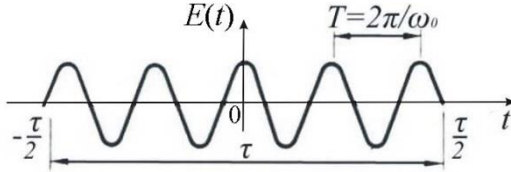


Рис. 2.5

где ω_0 – круговая частота волнового цуга. Вид этого цуга показан на рис. 2.5. Поскольку функция $E(t)$ непериодическая, то для определения спектрального состава цуга следует произвести преобразование Фурье функции $E(t)$, описывающей цуг. Подставляя функцию $E(t)$ цуга в формулу (2.7) и учитывая, что эта функция определена только внутри интервала от $-\tau/2$ до $\tau/2$, получим

$$\begin{aligned}
 E(\omega) &= E_0 \int_{-\tau/2}^{\tau/2} \exp[i(\omega - \omega_0)t] dt = \\
 &= 2E_0 \frac{\sin[(\omega - \omega_0)\tau/2]}{\omega - \omega_0}.
 \end{aligned}$$

Представим это выражение в виде

$$E(\omega) = E_0 \tau \frac{\sin[\pi(\omega - \omega_0)\tau/2\pi]}{\pi(\omega - \omega_0)\tau/2\pi}. \quad (2.11)$$

Выражение, стоящее за множителем $E_0\tau$ в (2.11), можно записать в виде так называемой sinc-функции

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

аргумента $(\omega - \omega_0)\tau/2\pi$. График функции $\text{sinc}(x)$ показан на рис. 2.6.

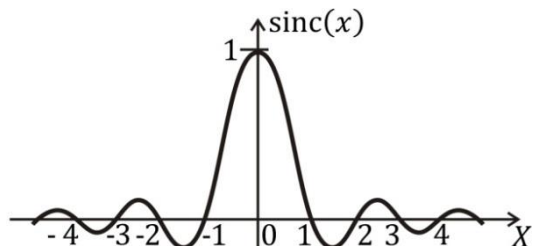


Рис. 2.6

Как видим, sinc-функция имеет главный максимум, равный единице, при $x = 0$. В точках $x = \pm 1, \pm 2, \pm 3, \dots$ функция $\text{sinc}(x)$ обращается в нуль. В промежутках между этими точками она имеет второстепенные максимумы (и минимумы), много меньшие по величине главного (центрального) максимума и быстро убывающие с увеличением по модулю аргумента. Положения этих максимумов определяются из решения трансцендентного уравнения

$$\pi x \cos \pi x - \sin \pi x = 0.$$

Введя sinc-функцию, выражение для $E(\omega)$ можно записать в виде

$$E(\omega) = E_0 \tau \text{sinc}[(\omega - \omega_0)\tau / 2\pi]. \quad (2.12)$$

Функция (2.12) и определяет распределение комплексных

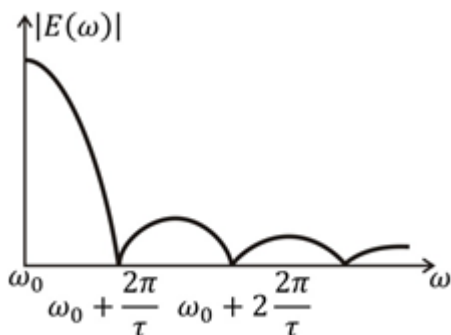


Рис. 2.7

амплитуд
гармонических

составляющих цуга, т.е. амплитудный спектр незатухающего волнового цуга.

В рассматриваемом случае функция $E(\omega)$ – действительная величина, принимающая положительные и отрицательные значения. Она имеет главный максимум при $\omega = \omega_0$ и симметрична относительно начала координат, помещенного в точку $\omega = \omega_0$. При $(\omega - \omega_0)\tau / 2\pi = \pm m$, где $m = 1, 2, 3, \dots$ – целое число, т.е. при

$$\omega = \omega_0 \pm m(2\pi/\tau),$$

спектр амплитуд проходит через нуль. При переходе через нулевые точки синус в выражении (2.12) меняет знак и, следовательно, фаза соответствующих гармоник изменяется на π .

Нули в спектре обусловлены взаимным гашением гармонических составляющих цуга, а горбы и впадины – их взаимным усилением. Наибольшую амплитуду $E_0\tau$ имеет гармоническая составляющая с частотой, равной частоте волнового цуга ω_0 . При удлинении цуга (увеличении τ) расстояние между нулевыми значениями функции $E(\omega)$ сокращается, а ее максимальное значение $E(\omega_0) = E_0\tau$ увеличивается. При укорочении цуга, наоборот, расстояние между нулевыми значениями $E(\omega)$ увеличивается, а максимальное значение $E(\omega_0)$ уменьшается. Амплитудно-частотная характеристика (модуль функции $E(\omega)$) спектра синусоидального цуга показана на рис. 2.7.

Распределение энергии излучения по спектру частот определяется спектральной плотностью энергии:

$$|E(\omega)|^2 = |E_0|^2 \tau^2 \text{sinc}^2[(\omega - \omega_0)\tau / 2\pi].$$

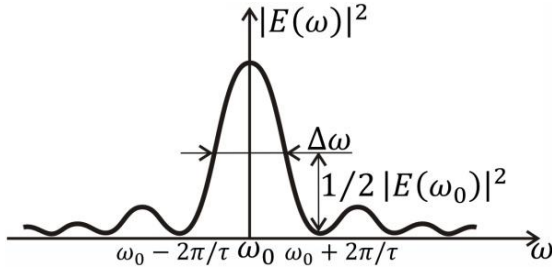


Рис. 2.8

График распределения плотности энергии по спектру частот показан на рис. 2.8. Видно, что спектр синусоидального волнового цуга является непрерывным, т.е. состоящим из бесконечного числа гармонических составляющих, частоты которых непрерывно заполняют бесконечный интервал. При $\omega = \omega_0$ кривая $|E(\omega)|^2$ имеет максимум, равный $|E_0|^2 \tau^2$. С увеличением разности $|\omega - \omega_0|$ она спадает в обе стороны, обращаясь в нуль при частотах $\omega_{1,2} = \omega_0 \pm 2\pi/\tau$. Затем кривая $|E(\omega)|^2$ вновь возрастает, оставаясь, однако, существенно меньше своего значения при $\omega = \omega_0$, затем убывает, вновь возрастает и т.д. При этом второй максимум функции $|E(\omega)|^2$ составляет $2,5 \cdot 10^{-3}$, третий $4 \cdot 10^{-4}$ и т.д. значения $|E(\omega_0)|^2$. Поэтому основное значение имеет только главный максимум, заполняющий интервал частот $[\omega_0 - 2\pi/\tau, \omega_0 + 2\pi/\tau]$. В пределах этого интервала сосредоточена практически вся энергия излучения.

Поэтому можно считать, что спектр ограниченного во времени излучения представляет собой полосу сплошного (непрерывного) спектра с центром при $\omega = \omega_0$. За ширину $\Delta\omega$ полосы принимают такой интервал частот, симметричный относительно

ω_0 , на границах которого значение $|E(\omega)|^2$ уменьшается вдвое по сравнению с его значением в максимуме $|E(\omega_0)|^2$.

Приблизненно ширина спектра $\Delta\omega$ равна расстоянию от центрального максимума до первого минимума, т.е. величине $(\omega_1 - \omega_0) = 2\pi/\tau$. Следовательно,

$$\Delta\omega \approx 2\pi/\tau. \quad (2.13)$$

Это важное соотношение между шириной спектра $\Delta\omega$ цуга и его длительностью τ имеет общий характер. В физике колебаний и волн его называют соотношением неопределенностей частота – время. Оно показывает, что конечность продолжительности излучения приводит к его некогерентности.

Если в формуле, описывающей цуг, положить $\omega_0 = 0$, получим прямоугольный импульс высотой E_0 и длительностью τ . Поместив начало отсчета времени посередине импульса, его можно представить функцией

$$E(t) = \begin{cases} E_0, & |t| \leq \tau/2; \\ 0, & |t| > \tau/2. \end{cases}$$

Спектр такого импульса проще всего получить, положив в формулах для $E(\omega)$ и $|E(\omega)|^2$ цуга значение ω_0 равным нулю. Будем иметь для спектра амплитуд прямоугольного импульса и его спектральной плотности энергии соответственно:

$$E(\omega) = E_0 \tau \text{sinc}(\omega\tau/2\pi),$$

$$|E(\omega)|^2 = |E_0|^2 \tau^2 \text{sinc}^2(\omega\tau/2\pi).$$

Как видим, фурье-спектр прямоугольного импульса имеет такую же структуру, как и фурье-спектр волнового цуга (если каждую из составляющих спектра представить в виде суммы положительных

и отрицательных частот), но с центром при $\omega = 0$. Видим также, что $E(0) = E_0 \tau$. Следовательно, при нулевой частоте спектральная плотность прямоугольного импульса, как и должно быть, равна площади импульса.

Следует отметить, что по свойству взаимозаменяемости t и ω в преобразованиях Фурье для четных функций фурье-спектр sinc -функции будет представлять собой прямоугольный импульс.

2.2.3. Спектр периодической последовательности цугов

Рассмотрим теперь периодическую последовательность идентичных импульсов в виде волновых цугов длительности τ , следующих друг за другом через промежутки времени Δt (рис. 2.9). Так как рассматриваемый процесс – периодический, то для определения спектра его следует представить рядом Фурье.

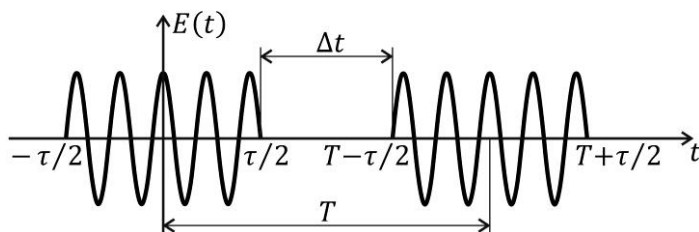


Рис. 2.9

Воспользуемся комплексной формой ряда. Начало отсчета времени поместим в центре одного из цугов. При выбранном начале отсчета функция $E(t)$, описывающая последовательность импульсов является четной, период повторения $T = \tau + \Delta t$. С учетом этого по формуле (2.3) находим коэффициенты разложения E_m :

$$E_m = \frac{1}{T} \int_{-T/2}^{T/2} E(t) e^{i\omega_m t} dt = E_0 \frac{1}{T} \int_{-\tau/2}^{\tau/2} \exp[i(\omega_m - \omega_0)t] dt =$$

$$= E_0 \frac{\tau}{T} \cdot \frac{\sin[(\omega_m - \omega_0)\tau/2]}{(\omega_m - \omega_0)\tau/2},$$

или

$$E_m = E_0 \frac{\tau}{T} \operatorname{sinc}[(\omega_m - \omega_0)\tau/2\pi], \quad (2.14)$$

где $\omega_m = m\omega_1$, $m = 1, 2, 3, \dots$, $\omega_1 = 2\pi/T$ – частота основной гармоники. Постоянная составляющая может быть найдена из соотношения

$$\begin{aligned} C_0 &= \frac{1}{T} \int_{-T/2}^{T/2} E(t) dt = E_0 \frac{1}{T} \int_{-\tau/2}^{\tau/2} \exp(-i\omega_0 t) dt = \\ &= E_0 \frac{\tau}{T} \operatorname{sinc}(\omega_0\tau/2\pi). \end{aligned}$$

Модуль выражения (2.14) и определяет амплитудный спектр рассматриваемой последовательности цугов. Спектр амплитуд – дискретный, его вид показан на рис. 2.10. Расстояние между соседними спектральными линиями равно основной частоте ω_1 :

$$\Delta\omega = \Delta m \cdot \omega_1 = \omega_1 = 2\pi/T.$$

При $(\omega_m - \omega_0)\tau/2\pi = \pm m$, т.е. при $\omega_m - \omega_0 = \pm m(2\pi/\tau)$,

где $m = 1, 2, 3, \dots$, все амплитуды E_m обращаются в нуль. Огибающая спектра (величин $|E_m|$) имеет вид амплитудно-частотной характеристики одиночного волнового цуга, определяемой модулем выражения (2.12). На рис. 2.10 эта огибающая изображена пунктирной кривой. При $T \rightarrow \infty$ (случай непериодического излучения) расстояние между соседними спектральными линиями

$\Delta\omega = 2\pi/T$ становится бесконечно малым ($\Delta\omega \rightarrow 0$), а спектр – сплошным. Частота ω_m при этом заменяется непрерывно изменяющейся текущей частотой ω . Спектр такого сигнала превращается в спектр одиночного волнового цуга, изображенного в виде огибающей на рис. 2.10.

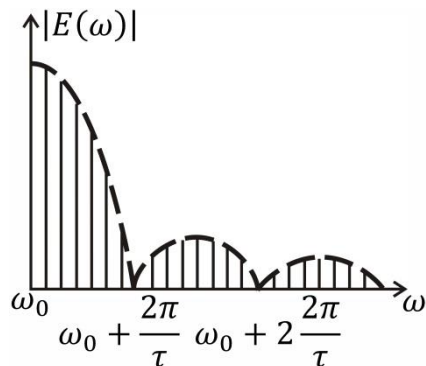


Рис. 2.10

2.2.4. Спектр монохроматического излучения

Из соотношения неопределенностей частота – время (2.13) следует, что если бы излучение продолжалось неограниченно долго ($\tau = \infty$), то его спектр представлял бы собой бесконечно узкую линию ($\Delta\omega = 0$), соответствующую частоте ω_0 . Рассмотрим этот вопрос подробнее. Перепишем выражение (2.11) в виде

$$E(\omega) = E_0 \frac{\sin[\pi(\omega - \omega_0)\tau/2\pi]}{\pi(\omega - \omega_0)/2\pi}.$$

Перейдя к пределу при $\tau \rightarrow \infty$ и учитывая модель дельта-функции Дирака вида

$$\delta(x) = \lim_{\varepsilon \rightarrow \infty} \frac{\sin \varepsilon \pi x}{\pi x},$$

будем иметь

$$\lim_{\tau \rightarrow \infty} E(\omega) = E_0 \lim_{\tau \rightarrow \infty} \frac{\sin[\pi(\omega - \omega_0)\tau/2\pi]}{\pi(\omega - \omega_0)/2\pi} = E_0 \delta\left(\frac{\omega - \omega_0}{2\pi}\right),$$

или, учитывая свойство δ -функции

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

$$\lim_{\tau \rightarrow \infty} E(\omega) = 2\pi E_0 \delta(\omega - \omega_0).$$

Отсюда следует, что предел $E(\omega)$ при $\tau \rightarrow \infty$ равен нулю при всех значениях ω за исключением значения $\omega = \omega_0$, при котором

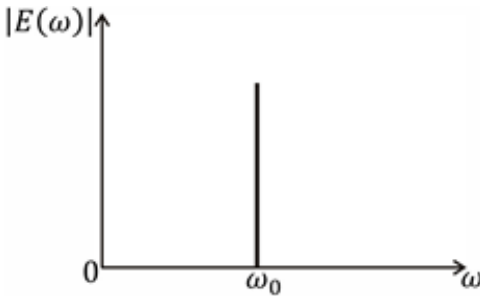


Рис. 2.11

$E(\omega)$ обращается в бесконечность. Это и означает, что амплитудный спектр излучения бесконечной длительности, т.е. монохроматического излучения

$$E(t) = E_{00} \cos \omega_0 t,$$

представляет собой

одну спектральную линию, соответствующую частоте $\omega = \omega_0$. Этот спектр представлен на рис. 2.11.

Проведенное рассмотрение показывает, что преобразованием Фурье гармонической функции $\cos \omega_0 t$ является δ -функция, умноженная на 2π : $F\{\cos \omega_0 t\} = 2\pi \delta(\omega - \omega_0)$. При этом легко видеть, что

$$\begin{aligned}
2\pi \delta(\omega - \omega_0) &= \int_{-\infty}^{\infty} \exp[i(\omega - \omega_0)t] dt = \\
&= \lim_{\tau \rightarrow \infty} \int_{-\tau/2}^{\tau/2} \exp[i(\omega - \omega_0)t] dt. \quad (2.15)
\end{aligned}$$

Математический объект, называемый δ (дельта)-функцией Дирака, часто используется в оптике, поэтому с этой функцией следует познакомиться поближе. Дельта-функция Дирака не является обычной функцией (такие функции в математике называют обобщенными функциями). Дельта-функцию можно получить посредством различных аналитических функций, совершая какой-либо предельный переход. Ее можно представить как предельную форму функции, имеющей вид прямоугольника шириной a и высотой $h = 1/a$ при $a \rightarrow 0$. В этом случае площадь прямоугольника остается равной

$$a \cdot h = 1,$$

ширина a стремится к нулю, а высота h — к бесконечности. Аналитически эту прямоугольную функцию можно записать в виде

$$f(x, a) = \begin{cases} 1/a, & |x| \leq a; \\ 0, & |x| > a, \end{cases}$$

$$\delta(x) = \lim_{a \rightarrow 0} f(x, a) = \lim_{a \rightarrow 0} \frac{1}{a}.$$

Таким же образом получается дельта-функция во временной об-

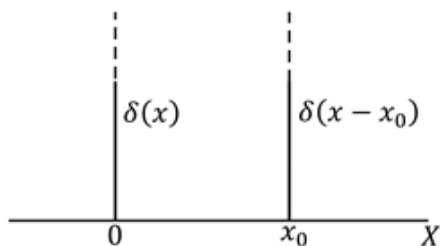


Рис. 2.12

ласти, если высоту прямоугольного импульса принять равной обратной величине его длительности τ при $\tau \rightarrow \infty$.

На рис. 2.12 изображены дельта-функции, расположенные в точке x_0 (функция $\delta(x - x_0)$) и в

начале координат (функция $\delta(x)$). Как видим, δ -функция равна нулю всюду, за исключением одной характерной точки ($x = x_0$ для $\delta(x - x_0)$ и $x = 0$ для $\delta(x)$), где она равна бесконечности. Можно сказать, что дельта-функция представляет собой импульс с нулевой (временной или пространственной) длительностью (протяженностью) и бесконечной амплитудой. Часто такой импульс называют единичным. Вид дельта-функции не зависит от ее расположения на оси X . Значение δ -функции определяется следующими соотношениями

$$\delta(x - x_0) = \begin{cases} 0, & x \neq x_0; \\ \infty, & x = x_0, \end{cases}$$

$$\int_{-\infty}^{\infty} \delta(x - x_0) dx = 1,$$

$$\int_{-\infty}^{\infty} f(x) \delta(x - x_0) dx = f(x_0). \quad (2.16)$$

$$\int_{-\infty}^x \delta(x) dx = \begin{cases} 0, & x < 0; \\ 1/2, & x = 0; \\ 1, & x > 0. \end{cases}$$

Соотношение (2.16) называется фильтрующим свойством дельта-функции. Этим соотношением $\delta(x - x_0)$ определяет отдельное значение $f(x_0)$ произвольной функции $f(x)$.

Дельта-функция является четной: $\delta(x) = \delta(-x)$. Это означает, что

$$\int_{-\infty}^{\infty} f(x) \delta(x) dx = \int_{-\infty}^{\infty} f(x) \delta(-x) dx = f(0) .$$

Во всех этих соотношениях вместо переменной x может стоять время t или частоты ω и k в зависимости от того, в какой области рассматривается функция – в координатной, временной или частотной.

Используя фильтрующее свойство δ -функции, можно найти ее фурье-спектр:

$$F\{\delta(t - t_0)\} = \int_{-\infty}^{\infty} \delta(t - t_0) e^{i\omega t} dt = \exp(i\omega t_0).$$

Положив в этом выражении $t_0 = 0$, получим фурье-спектр δ -функции:

$$F\{\delta(t)\} = 1.$$

В обоих случаях модуль равен единице. Это означает, что амплитудно-частотная характеристика дельта-функции постоянна и равна единице. Ее фазочастотная характеристика в случае $t_0 \neq 0$ имеет вид

$$\varphi(\omega) = \omega t_0,$$

т.е. линейно возрастает с частотой ω , а в случае $t_0 = 0$ – равна нулю при всех значениях частоты.

Заметим, что поскольку $\delta(\omega) = 1$, то, подставив это в формулу обратного преобразования Фурье (2.6), можно сразу прийти к соотношению (2.15).

Дельта-функция, называемая также импульсной функцией, служит для представления мгновенных импульсов и точечных воздействий – точечных источников света. Точечный источник света, расположенный в точке (x_0, y_0) , можно представить двумерной дельта-функцией

$$\delta(x - x_0, y - y_0) = \delta(x - x_0) \cdot \delta(y - y_0).$$

Для представления идеализированного точечного источника света часто используется и такая двумерная модель дельта-функции:

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

2.2.5. Спектр модулированного колебания

Для передачи информации на расстояние применяются электромагнитные сигналы, обладающие способностью распространяться в виде свободных электромагнитных волн в среде, разделяющей отправителя и получателя информации. Такими сигналами являются модулированные колебания, получающиеся из гармонического колебания

$$E(t) = E_0 \cos(\omega_0 t + \alpha),$$

если какой-либо из его постоянных параметров (амплитуду E_0 , начальную фазу α или частоту ω_0) заставить изменяться во времени под действием управляющего сигнала, в котором заключено передаваемое сообщение. Используемое для этих целей гармоническое колебание $E(t)$ называется несущим колебанием, а операция изменения его параметров под воздействием управляющего сигнала – модуляцией. Соответственно изменяющемуся параметру несущего колебания различают амплитудную, фазовую и частотную модуляции. При этом получают соответственно амплитудно-модулированное, фазомодулированное и частотно-модулированное колебания.

Амплитудная модуляция. При амплитудной модуляции под воздействием управляющего сигнала изменяется только амплитуда несущего колебания (а фаза и частота остаются неизменными). Амплитудно-модулированное колебание можно записать в виде

$$E(t) = A(t) \cos(\omega_0 t + \alpha), \quad (2.17)$$

где $A(t)$ – медленно меняющаяся (по сравнению с изменением $\cos(\omega_0 t + \alpha)$) функция (медленно меняющаяся амплитуда). Амплитудно-модулированные колебания широко используются в радиовещании, радиосвязи и при передаче изображения в телевидении. Это наиболее простой способ заложения информации в электромагнитное излучение. В простейшем случае

$$A(t) = E_0 + k_1 E_c(t),$$

где модулирующий (управляющий) сигнал $E_c(t)$ является однотональным сигналом, т.е. представляющим собой гармоническое колебание с амплитудой E_{01} и частотой ω_1 : $E_c(t) = E_{01} \cos \omega_1 t$; k_1 – коэффициент пропорциональности. Тогда

$$A(t) = E_0 (1 + m \cos \omega_1 t).$$

Величину ω_0 называют несущей частотой, ω_1 – частотой модуляции (при этом $\omega_1 \ll \omega_0$). Отношение

$$m = k_1 E_{01} / E_0 = \Delta A / E_0$$

называют коэффициентом глубины амплитудной модуляции (или просто глубиной амплитудной модуляции); он показывает, в каких пределах может изменяться амплитуда модулированного колебания под воздействием управляющего сигнала. Здесь $\Delta A = k_1 E_{01}$; постоянная E_0 представляет собой среднее по времени

значение амплитуды: $E_0 = \langle A(t) \rangle$. Заметим, что максимальное и минимальное значения амплитуды соответственно равны

$$A_{\max} = E_0 (1 + m), A_{\min} = E_0 (1 - m),$$

а коэффициент модуляции

$$m = \frac{A_{\max} - A_{\min}}{A_{\max} + A_{\min}}.$$

Для неискаженной передачи информации необходимо, чтобы $A_{\min} \geq 0$ и, следовательно, $m \leq 1$; значение $m = 1$ получается при $A_{\min} = 0$. Отсутствию модуляции соответствует значение $m = 0$, которое получается при $A_{\min} = A_{\max}$.

Подставив в (2.17) выражение амплитуды $A(t)$, получим

$$E(t) = E_0 (1 + m \cos \omega_1 t) \cos(\omega_0 t + \alpha). \quad (2.18)$$

Такое амплитудно-модулированное колебание показано на рис. 2.13. Огибающая $E(t)$ изменяется по закону модулирующего сигнала $A(t)$, $\Delta A = (A_{\max} - A_{\min})/2$ – амплитуда изменения огибающей.

Сложное негармоническое колебание (2.18) может быть представлено в виде совокупности трех строго гармонических колебаний с частотами ω_0 , $\omega_0 - \omega_1$ и $\omega_0 + \omega_1$. Действительно, раскрыв скобки и воспользовавшись известной тригонометрической формулой произведения косинусов, получим

$$E(t) = E_0 \cos(\omega_0 t + \alpha) + \frac{1}{2} E_0 m \cos [(\omega_0 - \omega_1)t + \alpha] + \frac{1}{2} E_0 m \cos [(\omega_0 + \omega_1)t + \alpha]. \quad (2.19)$$

Таким образом, амплитудно-модулированное колебание с однотоновальным управляющим сигналом представляется суммой трех гармонических колебаний: с основной (несущей) частотой

ω_0 и двумя боковыми частотами – нижней $\omega_0 - \omega_1$ и верхней $\omega_0 + \omega_1$ и амплитудами E_0 , $mE_0/2$ и $mE_0/2$ соответственно.

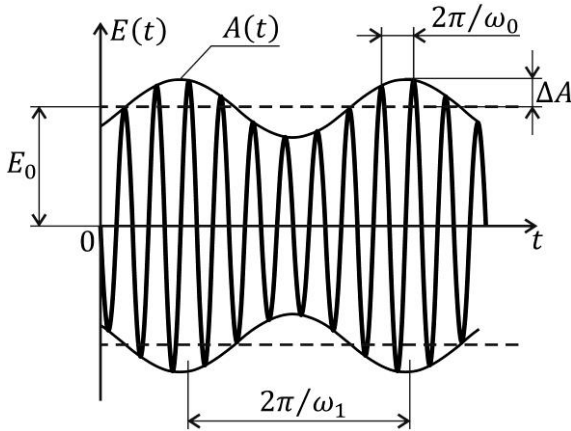


Рис. 2.13

Каждому слагаемому суммы (2.19) соответствует одна составляющая спектра амплитудно-модулированного сигнала. Заметим, что наибольшую амплитуду имеет колебание на несущей частоте ω_0 . Эта гармоническая составляющая вносит основной вклад в результирующее колебание. Амплитуды боковых колебаний зависят не только от среднего значения амплитуды E_0 , но и от коэффициента глубины модуляции m . Энергия колебания, усредненная по промежутку времени, равному периоду модуляции $T_1 = 2\pi/\omega_1$, равна (с точностью до несущественного коэффициента пропорциональности)

$$E_0^2 \langle (1 + m \cos \omega_1 t)^2 \rangle = E_0^2 + m^2 E_0^2 / 2$$

и, как и следовало ожидать, равна сумме энергий трех гармонических составляющих с частотами ω_0 , $\omega_0 - \omega_1$ и $\omega_0 + \omega_1$:

$$E_0^2 + \frac{1}{4}m^2E_0^2 + \frac{1}{4}m^2E_0^2.$$

Заметим также, что средняя за период мощность превышает мощность несущего колебания в $(1 + m^2/2)$ раз.

Применив к (2.19) преобразование Фурье и учтя соотношение (2.15), найдем фурье-спектр рассматриваемого амплитудно-модулированного колебания:

$$E(\omega) = \int_{-\infty}^{\infty} E(t) e^{i\omega t} dt = 2\pi E_0 \delta(\omega - \omega_0) + \\ + \pi E_0 m \delta[\omega - (\omega_0 - \omega_1)] + \pi E_0 m \delta[\omega - (\omega_0 + \omega_1)].$$

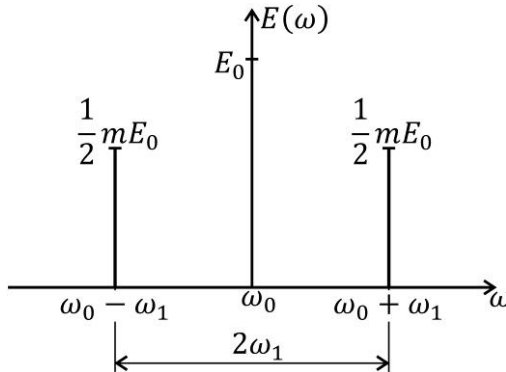


Рис. 2.14

Этот спектр представлен на рис. 2.14. Как видим, амплитудный спектр рассматриваемого простейшего амплитудно-модулированного колебания является линейчатым и состоит из трех линий, соответствующих частотам ω_0 , $\omega_0 - \omega_1$ и $\omega_0 + \omega_1$. Спектр симметричен относительно несущей частоты ω_0 ; ширина спектра

$$\Delta\omega = (\omega_0 + \omega_1) - (\omega_0 - \omega_1) = 2\omega_1$$

– равна удвоенной частоте модуляции.

Если управляющий сигнал содержит N гармонических составляющих, т.е. если

$$E_c(t) = \sum_{n=1}^N E_{0n} \cos(\omega_n t + \alpha_n),$$

$$A(t) = E_0 \left[1 + \sum_{n=1}^N m_n \cos(\omega_n t + \alpha_n) \right],$$

то амплитудно-модулированный сигнал может быть представлен в виде

$$E(t) = E_0 \cos(\omega_0 t + \alpha) + \frac{1}{2} E_0 \left\{ \sum_{n=1}^N m_n \cos[(\omega_0 + \omega_n)t + \alpha + \alpha_n] + \sum_{n=1}^N m_n \cos[(\omega_0 - \omega_n)t + \alpha - \alpha_n] \right\}.$$

Спектр в этом случае будет содержать несущую составляющую с частотой ω_0 и боковые гармоники с частотами $\omega = \omega_0 \pm \omega_n$, $n = 1, 2, 3, \dots, N$. Ширина спектра будет равна удвоенному значению наибольшей частоты гармоники:

$$\Delta\omega = (\omega_0 + \omega_{\max}) - (\omega_0 - \omega_{\max}) = 2\omega_{\max}.$$

Фазовая модуляция. При фазовой модуляции амплитуда и частота несущего колебания остаются постоянными, но изменяется со временем (в соответствии с законом изменения управляющего сигнала) его начальная фаза. Поле E при этом записывается как

$$E(t) = E_0 \cos[\omega_0 t + \alpha(t)].$$

В простейшем случае

$$\alpha(t) = \alpha + k_2 E_{01} \cos \omega_1 t = \alpha + m \cos \omega_1 t.$$

Здесь k_2 – коэффициент пропорциональности, $m = k_2 E_{01}$ – постоянная, определяющая диапазон изменения начальной фазы (от $-m$ до m) и называемая глубиной (или индексом) модуляции фазы. В этом случае

$$\begin{aligned} E(t) &= E_0 \cos[\omega_0 t + m \cos \omega_1 t + \alpha] = \\ &= E_0 \cos[(\omega_0 t + \alpha) + m \cos \omega_1 t]. \end{aligned}$$

Используя формулу косинуса суммы двух углов, это выражение можно записать в виде

$$\begin{aligned} E(t) &= E_0 [\cos(\omega_0 t + \alpha) \cos(m \cos \omega_1 t) - \\ &\quad - \sin(\omega_0 t + \alpha) \sin(m \cos \omega_1 t)]. \end{aligned}$$

Колебание, описываемое этим выражением, имеет очень сложный спектр с большим числом слагаемых гармонических колебаний. Его можно разложить по гармоническим составляющим, амплитуды которых определяются функциями Бесселя первого рода $J_n(m)$ порядка $n = 0, 1, 2, \dots$:

$$\begin{aligned} E(t) &= E_0 J_0(m) \cos(\omega_0 t + \alpha) + \\ &\quad + E_0 \sum_{n=1}^{\infty} J_n(m) \{ \sin[(\omega_0 + n\omega_1)t + \alpha] + \\ &\quad + (-1)^n \sin[(\omega_0 - n\omega_1)t + \alpha] \}. \end{aligned}$$

Это колебание содержит бесконечное число составляющих с частотами $\omega = \omega_0 \pm n\omega_1$, $n = 0, 1, 2, \dots$

Простой случай имеет место, когда глубина модуляции фазы очень мала ($m \ll 1$). При этом $\cos(m \cos \omega_1 t) \approx 1$, $\sin(m \cos \omega_1 t) \approx m \cos \omega_1 t$, и тогда

$$E(t) = E_0 \cos(\omega_0 t + \alpha) + mE_0 \sin(\omega_0 t + \alpha) \cos \omega_1 t.$$

Записав второй член в виде суммы косинусов, придем к выражению

$$E(t) = E_0 \cos(\omega_0 t + \alpha) + \frac{mE_0}{2} \cos \left[(\omega_0 + \omega_1)t + \alpha + \frac{\pi}{2} \right] + \\ + \frac{mE_0}{2} \cos \left[(\omega_0 - \omega_1)t + \alpha + \frac{\pi}{2} \right].$$

Таким образом, в простейшем случае фазовой модуляции малой глубины спектр фазомодулированного сигнала, как и спектр амплитудно-модулированного сигнала, содержит три компоненты: одну на несущей частоте ω_0 и две на частотах, сдвинутых выше и ниже ω_0 на частоту модуляции ω_1 . Различие между амплитудно-модулированным и фазомодулированным сигналами состоит лишь в фазовом сдвиге на $\pi / 2$. Однако это различие не приводит к кардинальному отличию в форме результирующего сигнала.

Частотная модуляция. При частотной модуляции в соответствии с законом изменения управляющего сигнала изменяется частота несущего сигнала. В простейшем случае гармонического управляющего сигнала

$$\omega_0(t) = \omega_0 + k_3 E_{01} \cos \omega_1 t = \omega_0 + \Delta\omega \cdot \cos \omega_1 t,$$

где k_3 – коэффициент пропорциональности, $\Delta\omega = k_3 E_{01}$ – величина, называемая девиацией частоты. Мгновенная частота колебаний представляет собой первую производную фазы по

времени: $\omega = d\varphi(t)/dt$. Поэтому фаза частотно-модулированного колебания будет изменяться по закону

$$\varphi(t) = \int_0^t \omega(t) dt + \alpha = \omega_0 t + m \sin \omega_1 t + \alpha,$$

где $m = \frac{\Delta\omega}{\omega_1}$ – глубина (или индекс) частотной модуляции.

Видим, что при частотной модуляции, как и при фазовой, изменяется один и тот же параметр – фаза колебания. Но если при фазовой модуляции изменяется непосредственно фаза колебания, то при частотной – ее первая произвольная по времени, т.е. скорость изменения фазы. Спектры обоих колебаний, однако, не должны существенно отличаться друг от друга. В случае глубины частотной модуляции $m \ll 1$, проведя выкладки, аналогичные тем, которые были проделаны в предыдущем случае, придем к выражению

$$E(t) = E_0 \cos(\omega_0 t + \alpha) + \frac{mE_0}{2} \cos[(\omega_0 + \omega_1)t + \alpha] - \frac{mE_0}{2} \cos[(\omega_0 - \omega_1)t + \alpha].$$

Как видим, спектр частотно-модулированного сигнала отличается от спектра амплитудно-модулированного сигнала только знаком одной из составляющих, т.е. сдвигом ее фазы на π . От спектра фазомодулированного колебания он отличается отсутствием сдвига фазы на $\pm \pi / 2$ (где знак плюс относится к верхней боковой гармонике, а знак минус – к нижней).

При произвольных значениях индекса частотной модуляции m спектр частотно-модулированного сигнала содержит бесконечное количество составляющих с частотами $\omega = \omega_0 \pm n\omega_1$, где n – целое положительное число, включая нуль. Амплитуды этих составляющих, как и в случае фазомодулированного колебания,

различны и определяются функциями Бесселя первого рода n -го порядка. Эффективная ширина спектра частотно-модулированного колебания при больших индексах модуляции примерно равна удвоенному значению девиации частоты $2\Delta\omega$, а при малых значениях – $2\omega_1$.

2.2.6. Спектр затухающего цуга

Все реальные свободные колебания являются затухающими. Затухающие колебания осциллятора порождают затухающий волновой цуг. Такие цуги излучают и возбужденные атомы вещества. Поэтому представляет интерес рассмотреть фурье-спектр и затухающего цуга. Напряженность поля затухающего цуга в точке $z = 0$ изменяется по закону

$$E(t) = \begin{cases} 0, & t < 0; \\ E_0 e^{-\beta t} \cos \omega_0 t, & t > 0, \end{cases}$$

где β – коэффициент затухания, ω_0 – так называемая частота затухающих колебаний. Вид такого затухающего цуга показан на рис. 2.15.

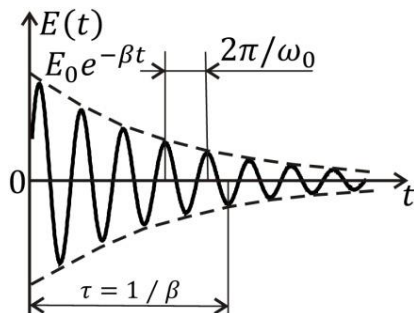


Рис. 2.15

Для получения спектра этого непериодического излучения подставим выражение $E(t)$ в формулу (2.7). При вычислении полученного интеграла удобно функцию $\cos \omega_0 t$ записать в комплексной форме $e^{-i\omega_0 t}$. Тогда получим

$$E(\omega) = E_0 \int_0^{\infty} e^{-\beta t} \cdot e^{-i\omega_0 t} \cdot e^{i\omega t} dt = E_0 \int_0^{\infty} e^{-[i(\omega_0 - \omega) + \beta]t} dt = \\ = \frac{E_0}{i(\omega_0 - \omega) + \beta}.$$

Найдя модуль функции $E(\omega)$ ($|E(t)| = \sqrt{E(t)E^*(t)}$), получим амплитудно-частотную характеристику затухающего цуга:

$$|E(\omega)| = \frac{E_0}{\sqrt{(\omega - \omega_0)^2 + \beta^2}}.$$

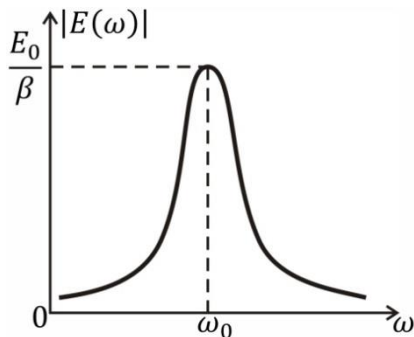


Рис. 2.16

Функция эта имеет вид резонансной кривой (рис. 2.16). Она сначала возрастает от значения $|E(0)| = E_0 / \sqrt{\omega_0^2 + \beta^2}$, при

$\omega = \omega_0$, т.е. при частоте, равной собственной частоте осциллятора, достигает максимума $|E(\omega_0)| = E_0 / \beta$, а затем с ростом разности $|\omega - \omega_0|$ спадает до нуля, не имея других максимумов. При слабом затухании, т.е. при $\beta \ll \omega_0$, или, иначе, при $\omega_0 \tau \gg 1$, где $\tau = 1 / \beta$ – время затухания колебаний, эта функция имеет острый максимум при $\omega = \omega_0$. Это означает, что слабо затухающий осциллятор излучает в узком интервале частот вблизи своей собственной частоты ω_0 .

Фазочастотную характеристику затухающего цуга найдем, представив $E(\omega)$ в виде суммы действительного и мнимого слагаемых. Записав затем это комплексное число в тригонометрической форме, найдя $\sin \varphi$ и $\cos \varphi$ и, поделив друг на друга, получим $\operatorname{tg} \varphi = (\omega - \omega_0) / \beta$. Из этого соотношения находим

$$\varphi(\omega) = \operatorname{arctg} \frac{\omega - \omega_0}{\beta}.$$

Это и есть фазочастотная характеристика затухающего цуга. Кривая, изображающая эту характеристику, начинается в точке $(0, -\operatorname{arctg}(\omega_0 / \beta))$, пересекает ось частот при $\omega = \omega_0$ и далее, монотонно возрастая, стремится к своему предельному значению $\varphi = \pi / 2$. Точка $(\omega_0, 0)$ является точкой перегиба кривой $\varphi(\omega)$. Вид этой кривой показан на рис. 2.17. В случае незатухающего цуга (при $\beta = 0$) фазочастотная характеристика изображалась бы горизонтальной прямой $\varphi(\omega) = \pi / 2 = \operatorname{const}$.

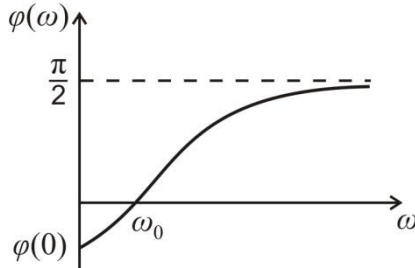


Рис. 2.17

Спектральная плотность энергии излучения

$$|E(\omega)|^2 = \frac{E_0^2}{(\omega - \omega_0)^2 + \beta^2}.$$

Кривая, изображающая это распределение, показана на рис. 2.18. Это распределение имеет вид квадрата резонансной кривой. Величина $|E(\omega)|^2$ сначала возрастает от значения $|E(0)|^2$, при $\omega = \omega_0$ она достигает максимума, а затем, убывая, стремится к нулю. Распределения такого вида принято называть лоренцовым контуром. Спад кривой $|E(\omega)|^2$ будет происходить тем быстрее, чем меньше коэффициент затухания β . При слабом затухании максимум кривой будет острым. Таким образом, спектр слабо затухающего цуга представляет собой узкую полосу сплошного спектра с центром при $\omega = \omega_0$.

Как легко убедиться, спектральная плотность энергии излучения уменьшается вдвое (по сравнению с $|E(\omega_0)|^2$) при частотах $\omega_{1,2} = \omega_0 \pm \beta = \omega_0 \pm 1/\tau$, где учтено, что $\beta = 1/\tau$. Поэтому в качестве ширины спектрального распределения $\Delta\omega$ можно принять величину примерно равную разности $\omega_2 - \omega_1$ (рис. 2.18). Тогда получим

$$\Delta\omega \approx 2 / \tau. \quad (2.20)$$

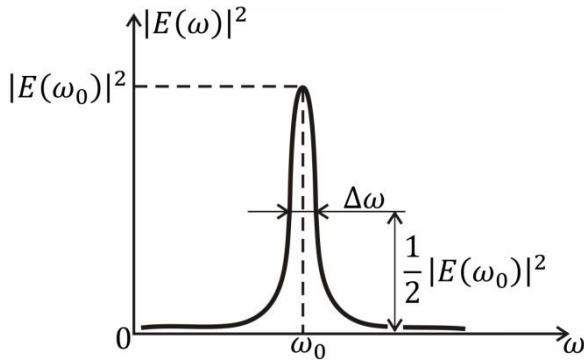


Рис. 2.18

Как видим, ширина полосы излучаемых частот $\Delta\omega$ связана с характерной длительностью цуга τ примерно таким же соотношением, как и для незатухающего цуга. Это является еще одной иллюстрацией соотношения неопределенностей частота – время. Чем больше время затухания τ , тем меньше ширина спектра $\Delta\omega$. В пределе $\tau \rightarrow \infty$ (синусоидальная волна) спектр излучения стягивается в бесконечно узкую линию, соответствующую частоте $\omega = \omega_0$. При увеличении затухания контур расширяется, а величина максимума снижается.

Затухающий осциллятор, как уже отмечалось, может служить классической моделью атома или молекулы в режиме излучения. Время затухания τ называют временем излучения атома. По порядку величины это время составляет $2,5 \cdot 10^{-8}$ с. Кривую, изображенную на рис. 2.18, в классической теории излучения атома называют контуром спектральной линии, а ширину частотного распределения $\Delta\omega$ – шириной спектральной линии. Следовательно, спектральная линия излучения атома имеет конечную ширину. В соответствии с формулой (2.20) эта ширина будет тем меньше, чем больше время затухания. Если бы

колеблющийся оптический электрон вовсе не испытывал затухания, то спектр излучения атома состоял бы из одной бесконечно узкой спектральной линии, характеризуемой вполне определенным значением частоты $\omega = \omega_0$. Ширина спектральной линии, обусловленная радиационным затуханием, $\Delta\omega \approx 2 / 2,5 \cdot 10^{-8} = 0,8 \cdot 10^8 \text{ с}^{-1}$. Эта ширина, называемая естественной шириной спектральной линии, очень мала и составляет примерно $2 \cdot 10^{-8}$ от частоты самой линии.

2.3. Пространственный спектр излучения

2.3.1. Спектр одномерного сигнала

До сих пор мы рассматривали применение математического аппарата Фурье к анализу частотного спектра полей, зависящих только от времени. Рассмотрим теперь случай, когда поле является функцией пространственных переменных. Обратимся сначала к простейшему случаю, когда поле E является функцией только одной пространственной переменной (z). Положив в функции поля $E(z, t)$ значение $t = 0$, получим зависимость напряженности электрического поля от координаты z в фиксированный момент времени $t = 0$: $E(z) = E(z, 0)$. Эту функцию одной пространственной переменной z тоже можно разложить либо в ряд Фурье, либо в интеграл Фурье (предполагается, что, как и в рассмотренном выше случае, для функции $E(t)$, условия Дирихле выполнены). Теперь разложение будет производиться по гармоническим функциям $\exp(ikz)$.

Разложение периодической функции $E(z)$ в пространственный комплексный ряд Фурье будет иметь вид

$$E(z) = \sum_{m=-\infty}^{\infty} E_m e^{ik_m z}, \quad (2.21)$$

где коэффициенты разложения

$$E_m = \frac{1}{d} \int_{-d/2}^{d/2} E(z) e^{-ik_m z} dz. \quad (2.22)$$

Здесь $k_m = mk_1$, $m = 1, 2, 3, \dots$, $k_1 = 2\pi/d$, d – пространственный период функции $E(z)$. Члены ряда (2.21) называются пространственными гармониками. Коэффициенты E_m представляют собой амплитуды пространственных гармоник, k_m – круговые частоты пространственных гармоник; k_1 – круговая пространственная частота первой (основной) гармоники, $1/d = \omega_1$ – пространственная частота первой гармоники.

Легко видеть, что формулы (2.21) и (2.22) получаются из формул (2.2) и (2.3), если в этих формулах заменить время t на пространственную координату z , временную частоту $\omega_m = m\omega_1$ на пространственную частоту $k_m = mk_1 = m \frac{2\pi}{d}$, а i на $-i$, и наоборот. Аналогично с учетом указанных замен записывается пространственный ряд Фурье и в тригонометрической форме.

Непериодическую функцию $E(z)$ можно разложить в пространственный интеграл Фурье, являющийся непрерывной суммой плоских волн различных пространственных частот:

$$E(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} E(k) e^{ikz} dk. \quad (2.23)$$

При этом

$$E(k) = \int_{-\infty}^{\infty} E(z) e^{-ikz} dz. \quad (2.24)$$

Интегралы (2.23) и (2.24) называют соответственно обратным и прямым одномерным пространственным преобразованием Фурье. Координата z и компонента $k_z = k$ волнового вектора \mathbf{k} –

сопряженные параметры преобразований. В интеграле (2.23) компоненты с $k > 0$ соответствуют волнам, распространяющимся в положительном направлении оси Z , а компоненты с $k < 0$ – волнам, распространяющимся в отрицательном направлении оси Z . Используя пространственную частоту w , их можно записать в симметричной форме:

$$E(z) = \int_{-\infty}^{\infty} E(w) e^{i2\pi w z} dw, \quad (2.25)$$

$$E(w) = \int_{-\infty}^{\infty} E(z) e^{-i2\pi w z} dz. \quad (2.26)$$

Заметим, что временное и пространственное преобразования Фурье получаются друг из друга заменой временных величин ω , t (или ν , t) пространственными величинами k , z (или w , z) и i на $-i$, и наоборот. Величины $|E(k)|$ и $|E(k)|^2$ имеют тот же смысл, что и величины $|E(\omega)|$ и $|E(\omega)|^2$ с учетом аналогии между k и ω .

В качестве примера применения пространственного преобразования Фурье найдем пространственный спектр рассмотренного в п. 2.2.2 незатухающего волнового цуга. Для этого следует рассмотреть пространственное распределение поля цуга. Если цуг расположен вдоль оси Z и его протяженность равна l , то функция, описывающая поле цуга, будет иметь вид

$$E(z) = \begin{cases} E_0 \exp(ik_0 z), & |z| \leq l/2; \\ 0, & |z| > l/2, \end{cases}$$

где k_0 – волновое число цуга. Произведя преобразование Фурье по формуле (2.24), получим функцию распределения амплитуд гармонических составляющих по круговым пространственным частотам (пространственный спектр амплитуд):

$$|E(k)| = |E_0| l \operatorname{sinc} \left[\frac{(k - k_0)l}{2\pi} \right].$$

По форме эта функция имеет такой же вид, как и функция $|E(\omega)|$ временного цуга. Такой же, как и $|E(\omega)|^2$, будет иметь вид и распределение плотности энергии по спектру пространственных частот (пространственный энергетический спектр):

$$|E(k)|^2 = |E_0|^2 l^2 \operatorname{sinc}^2 \left[\frac{(k - k_0)l}{2\pi} \right] = |E(k_0)|^2 \operatorname{sinc}^2 \left[\frac{(k - k_0)l}{2\pi} \right],$$

где $|E(k_0)|^2 = |E_0|^2 l^2$ – максимальное значение спектральной плотности энергии.

По аналогии с частотным спектром в качестве ширины спектра пространственных частот можно приближенно принять расстояние от точки максимума k_0 до первого минимума $k_1 = k_0 + 2\pi/l$, т.е. величину $k_1 - k_0$. Тогда получим

$$\Delta k \approx 2\pi/l. \quad (2.27)$$

Соотношение (2.27) связывает пространственную протяженность цуга l с шириной интервала его пространственных частот Δk . Его называют соотношением неопределенностей пространственная частота – длина. Как и соотношение (2.13), оно справедливо для всех волновых процессов и показывает, что чем меньше пространственная протяженность волнового процесса, тем шире интервал пространственных частот Δk волн, которые его образуют. Из соотношений (2.13) и (2.27), в частности, следует, что в случае строго монохроматической плоской волны, для которой $\Delta k = 0$ и $\Delta \omega = 0$, имеем $l = \infty$ и $\tau = \infty$. Это означает,

что строго плоская монохроматическая волна имеет бесконечную протяженность и длительность. Видим, что ограничение длительности и протяженности волны приводит к ее немонохроматичности; разброс направлений волновых векторов относительно среднего направления при этом не происходит.

В качестве другого примера одномерного сигнала рассмотрим цуг с гауссовым распределением амплитуды. Такой цуг в момент времени $t = 0$ описывается функцией

$$E(z) = \frac{E_0}{\sqrt{2\pi\alpha}} e^{-\frac{z^2}{2\alpha}} \cos k_0 z,$$

где E_0 , α и k_0 – постоянные параметры. Параметр α определяется из того условия, что на расстоянии z_0 амплитуда цуга уменьшается в e раз. Тогда получим $2\alpha = z_0^2$. Фурье-спектр такого цуга определится как

$$\begin{aligned} E(k) &= \frac{E_0}{\sqrt{2\pi\alpha}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2\alpha}} \cos k_0 z e^{-ikz} dz = \\ &= \frac{E_0}{\sqrt{2\pi\alpha}} \int_{-\infty}^{\infty} \exp\left[-\frac{z^2}{4\alpha} - i(k - k_0)z\right] dz. \end{aligned}$$

Введем обозначения $1 / (2\alpha) = a$, $i(k - k_0) = b$ и вычислим

интеграл $\int_{-\infty}^{\infty} \exp(-az^2 - bz) dz$. Имеем:

$$\begin{aligned} \int_{-\infty}^{\infty} \exp(-az^2 - bz) dz &= e^{\frac{b^2}{2a}} \int_{-\infty}^{\infty} \exp\left[-a\left(z - \frac{b}{2a}\right)^2\right] dz = \\ &= e^{\frac{b^2}{2a}} \int_{-\infty}^{\infty} e^{-a\xi^2} d\xi = \sqrt{\frac{\pi}{a}} e^{\frac{b^2}{2a}}. \end{aligned} \quad (2.28)$$

Возвращаясь к старым обозначениям, будем иметь

$$E(k) = E_0 \exp[-\alpha(k - k_0)^2] .$$

Интенсивность излучения

$$I(k) = I_0 \exp[-2\alpha(k - k_0)^2] .$$

Профиль интенсивности в координатах k , I имеет вид гауссовой кривой с вершиной в точке k_0 . Интенсивность уменьшается в e раз при $2\alpha(k - k_0)^2 = 1$. Учитывая, что $2\alpha = z_0^{-2}$, это равенство можно переписать в виде $[z_0(k - k_0)]^2 = 1$. Примем величину $2z_0$ в качестве пространственной протяженности цуга l , а величину $2(k - k_0)$ в качестве ширины интервала его пространственных частот Δk . Тогда приходим к соотношению неопределенностей пространственная частота – длина в виде $\Delta k = 4 / l$. Полученное соотношение практически не отличается от соотношения (2.27).

2.3.2. Спектр двумерного сигнала

Перейдем теперь к рассмотрению пространственного спектра комплексной амплитуды поля как функции двух пространственных переменных x и y . Такие функции играют основную роль в оптике, где информация часто представляется распределением амплитуды и фазы светового поля в некоторой плоскости XU .

Пусть в плоскости XU задано распределение комплексной амплитуды некоторого монохроматического светового поля $E(x, y)$. Аппарат спектрального разложения можно применить и к функции двух пространственных переменных (при выполнении соответствующих условий). Тогда по аналогии с (2.23) можно написать разложение функции $E(x, y)$ в двойной интеграл Фурье:

$$E(x, y) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(k_x, k_y) e^{i(k_x x + k_y y)} dk_x dk_y. \quad (2.29)$$

При этом по аналогии с (2.24) будем иметь

$$E(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) e^{-i(k_x x + k_y y)} dx dy. \quad (2.30)$$

Величина $E(k_x, k_y) \frac{dk_x dk_y}{(2\pi)^2}$ определяет вклад в $E(x, y)$ плоских

монохроматических составляющих из интервала пространственных частот от k_x до $k_x + dk_x$ и от k_y до $k_y + dk_y$. Формулы (2.29) и (2.30) можно записать в симметричной форме по аналогии с (2.25) и (2.26), используя пространственные частоты u и v :

$$E(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) e^{i2\pi(ux + vy)} du dv, \quad (2.31)$$

$$E(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) e^{-i2\pi(ux + vy)} dx dy. \quad (2.32)$$

При этом $E(-u, -v) = E^*(u, v)$, так как функция $E(x, y)$ вещественная. Спряженными переменными преобразований Фурье являются здесь величины x, y и k_x, k_y (или x, y и u, v).

Выясним физический смысл написанных разложений функции $E(x, y)$. Для этого рассмотрим плоскую волну с комплексной амплитудой (1.27). Такая волна создает в плоскости $z = 0$ распределение поля

$$E_1(x, y) = E_0 \exp [i2\pi(ux + vy)].$$

Будем считать, что амплитуда волны E_0 равна единице. Тогда легко видеть, что экспоненциальный множитель в интеграле (2.31) представляет собой элементарную плоскую волну (гармоническую составляющую) с комплексной амплитудой

$$E_0(x, y) = \exp[i2\pi(ux + vy)]$$

и с пространственными частотами u , v и

$$w = \sqrt{1/\lambda^2 - u^2 - v^2},$$

т.е. с направляющими косинусами

$$\cos\alpha = \lambda u, \cos\beta = \lambda v, \cos\gamma = \lambda w = \sqrt{1 - (\lambda u)^2 - (\lambda v)^2}. \quad (2.33)$$

Следовательно, плоские волны разных направлений – это волны разных пространственных частот.

Двумерную элементарную комплексную волну $E_1(x, y)$ физически можно представить гармонической функцией в одном направлении (в плоскости XY) и постоянной в направлении, перпендикулярном первому. Положим $2\pi(u + vy) = m \cdot 2\pi$, где m – целое число, включая нуль. Тогда получим

$$ux + vy - m = 0. \quad (2.34)$$

Эти уравнения определяют семейство прямых (рис. 2.19), вдоль которых фаза функции $\exp[i2\pi(ux + vy)]$ равна нулю (прямых нулевой фазы). Эти прямые пересекают оси X и Y в точках $x_m = m/u$ и $y_m = m/v$ соответственно. Пространственный период d гармонического изменения функции

$$\exp[i2\pi(ux + vy)]$$

равен расстоянию между соседними линиями нулевой фазы. Это расстояние можно найти, если учесть, что величина

$m/\sqrt{u^2 + v^2}$ определяет расстояние прямой (2.34) от начала координат. Тогда расстояние между двумя соседними прямыми ($\Delta m = 1$) – период d функции $\exp [i2\pi (ux + vy)]$ – найдется как

$$d = 1/\sqrt{u^2 + v^2} .$$

Частоты u и v , являющиеся коэффициентами при x и y прямых (2.34), определяют координаты нормали \vec{N} к этим прямым. Для любой определенной пары частот u и v , т.е. для любой элементарной плоской волны в интеграле (2.29), направление вектора нормали \vec{N} определяется углом α между этим вектором и осью X . При этом

$$\operatorname{tg} \alpha = N_y / N_x = v/u .$$

Таким образом, двойной интеграл (2.29) или (2.31) является представлением комплексной амплитуды поля $E(x, y)$ в виде бесконечно большого набора гармонических составляющих всех пространственных частот k_x, k_y или u, v (спектра плоских волн, распространяющихся в различных направлениях).

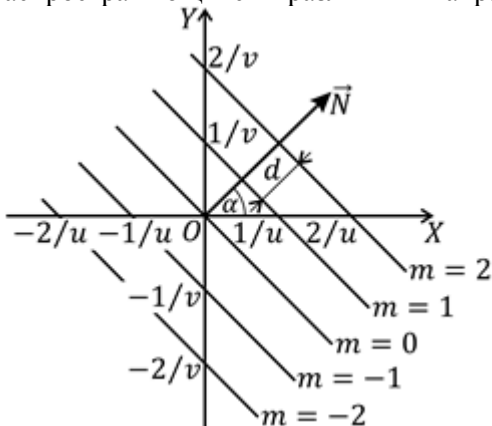


Рис. 2.19

При этом каждой пространственной гармонике в спектре поля $E(x, y)$ соответствует плоская волна в координатном пространстве. Каждая такая волна характеризуется своими направляющими косинусами, которые определяют комплексную амплитуду волны и

направление ее распространения. Следовательно, разлагая поле в плоскости $z = 0$ по пространственным гармоникам, мы тем самым разлагаем поле во всем пространстве (т.е. при $z > 0$) по плоским волнам. Это утверждение справедливо для всех пространственных частот, удовлетворяющих условию

$$u^2 + v^2 \leq 1 / \lambda^2.$$

Функция $E(u, v)$ (или функция $E(k_x, k_y)$) представляет собой двумерный фурье-образ или двойное преобразование Фурье функции $E(x, y)$ и называется пространственным или угловым спектром комплексной амплитуды поля $E(x, y)$. Как видим, спектр пространственных частот оптического сигнала $E(x, y)$, заданного в плоскости XU , представляется в виде суперпозиции плоских волн, распространяющихся под различными углами от плоскости XU . Таким образом, если пространственный сигнал $E(x, y)$ представлен в виде амплитуд и фаз поля монохроматической световой волны в некоторой плоскости XU , то спектр его пространственных частот отображается распределением амплитуд и фаз плоских световых волн, распространяющихся под различными углами к оси Z . При распространении светового сигнала в свободном пространстве его спектральные характеристики проявляются в распределении света по углам α и β .

Квадрат модуля функции $E(u, v)$ (или $E(k_x, k_y)$), т.е. величина $|E(u, v)|^2$ (или $|E(k_x, k_y)|^2$) определяет энергетический спектр комплексного светового поля. Теорема Парсеваля для двумерного пространственного преобразования Фурье имеет вид

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |E(x, y)|^2 dx dy = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |E(k_x, k_y)|^2 dk_x dk_y$$

или

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |E(x, y)|^2 dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |E(u, v)|^2 du dv.$$

Величина $\left| E(k_x, k_y) \right|^2 dk_x dk_y$ определяет энергию (и интенсивность) волны в интервале пространственных частот от k_x до $k_x + dk_x$ и от k_y до $k_y + dk_y$.

Интегралы (2.29) и (2.31) представляют собой двойные обратные преобразования Фурье. Обратное преобразование Фурье позволяет восстановить распределение поля $E(x, y)$ по спектральной характеристике $E(u, v)$. Отметим также, что функция

$$\exp[-i2\pi(ux + vy)]$$

называется ядром преобразования Фурье.

Если поле $E(x, y)$ есть функция с разделяющимися переменными, т.е. если ее можно представить в виде произведения

$$E(x, y) = E_1(x) \cdot E_2(y),$$

то

$$\begin{aligned} F\{E(x, y)\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) e^{-i2\pi(ux + vy)} dx dy = \\ &= \int_{-\infty}^{\infty} E_1(x) e^{-i2\pi ux} dx \cdot \int_{-\infty}^{\infty} E_2(y) e^{-i2\pi vy} dy = F\{E_1(x)\} \cdot F\{E_2(y)\}. \end{aligned}$$

Следовательно, двумерное преобразование Фурье функции с разделяющимися переменными равно произведению одномерных преобразований Фурье функций-сомножителей.

Представление поля в виде суперпозиции отдельных гармоник – не только удобный математический аппарат, но и способ описания, наиболее близкий к реальным условиям эксперимента. Формулы (2.29) – (2.32) являются исходными при анализе

оптических систем методом Фурье. Ниже мы убедимся, что функция $E(x, y)$ и ее фурье-образ $E(u, v)$ определяют соответственно распределение светового поля в плоскости XU какого-либо оптического прибора (диафрагмы, линзы, зеркала и т.д.) и другой плоскости $X'Y'$ пространства, причем пространственные частоты

$$u = x' / \lambda z, v = y' / \lambda z,$$

где z – либо фокусное расстояние выходной линзы оптической системы, либо расстояние между указанными плоскостями.

Спектр пространственного сигнала можно наблюдать в виде действительного распределения амплитуд и фаз света в некоторой области пространства. Пространственный спектр можно воспроизвести, например, с помощью линзы. Каждая плоская волна, падающая на линзу, фокусируется ею в точку (точнее, в дифракционное пятно) в фокальной плоскости линзы. Распределение света в фокальной плоскости линзы будет в некотором масштабе воспроизводить фурье-спектр $E(u, v)$ поля $E(x, y)$. Линза, таким образом, осуществляет в своей фокальной плоскости пространственное преобразование Фурье, подобно тому, как, например, призма или дифракционная решетка производит разложение падающего излучения в частотный спектр. Преобразование Фурье временного сигнала как экспериментальная процедура осуществляется посредством сложных устройств – электронных анализаторов. Та же процедура для пространственных сигналов, как видим, может быть осуществлена с помощью простой линзы.

2.4. Квазимонохроматический свет

Как уже неоднократно отмечалось, реальные световые волны плоскими монохроматическими не являются. Однако любую реальную волну можно представить как суперпозицию плоских монохроматических волн, частоты и волновые числа которых

непрерывно заполняют интервалы шириной соответственно $\Delta\omega$ и Δk . Замена же суперпозиции волн одной плоской монохроматической волной допустима лишь в некоторых конкретных случаях и при определенных оговорках.

Следует отметить, что с помощью неограниченной в пространстве и во времени плоской монохроматической волны (если бы такая волна реально существовала) – волны, имеющей одинаковые свойства в любой точке и в любой момент времени, невозможно передать никакой информации, кроме информации о ее амплитуде, частоте и поляризации. Поэтому плоская монохроматическая волна не может быть использована в качестве сигнала. Чтобы с помощью такой волны можно было передать информацию, ее параметры следует соответствующим образом изменять. Ее следует, например, оборвать на какое-либо время или изменять ее амплитуду, частоту или фазу в соответствии с передаваемым сообщением. Но такая промодулированная волна уже не является строго плоской монохроматической волной. Ее можно считать квазимонохроматической волной. К рассмотрению таких волн мы теперь и переходим.

Представим себе волновой процесс, у которого ширина частотного интервала $\Delta\omega$ и соответствующего интервала волновых чисел Δk мала, т.е. когда $\Delta\omega \ll \omega_0$ и $\Delta k \ll k_0$, где ω_0 – некоторая центральная частота из интервала $(\omega_0 - \Delta\omega/2, \omega_0 + \Delta\omega/2)$, а k_0 – волновое число из интервала $(k_0 - \Delta k/2, k_0 + \Delta k/2)$. В этом случае частоты и волновые числа компонент, образующих волновой процесс, мало отличаются друг от друга и сгруппированы вблизи некоторой центральной частоты ω_0 и волнового числа k_0 . Такой волновой процесс и называют квазимонохроматической волной. Его называют также группой волн или волновым пакетом. Примерами квазимонохроматических волн могут служить рассмотренные выше незатухающий цуг, если его длительность τ такова, что

$$\omega_0 \tau = 2\pi(\omega_0 / \Delta\omega) \gg 1$$

(использовано соотношение неопределенностей (2.13)), и затухающий пуг, если его время затухания τ удовлетворяет условию $\omega_0 \tau = 2(\omega_0 / \Delta\omega) \gg 1$ (использовано соотношение (2.20)). Условие $\Delta\omega / \omega_0 \ll 1$ или равносильное ему, условие $\omega_0 \tau \gg 1$ выполняется и для атома (где величина $\omega_0 \tau = 2 \cdot 10^{-8}$), поэтому излучение атома с большой точностью также можно считать квазимонохроматическим.

Предположим для простоты, что группа волн распространяется в положительном направлении оси Z . Рассмотрим совокупность волн, круговые частоты которых заключены в интервале от ω до $\omega + d\omega$. Амплитуды волн, входящих в эту элементарную группу, равны $E(\omega) d\omega / 2\pi$. Тогда, если $dE(z, t)$ – результирующее возмущение для указанной бесконечно малой группы волн, то

$$dE(z, t) = \frac{1}{2\pi} E(\omega) \exp[i(kz - \omega t)] d\omega,$$

а возмущение в конечном интервале частот от $\omega_0 - \Delta\omega/2$ до $\omega_0 + \Delta\omega/2$

$$E(z, t) = \frac{1}{2\pi} \int_{\omega_0 - \Delta\omega/2}^{\omega_0 + \Delta\omega/2} E(\omega) e^{i(kz - \omega t)} d\omega. \quad (2.35)$$

Аналогично можно получить

$$E(z, t) = \frac{1}{2\pi} \int_{k_0 - \Delta k/2}^{k_0 + \Delta k/2} E(k) e^{i(kz - \omega t)} dk. \quad (2.36)$$

Интегралы (2.35) и (2.36) представляют собой разложение волновой функции поля $E(z, t)$ по плоским волнам $\exp[i(kz - \omega t)]$ в конечном интервале частот и волновых чисел. Величины k и ω в выражениях (2.35) и (2.36) функционально связаны друг с другом: $\omega = \omega(k)$ (а значит, и $k = k(\omega)$). В теории волн эту зависимость называют законом дисперсии. Для вакуума закон дисперсии имеет вид $\omega = ck$. Знание этого закона достаточно для исследования поведения любой волны с течением времени.

Разлагая частоту ω как функцию волнового числа k по степеням разности $k - k_0$ и ограничиваясь линейным членом (ввиду малости $k - k_0$), получим

$$\omega = \omega_0 + \left(\frac{d\omega}{dk} \right)_0 (k - k_0),$$

где индекс 0 у производной указывает, что производная берется в точке $k = k_0$. Кроме того, представим k в виде

$$k = k_0 + (k - k_0).$$

Тогда получим

$$E(z, t) = e^{i(k_0 z - \omega_0 t)} \int_{k_0 - \Delta k/2}^{k_0 + \Delta k/2} E(k) e^{i(kz - \omega t) - (k - k_0)z} dk, \quad (2.37)$$

где

$$u = \left(\frac{d\omega}{dk} \right)_0, \quad (2.38)$$

Поле $E(k)$ является медленно меняющейся функцией внутри интервала интегрирования (ввиду того, что все k близки к k_0), поэтому можно принять $E(k) \approx E(k_0)$, а постоянную величину $E(k_0)$ вынести за знак интеграла. Кроме того, приняв $k - k_0$ в

качестве новой переменной интегрирования ξ , т.е. положив $k - k_0 = \xi$, для функции $E(z, t)$, описывающей группу волн, будем иметь

$$E(z, t) = E(k_0) e^{i(k_0 z - \omega_0 t)} \int_{-\Delta k/2}^{\Delta k/2} e^{i(kz - ut)\xi} d\xi.$$

Выполнив интегрирование, получим

$$\begin{aligned} E(z, t) &= E(k_0) \Delta k \frac{\sin \frac{1}{2}(z - z_0) \Delta k}{\frac{1}{2}(z - z_0) \Delta k} e^{i(k_0 z - \omega_0 t)} = \\ &= E_0(z, t) e^{i(k_0 z - \omega_0 t)}, \end{aligned} \quad (2.39)$$

где выражение $E_0(z, t)$ включает в себя все множители, стоящие перед экспонентой, а

$$z_0 = ut. \quad (2.40)$$

Величина

$$\begin{aligned} |E_0(z, t)| &= \left| E(k_0) \Delta k \frac{\sin \frac{1}{2}(z - z_0) \Delta k}{\frac{1}{2}(z - z_0) \Delta k} \right| = \\ &= \left| E(k_0) \Delta k \operatorname{sinc} \frac{(z - z_0) \Delta k}{2\pi} \right| \end{aligned} \quad (2.41)$$

есть медленно (по сравнению с $\exp[i(k_0 z - \omega_0 t)]$) меняющаяся функция времени t и координаты z . Поэтому функцию $|E_0(z, t)|$ можно рассматривать как амплитуду квазимонохроматической волны с частотой ω_0 и волновым числом k_0 , а $k_0 z - \omega_0 t$ — как ее

фазу. Амплитуда имеет вид модуля sinc-функции. Она достигает максимума в точке z_0 . Эту точку называют центром группы. В центре группы фазы всех волн, образующих группу, одинаковы. Поэтому сложение этих волн и дает максимальную амплитуду волнового процесса. Согласно соотношениям (2.38) и (2.40), центр группы волн движется со скоростью

$$u = \frac{d\omega}{dk},$$

где индекс 0 опущен за ненадобностью. Полученная скорость представляет собой групповую скорость волны. Именно с этой скоростью переносится энергия (и информация) световым сигналом. В вакууме, где закон дисперсии определяется формулой $\omega = ck$, групповая скорость равна фазовой и равна скорости света в вакууме ($u = c$). В других средах групповая скорость отличается от фазовой и может быть как больше, так и меньше ее. Во всех средах волновой пакет расплывается с течением времени (так как компоненты различных частот имеют разную скорость) и только в вакууме волновой пакет перемещается со скоростью c как единое целое, без изменения своей формы (не расплываясь). Зависимость скорости волны (а значит, и показателя преломления n среды) от ее частоты ($n = n(\omega)$) называется дисперсией волны. Так как дисперсионные свойства однородной среды не могут зависеть от направления, то ω следует считать четной функцией k , т.е. что $\omega(-k) = \omega(k)$.

Расстояние от точки максимума $|E_0(z, t)|$ до первого минимума ($z_1 = z_0 + 2\pi / \Delta k$) можно принять за меру, определяющую размер группы. Обозначив его через Δz , будем иметь

$$\Delta z \Delta k \approx 2\pi.$$

Тем самым мы снова пришли к соотношению неопределенностей для волн: произведение линейных размеров группы волн Δz на

интервал волновых чисел Δk волн, из которых построена группа, есть величина постоянная и приближенно равная 2π . Видно, что волна, у которой волновые числа составляющих гармоник заполняют спектральный интервал Δk , оказывается практически локализованной в области пространства шириной Δz (отсюда и название – волновой пакет).

Из этого соотношения можно получить и соотношение неопределенностей частота – время $\Delta\omega \Delta t \approx 2\pi$. Пакет шириной Δz перемещается вдоль оси Z со скоростью u , однако момент прохождения им заданной точки на оси Z не может быть указан точно: неопределенность в определении этого момента оказывается порядка $\Delta t \approx \Delta z / u$. Волновой пакет характеризуется размытостью по частоте $\Delta\omega$, причем

$$\Delta\omega \approx \frac{d\omega}{dk} \Delta k = u \Delta k.$$

Из этих двух приближенных равенств находим, что

$$\Delta\omega \Delta t \approx \Delta k \Delta z \approx 2\pi.$$

Таким образом, квазимонохроматическую световую волну можно представить как волну монохроматическую с частотой ω_0 и волновым числом k_0 , модулированную по комплексной амплитуде или (если волновую функцию записывать в тригонометрической форме), по амплитуде и начальной фазе. В тригонометрической форме квазимонохроматическую волну, распространяющуюся вдоль оси Z , записывают в виде

$$E(z, t) = E_0(z, t) \cos[\omega_0 t \pm k_0 z + \alpha(z, t)],$$

или (объединив $k_0 z$ и $\alpha(z, t)$ и обозначив их сумму через $\varphi(z, t)$)

$$E(z, t) = E_0(z, t) \cos[\omega_0 t + \varphi(z, t)], \quad (2.42)$$

где $E_0(z, t)$ и $\varphi(z, t)$ – медленно (по сравнению с $\cos \omega_0 t$) изменяющиеся функции координаты и времени. Следует, однако, отметить, что изменения эти происходят хаотически, амплитуда и фаза являются случайными функциями времени. (В детерминированном модулированном электромагнитном сигнале функции $E_0(z, t)$ и $\varphi(z, t)$ являются точно известными). При этом случайные функции $E_0(z, t)$ и $\varphi(z, t)$ могут изменяться независимо друг от друга или, как говорят, быть статистически независимыми.

Квазимонохроматический свет реального источника можно считать хаотически модулированным по амплитуде и фазе. Время модуляции обычно имеет порядок длительности цуга τ . Время τ является характерным временем изменения амплитуды $E_0(z, t)$ волны и ее фазы $\varphi(z, t)$. Его называют временем корреляции или временем когерентности. Считается, что значения амплитуды E_0 (а также и фазы φ) в моменты времени t и $t + t_0$, т.е. $E_0(z, t)$ и $E_0(z, t + t_0)$ (а также $\varphi(z, t)$ и $\varphi(z, t + t_0)$) коррелированы при $t_0 < \tau$ и не коррелированы при $t_0 > \tau$.

Таким образом, квазимонохроматический свет можно характеризовать двумя временными масштабами: периодом $T = 2\pi / \omega_0$ центральной гармоники (периодом «несущего колебания») и временем корреляции τ , или временем когерентности.

Квазимонохроматическую волну, распространяющуюся в произвольном направлении, можно записать в виде

$$E(\mathbf{r}, t) = E_0(\mathbf{r}, t) \cos [\omega_0 t \pm \mathbf{k}_0 \cdot \mathbf{r} + \alpha(\mathbf{r}, t)],$$

или, объединив члены $\pm \mathbf{k}_0 \cdot \mathbf{r}$ и $\alpha(\mathbf{r}, t)$, в виде

$$E(\mathbf{r}, t) = E_0(\mathbf{r}, t) \cos [\omega_0 t + \varphi(\mathbf{r}, t)].$$

В комплексной форме квазимонохроматическая волна представится формулой

$$E(\mathbf{r}, t) = E_0(\mathbf{r}, t) \exp\{-i[\omega_0 t + \varphi(\mathbf{r}, t)]\} = \\ = E_0(\mathbf{r}, t) \exp[-i\varphi(\mathbf{r}, t)] \cdot \exp(-i\omega_0 t).$$

Комплексную функцию $E_0(\mathbf{r}, t)\exp[-i\varphi(\mathbf{r}, t)]$ обычно обозначают как $E(\mathbf{r}, t)$ и называют оптическим сигналом. По существу, оптический сигнал – это комплексная амплитуда электромагнитных колебаний или огибающая квазигармонических колебаний. Комплексная функция $E(x, y, z)$, названная в п. 1.5 оптическим сигналом, на самом деле представляет собой стационарный оптический сигнал, когда поле E не зависит от времени: $E = E(x, y, z) \equiv E(\mathbf{r})$. Если, кроме того, точка наблюдения P лежит в плоскости $z = 0$, то $E = E(x, y)$. В дальнейшем изложении под оптическим сигналом мы будем понимать либо распределение комплексной амплитуды в плоскости XY (комплексную функцию $E(x, y)$), либо распределение интенсивности в этой плоскости (действительную функцию $I(x, y)$).

Квазимонохроматические колебания поля, вызванные квазимонохроматической волной, в какой-либо фиксированной точке в комплексной форме можно представить в виде

$$E(t) = E_0(t)\exp(-i\omega_0 t), \quad (2.43)$$

где комплексная амплитуда

$$E_0(t) = E_{00}(t)\exp[-\alpha(t)].$$

В выражении (2.43) случайный (статистический) характер имеет первый сомножитель $E_0(t)$, второй же сомножитель $\exp(-i\omega_0 t)$ описывает гармонические колебания с циклической частотой ω_0 . Функция $E_0(t)$ описывает закон случайной амплитудно-фазовой модуляции гармонического колебания.

Соотношение (2.43) полезно получить и исходя из модели излучения источника. Излучение реального (не лазерного) источника в силу фундаментальных физических причин имеет статистический характер. Реальный источник – это совокупность огромного числа атомов. Каждый из этих атомов становится элементарным источником светового излучения, если вследствие какого-либо воздействия (столкновение с другими атомами, электронный удар и т.д.) он будет переведен в возбужденное состояние. Случайный статистический характер механизма возбуждения и определяет статистическую природу излучения. Любой из атомов может начать «высвечиваться» в произвольный, случайный момент времени. В излучении одновременно принимает участие большое число атомов. Каждый атом излучает цуг волн длительности τ и протяженности $c\tau$. Излучение всего источника можно рассматривать как совокупность большого числа цугов, имеющих случайные фазы и случайно распределенных во времени. Поэтому по принципу суперпозиции полное поле в точке наблюдения в любой момент времени t можно представить в виде суммы полей, создаваемых каждым атомом источника:

$$E(t) = \sum_j E_j(t - t_j),$$

где t_j – момент начала излучения j -го атома. Суммирование ведется по всем N цугам, присутствующим в излучении в момент времени t . Для каждого цуга волн можно записать разложение Фурье

$$E_j(t - t_j) = \frac{1}{2\pi} \int_{-\infty}^{\infty} E_j(\omega) e^{-i\omega(t - t_j)} d\omega.$$

Тогда получим

$$E(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\sum_j E_j(\omega) e^{i\omega t_j} \right) e^{-i\omega t} d\omega.$$

Введя среднюю частоту ω_0 и рассматривая колебание $E(t)$ как квазимонохроматическое, представим это выражение в виде

$$\begin{aligned}
 E(t) &= e^{-i\omega_0 t} \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\sum_j E_j(\omega) e^{i\omega t_j} \right) e^{-i(\omega - \omega_0)t} d\omega = \\
 &= e^{-i\omega_0 t} \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\omega) e^{-i(\omega - \omega_0)t} d\omega, \quad (2.44)
 \end{aligned}$$

где

$$f(\omega) = \sum_j E_j(\omega) e^{i\omega t_j}.$$

Интеграл в (2.44) (с коэффициентом $1 / 2\pi$) представляет собой величину, зависящую от времени. Обозначив ее через $E_0(t)$, приходим к выражению (2.39). Полученная здесь комплексная амплитуда $E_0(t)$ отображает колебания комплексной амплитуды большого числа цугов волн. Изменения $E_0(t)$ происходят медленно по сравнению со средним периодом колебаний $T_0 = 2\pi / \omega_0$.

Интенсивность полного поля квазимонохроматического излучения источника

$$I \sim \frac{1}{t_{np} - t_{np}/2} \int_{t_{np}/2}^{t_{np}} E^2(t) dt \approx \frac{1}{t_{np}} \int_{-\infty}^{\infty} E^2(t) dt.$$

Пределы интегрирования распространены до бесконечности, так как время наблюдения t_{np} много больше длительности цуга τ . Используя равенство Парсеваля (2.8), после некоторых преобразований можно получить

$$I \sim \frac{1}{t_{np}} \int_{-\infty}^{\infty} E^2(t) dt = \frac{N}{t_{np}} \cdot \frac{1}{\pi} \int_0^{\infty} |E(\omega)|^2 d\omega.$$

Откуда спектральная плотность интенсивности излучения источника

$$I(\omega) = \frac{N}{t_{\text{пр}}} |E(\omega)|^2.$$

Таким образом, спектральная плотность интенсивности излучения всего источника равна спектральной плотности энергии излучения отдельного цуга, умноженной на среднее число цугов $N / t_{\text{пр}}$, проходящих через точку наблюдения за единицу времени. Тем самым мы показали, что спектр излучения всего квазимонохроматического источника определяется спектром излучения отдельного волнового цуга, излучаемого источником.

Любое реальное излучение ограничено во времени. Ограниченному во времени электромагнитному импульсу длительности τ соответствует в пространстве волновой пакет протяженности

$$\Delta z = c\tau \approx 2\pi c / \Delta\omega = 2\pi / \Delta k,$$

где использованы соотношения (2.13) и $k = \omega / c$. Поэтому на практике обычно приходится иметь дело не с монохроматическими, а с квазимонохроматическими волнами (волновыми пакетами). Представление о квазимонохроматической волне можно использовать для описания большинства реальных волновых процессов.

2.5. Пучок света конечной ширины

Продолжим развитие концепции группы волн. Выше мы показали, что цуг волн конечной длины не является строго монохроматической волной, а, согласно теореме Фурье, может рассматриваться как совокупность монохроматических волн с различными волновыми числами. Если цуг достаточно длинный, то интервал значений волновых чисел Δk волн, образующих цуг,

очень мал ($\Delta k \ll k_0$), и такую совокупность, как отмечалось в п. 2.4, называют группой волн.

Все световые пучки, с которыми приходится иметь дело на практике, ограничены не только в направлении их распространения, но и в обоих поперечных направлениях. Покажем, что пучок света конечной ширины нельзя представить одной плоской волной и его следует рассматривать как результат сложения совокупности плоских волн, распространяющихся в различных направлениях. Если ширина пучка много больше длины волны, то его угловое расхождение (апертура) мало, и образуется группа плоских волн в указанном выше смысле. Рассмотрим для простоты световой пучок, который ограничен в направлении оси X и не ограничен в направлении оси Y . Запишем выражение неограниченной плоской волны, распространяющейся в направлении, перпендикулярном оси Y , и под углом θ к оси Z :

$$\begin{aligned} E_0 \exp[-i(\omega t - k \sin \theta x - k \cos \theta z)] &= \\ &= E_0 \exp[-i(\omega t - k_x x - k_z z)]. \end{aligned}$$

Будем рассматривать группу волн, у которых направления распространения лежат в интервале углов, соответствующих значениям x -компоненты волнового вектора \mathbf{k} в интервале от k_x до $k_x + dk_x$. Амплитуды волн, составляющих эту группу, будут $E(k_x)dk_x/2\pi$, а результирующее возмущение для рассматриваемой бесконечно малой группы волн

$$dE(x, z, t) = \frac{1}{2\pi} E(k_x) \exp[-i(\omega t - k_x x - k_z z)] dk_x.$$

Рассмотрим далее систему волн, основное направление распространения которых совпадает с осью Z в пределах углов θ , соответствующих значениям k_x от $-\Delta k_x/2$ до $\Delta k_x/2$. (Ось Z при этом является осью конуса с малым телесным углом при

вершине, в котором заключен рассматриваемый нами пучок света.)

Результирующее возмущение можно записать в виде

$$E(x, z, t) = \frac{1}{2\pi} \int_{-\Delta k_x/2}^{\Delta k_x/2} E(k_x) \exp[i(k_x x + k_z z - \omega t)] dk_x.$$

Положив здесь $z = 0$ и $t = 0$, получим поле в плоскости $z = 0$ в момент времени $t = 0$:

$$E(x) = \frac{1}{2\pi} \int_{-\Delta k_x/2}^{\Delta k_x/2} E(k_x) e^{ik_x x} dk_x. \quad (2.45)$$

Эта формула позволяет найти световое возмущение в различных точках оси X при известной функции $E(k_x)$ и ширине интервала пространственных частот Δk_x . Если считать, что функция $E(k_x) = 0$ для всех значений k_x вне интервала $(-\Delta k_x / 2, \Delta k_x / 2)$, то пределы интегрирования в интеграле (2.45) можно распространить до бесконечности. Тогда получим

$$E(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} E(k_x) e^{ik_x x} dk_x. \quad (2.46)$$

Это выражение представляет собой разложение функции $E(x)$ в интеграл Фурье, поэтому можно написать

$$E(k_x) = \int_{-\infty}^{\infty} E(x) e^{-ik_x x} dx. \quad (2.47)$$

Для строго параллельного пучка света, распространяющегося вдоль оси Z , функция $E(k_x)$ равна нулю для любых значений k_x , кроме значения $k_x = 0$. Из соотношения (2.45) следует, что в

этом случае поле $E(x)$ не зависит от x . Смещение одинаково для всех точек на оси X , а это означает, что световой пучок является бесконечно широким в направлении оси X . А поскольку он (по предположению) имеет бесконечную ширину и в направлении оси Y , то этот пучок можно считать плоской волной. Предположим теперь, что пучок не является строго параллельным, но

$$E(k_x) = C = \text{const} \text{ при } -\Delta k_x / 2 < k_x < \Delta k_x / 2,$$

$$E(k_x) = 0 \text{ при всех значениях } k_x \text{ вне указанного интервала,}$$

т.е. на плоскости k_x поле E имеет форму прямоугольника высотой C и шириной Δk_x . Используя формулу (2.45), найдем распределение поля вдоль оси X :

$$\begin{aligned} E(x) &= \frac{1}{2\pi} C \int_{-\Delta k_x/2}^{\Delta k_x/2} e^{ik_x x} dk_x = \frac{C}{\pi} \frac{\sin(x\Delta k_x/2)}{x} = \\ &= \frac{1}{2\pi} C \Delta k_x \frac{\sin(x\Delta k_x/2)}{x\Delta k_x/2}. \end{aligned}$$

Функция $E(x)$ имеет значительную величину (порядка единицы) в интервале значений $x\Delta k_x/2$ от $-\pi/2$ до $\pi/2$, т.е. в интервале между точками $x_1 = -\pi/\Delta k_x$ и $x_2 = \pi/\Delta k_x$. Если Δk_x мало, то x велико, и пучок имеет большую ширину. Наоборот, при больших значениях Δk_x величина x мала, и пучок получается узким. Расстояние между точками $x = \pm \pi/\Delta k_x$, т.е. величину $x_2 - x_1$, можно принять за ширину пучка Δx . Тогда получим

$$\Delta x \Delta k_x \approx 2\pi.$$

Это есть соотношение неопределенностей для ширины пучка вдоль оси X и соответствующей ширины интервала волновых чисел (а значит, и ширины интервала углов, в пределах которого заключен световой пучок). Все сказанное выше о соотношении между шириной пучка и соответствующей шириной волновых чисел вытекает из этого соотношения. В общем случае трехмерных волновых пакетов, размеры которых вдоль осей X , Y , Z соответственно равны Δx , Δy , Δz , для интервалов Δk_x , Δk_y , Δk_z , в которых заметно отличны от нуля фурье-компоненты поля, можно записать

$$\Delta x \Delta k_x \approx 2\pi, \quad \Delta y \Delta k_y \approx 2\pi, \quad \Delta z \Delta k_z \approx 2\pi.$$

Таким образом, произведение размеров волнового пакета на соответствующий разброс проекции волнового вектора есть величина постоянная и приближенно равная 2π . Из этих соотношений следует, что для получения световой волны, близкой к плоской волне, следует использовать волновой пакет больших размеров (в пределе при $\Delta k_x \rightarrow 0$, $\Delta k_y \rightarrow 0$, $\Delta k_z \rightarrow 0$, что соответствует плоской волне, будем иметь $\Delta x \rightarrow \infty$, $\Delta y \rightarrow \infty$, $\Delta z \rightarrow \infty$).

В качестве примера применения полученных соотношений неопределенностей рассмотрим вопрос о получении плоскопараллельного пучка световых лучей конечной ширины. Попытаемся вырезать такой пучок из параллельного пучка лучей, падающих нормально на отверстие (диафрагму) D заданных размеров в непрозрачном экране (рис. 2.18). Если размер отверстия равен a , то

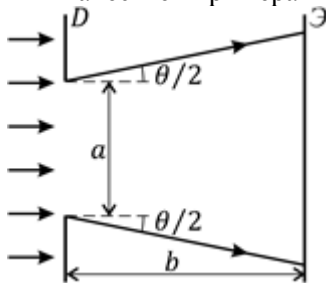


Рис. 2.18

$\Delta x = a$ и, следовательно, $\Delta k_x \approx 2\pi / a$. С другой стороны, как видно из рис. 2.18,

$$\Delta k_x = 2k \sin(\theta / 2),$$

где θ – угол расхождения пучка. А поскольку угол θ мал, то $\Delta k_x \approx 2\pi\theta / \lambda$. Откуда находим $\theta \approx \lambda / a$. Из этого соотношения видно, что строгая плоскопараллельность, при которой угол расхождения пучка $\theta = 0$, недостижима. При этом, чем меньше размер отверстия, тем больше угол расхождения пучка и тем, следовательно, больше пучок отличается от плоскопараллельного. При заданном размере отверстия пучок тем больше приближается к плоскопараллельному, чем меньше длина волны используемого света (в пределе $\theta \rightarrow 0$ при $\lambda \rightarrow 0$).

Ширина пучка на расстоянии b от отверстия

$$a' = a + \theta b \approx a + \lambda b / a.$$

Эта ширина имеет минимальное значение, равное $a'_{\min} = 2\sqrt{\lambda b}$, при ширине отверстия $a = \sqrt{\lambda b}$. Уширением пучка можно пренебречь, если оно много меньше исходной ширины пучка, т.е. когда $\lambda b \ll a^2$, или на расстоянии $b \ll a^2 / \lambda$.

Расходимость пучка обусловлена явлением дифракции света, которая возникает всегда, когда происходит ограничение фронта волны. Дифракция света будет подробно рассматриваться нами в главах 5 – 7.

Строго плоская световая волна может возбуждаться бесконечно протяженным источником света (или, как говорят, источником света с бесконечной апертурой). Но таких источников не существует, поэтому никакая реальная световая волна не является строго плоской волной. В такой волне амплитуда и фаза не являются постоянными, а изменяются в плоскости, перпендикулярной направлению распространения света, –

зависят от координат x и y . Тем не менее, оказывается возможным получить волну, обладающую свойствами, близкими к свойствам плоской волны, и от источника, имеющего конечную апертуру a , если эта апертура много больше длины световой волны λ ($a \gg \lambda$). В этом случае мы имеем дело с волнами, близкими к плоским волнам. Такие волны называют квазиплоскими волнами. На основании проведенного выше рассмотрения можно считать, что плоский фронт волны будет сохраняться (расхождением пучка можно пренебречь) на расстояниях $z \ll b$, где $b = a^2 / \lambda$ – так называемая дифракционная длина светового пучка.

Глава 3

ИНТЕРФЕРЕНЦИЯ СВЕТА

3.1. Интерференция световых монохроматических волн

Вопросы, рассмотренные в этой главе, непосредственного отношения к фурье-оптике не имеют. Однако изучение их необходимо для понимания когерентности. Кроме того, полученные здесь результаты будут использованы нами при изучении дифракции на периодических структурах и голографии.

Пусть в некоторой точке пространства P происходит наложение двух монохроматических световых волн, частоты которых соответственно равны ω_1 и ω_2 . Будем считать, что создаваемые этими волнами колебания электрических векторов \mathbf{E}_1 и \mathbf{E}_2 происходят в одном направлении. В этом случае векторный характер полей \mathbf{E}_1 и \mathbf{E}_2 можно не учитывать и считать, что

$$E_1(t) = E_{01} \exp[-i(\omega_1 t + \varphi_1)],$$

$$E_2(t) = E_{02} \exp[-i(\omega_2 t + \varphi_2)],$$

где фаза $\varphi_i = -\mathbf{k}_i \cdot \mathbf{r} + \alpha_i$ – для плоских волн и $\varphi_i = -k_i r + \alpha_i$ – для сферических волн, \mathbf{r} – радиус-вектор, определяющий положение в пространстве точки наблюдения P , r_i – расстояние, проходимое i -ой волной до точки наложения волн P , $i = 1, 2$. В

соответствии с принципом суперпозиции электрических полей, для результирующей напряженности поля в точке P будем иметь

$$E(P) = E_1(P) + E_2(P).$$

Принцип суперпозиции является следствием линейности уравнений Максвелла и вытекающих из них общего волнового уравнения и уравнения Гельмгольца: если поля $\mathbf{E}_1(\mathbf{r}, t)$ и $\mathbf{E}_2(\mathbf{r}, t)$ являются решением волнового уравнения, то его решением является и линейная комбинация этих полей, в частности, их сумма.

Так как частота света составляет примерно 10^{14} Гц, то используемые на практике детекторы светового излучения, например фотопленка, фотопластинка, фотодиоды, а также сетчатка глаза, вследствие своей инерционности не в состоянии реагировать на столь быстрые изменения светового поля. Обычно они не позволяют следить и за быстрыми изменениями световых потоков, обусловленных случайными изменениями амплитуд и фаз колебаний. Поэтому все детекторы светового излучения могут измерять только величины, пропорциональные квадрату напряженности световых полей. Они реагируют на интенсивность, которая представляет собой усредненную по времени плотность потока световой энергии и которая принимается равной среднему по времени наблюдения квадрату модуля (в комплексном представлении поля) напряженности электрического поля: $I = \langle |E|^2 \rangle = \langle E \cdot E^* \rangle$. Имея это в виду, получим

$$\begin{aligned} \langle |E|^2 \rangle &= \langle |E_1 + E_2|^2 \rangle = \langle (E_1 + E_2) \cdot (E_1^* + E_2^*) \rangle = \\ &= \langle |E_1|^2 \rangle + \langle |E_2|^2 \rangle + \langle E_1 \cdot E_2^* \rangle + \langle E_1^* \cdot E_2 \rangle. \end{aligned}$$

Здесь четвертый член является комплексно сопряженным третьему члену. Сумма таких слагаемых равна удвоенному значению реальной части одного из них:

$$\langle E_1 \cdot E_2^* \rangle + \langle E_1^* \cdot E_2 \rangle = 2 \operatorname{Re} \langle E_1 \cdot E_2^* \rangle.$$

Преобразуем это слагаемое:

$$\begin{aligned} & 2 \operatorname{Re} \langle E_1 \cdot E_2^* \rangle = \\ & = 2 |E_{01}| \cdot |E_{02}| \cdot \operatorname{Re} \langle \exp \{ -i[(\omega_1 - \omega_2)t + \varphi_1 - \varphi_2] \} \rangle = \\ & = 2 |E_{01}| \cdot |E_{02}| \langle \cos \delta \rangle, \end{aligned}$$

где $\delta = (\omega_1 - \omega_2)t + \varphi_1 - \varphi_2$ – разность фаз складываемых в точке P волн. Следовательно,

$$\langle |E|^2 \rangle = \langle |E_1|^2 \rangle + \langle |E_2|^2 \rangle + 2 |E_{01}| \cdot |E_{02}| \langle \cos \delta \rangle.$$

К такому же выражению можно прийти, используя формулу

$$\exp(i\alpha) + \exp(-i\alpha) = 2 \cos \alpha.$$

Учитывая, что $|E_{01}|^2 = I_1$, $|E_{02}|^2 = I_2$, $|E_{01}| \cdot |E_{02}| = \sqrt{I_1 I_2}$, для интенсивности света в точке P будем иметь

$$I(P) = I_1(P) + I_2(P) + I_{12}(P),$$

где

$$I_{12}(P) = 2 \sqrt{I_1(P) I_2(P)} \langle \cos \delta \rangle.$$

Слагаемое $I_{12}(P)$ называют интерференционным членом. Им определяются все интерференционные эффекты, связанные с характером изменения интенсивности в пространстве при наложении волн.

Если $\omega_1 \neq \omega_2$, то разность фаз складываемых волн δ будет зависеть от времени: $\delta = \delta(t)$. Среднее значение косинуса, а с ним и интерференционный член $I_{12}(P)$, обратятся в нуль и тогда

$$I(P) = I_1(P) + I_2(P),$$

т.е. имеет место суммирование интенсивностей. Если же $\omega_1 = \omega_2$ и разность начальных фаз $\alpha_1 - \alpha_2$ не зависит от времени (для монохроматических волн второе условие всегда имеет место), то $\langle \cos \delta \rangle = \cos \delta$ – тождественно не равен нулю. Тождественно не равным нулю будет и интерференционный член $I_{12}(P)$. В этом случае

$$I(P) = I_1(P) + I_2(P) + 2\sqrt{I_1(P)I_2(P)} \cos \delta, \quad (3.1)$$

где разность фаз $\delta = \varphi_1 - \varphi_2$ не зависит от времени. В зависимости от величины разности фаз δ , которая определяется положением точки наблюдения P , множитель $\cos \delta$ в интерференционном члене $I_{12}(P)$ может принимать как положительные, так и отрицательные значения. Поэтому результирующая интенсивность $I(P)$ в точке P будет либо больше, либо меньше суммы интенсивностей складываемых волн. В одних точках волны при наложении будут усиливать друг друга, а в других – ослаблять. В точках, для которых $\cos \delta = 1$, разность фаз $\delta = 2m\pi$, где m – целое число, включая нуль, интенсивность будет достигать максимального значения

$I_{\max}(P) = (\sqrt{I_1} + \sqrt{I_2})^2$, а для которых $\cos \delta = -1$, разность фаз $\delta = (2m + 1)\pi$, – достигать минимального значения

$I_{\min}(P) = (\sqrt{I_1} - \sqrt{I_2})^2$. При этом при равных интенсивностях ($I_1 = I_2 = I_0$) $I_{\max} = 4I_0$, а $I_{\min} = 0$.

Таким образом, при сложении двух монохроматических световых волн результирующая интенсивность в каждой точке пространства P равна сумме интенсивностей складываемых волн, если их частоты различны, и отличается от суммы интенсивностей, если частоты волн одинаковы. Во втором случае в зависимости от конкретного значения разности фаз $\delta = \varphi_1 - \varphi_2$,

которое определяется положением точки наблюдения P , интенсивность $I(P)$ может принимать различные, не равные сумме интенсивностей, значения. Имеет место перераспределение интенсивности света в пространстве так, что в одних точках пространства световые волны усиливают друг друга, а в других ослабляют. Это явление устойчивого во времени перераспределения колебаний светового поля в пространстве, в результате которого в одних местах колебания усиливаются, а в других ослабляются, называется интерференцией света. Область перекрытия волн называется интерференционным полем. В точках, где волны усиливают друг друга, образуются максимумы интенсивности света, а где ослабляют – минимумы интенсивности. Их называют соответственно интерференционными максимумами и интерференционными минимумами. Интенсивность всех интерференционных максимумов одинакова. На экране, расположенном подходящим образом в области наложения волн (в волновом поле), этим максимумам и минимумам будут соответствовать светлые и темные полосы, называемые интерференционными полосами. Расстояние h между соседними максимумами или минимумами интенсивности (светлыми или темными полосами) называется шириной интерференционной полосы. Таким образом, $h = x_m - x_{m-1}$, где x_i ($i = m, m - 1$) – координата i -го максимума или минимума. Устойчивую во времени картину чередования темных и светлых полос называют интерференционной картиной. Конкретный вид и форма интерференционных полос, а значит, и вид интерференционной картины зависят от интерферирующих волн, т.е. от комплексных амплитуд $E_1(\mathbf{r})$ и $E_2(\mathbf{r})$, или, иначе, – от пространственного распределения амплитуд и фаз интерферирующих волн.

Соотношение (3.1) в теории двухлучевой интерференции монохроматических волн является основным. Из него следует, что интерференционный эффект определяется зависимостью разности фаз δ от положения точки наблюдения. Чередование темных и

светлых полос, характерное для интерференции света, связано с изменением знака $\cos\delta$.

Световые волны, способные к интерференции, называются когерентными, в противном случае – некогерентными. При наложении когерентных волн в точке P устанавливаются гармонические колебания светового поля с постоянной амплитудой, различной в разных точках, а при наложении некогерентных – негармонические колебания с нерегулярно изменяющимися во времени амплитудами. Из сказанного ясно, что световые волны являются когерентными, если их разность фаз δ не зависит от времени. Для монохроматических волн это имеет место, если их частоты одинаковы.

Полученные выше соотношения для двух источников можно обобщить на случай многих источников. Пусть свет в некоторую точку пространства приходит от большого числа источников. Тогда если источники света являются когерентными, то результирующая интенсивность в указанной точке

$$I_{\text{ког}} = \left\langle \left| \sum_i E_i \right|^2 \right\rangle,$$

а если некогерентными –

$$I_{\text{неког}} = \left\langle \sum_i |E_i|^2 \right\rangle.$$

Суммирование в обоих случаях происходит по всем источникам. В результате можно сформулировать следующее утверждение: при когерентных источниках суммируются амплитуды полей и модуль суммы возводится в квадрат; при некогерентных источниках суммируются квадраты модулей, т.е. интенсивности полей. Следует при этом отметить, что переход от комплексной амплитуды к интенсивности, регистрируемой прибором, происходит только в детекторе светового излучения.

Рассмотрим несколько характерных случаев интерференции монохроматических волн всех видов (плоских, сферических, цилиндрических).

3.2. Интерференция волн различных видов

3.2.1. Интерференция плоских волн

Рассмотрим сначала интерференцию двух плоских монохроматических волн одинаковых частот, распространяющихся в одном направлении. Пусть направление распространения волн совпадает с положительным направлением оси Z . Комплексные амплитуды этих волн будут иметь вид

$$E_1 = E_{01} \exp[i(kz - \alpha_1)], \quad E_2 = E_{02} \exp[i(kz - \alpha_2)],$$

где E_{01} и E_{02} – амплитуды волн, а α_1 и α_2 – их начальные фазы. При интерференции волн складываются их комплексные амплитуды, $E = E_1 + E_2$, а интенсивность находится как квадрат модуля результирующего поля:

$$I = |E_1 + E_2|^2.$$

На основании этого получим

$$I = E_0^2 = E_{01}^2 + E_{02}^2 + 2E_{01}E_{02} \cos(\alpha_2 - \alpha_1).$$

Как видим, при сложении плоских волн, распространяющихся в одном направлении, результирующая интенсивность не зависит от положения точки наблюдения (не зависит от координаты z), т.е. оказывается одинаковой во всех точках интерференционного поля. Значение интенсивности результирующего поля определяется амплитудами складываемых волн и разностью их начальных фаз $\delta = \alpha_2 - \alpha_1$. Если эта разность фаз $\delta = 2m\pi$, где m – целое положительное или отрицательное число, т.е. если складываемые волны находятся в одинаковой фазе, то $\cos\delta = 1$, интенсивность во всех точках будет иметь максимальное

значение, волны во всех точках будут усиливать друг друга. Амплитуда результирующей волны будет постоянна и равна сумме амплитуд складываемых волн: $E_0 = E_{01} + E_{02}$. Если же разность фаз $\delta = (2m + 1)\pi$, т. е. если складываемые волны находятся в противофазе, то $\cos\delta = -1$, интенсивность во всех точках будет иметь минимальное значение, волны всюду будут ослаблять друг друга. Амплитуда результирующей волны в этом случае будет равна разности амплитуд складываемых волн: $E_0 = E_{01} - E_{02}$. В этом случае при равенстве амплитуд складываемых волн результирующая амплитуда $E_0 = 0$ – волны при наложении будут полностью гасить друг друга, результирующая волна не возникает.

Рассмотрим теперь случай, когда источники S_1 и S_2 плоских волн расположены в точках с координатами z_1 и z_2 . Начальные фазы волн будем считать равными нулю. В этом случае

$$E_1 = E_{01} \exp[ik(z - z_1)], \quad E_2 = E_{02} \exp[ik(z - z_2)],$$

а интенсивность

$$I = E_0^2 = E_{01}^2 + E_{02}^2 + 2E_{01}E_{02} \cos kd,$$

где $d = z_2 - z_1$ – расстояние между источниками волн. Теперь разность фаз $\delta = kd$ зависит от расстояния между источниками волн. Мы видим, что и в этом случае амплитуда не зависит от положения точки, в которой мы исследуем результирующее колебание. Все точки колеблются с одинаковой амплитудой, зависящей лишь от расстояний между точками, в которых фазы складываемых волн равны нулю (точки S_1 и S_2). Амплитуда будет максимальной, равной амплитуде складываемых волн, когда расстояние между источниками будет таким, что $\delta = 2m\pi$. При этом $d = m\lambda$. Минимальной амплитуда будет, когда расстояние между источниками будет удовлетворять условию $\delta = (2m + 1)\pi$, т.е. когда $d = (m + 1/2)\lambda$. Минимальное значение амплитуды будет равно разности амплитуд складываемых волн.

Физически интересный случай реализуется при наложении волн одинаковых амплитуд, распространяющихся навстречу друг другу. Будем считать, что начальные фазы волн равны нулю. Тогда результирующая комплексная амплитуда

$$E_0 = E_{00} \exp(ikz) + E_{00} \exp(-ikz),$$

где первое и второе слагаемые описывают комплексные амплитуды волн, распространяющихся соответственно в положительном и отрицательном направлениях оси Z . Откуда

$$E(z) = 2E_{00} \cos kz.$$

Если учесть временной фазовый множитель $\exp(-i\omega t)$, то для результирующего поля в точке z в момент времени t будем иметь

$$E(z, t) = 2E_{00} e^{-i\omega t} \cos kz,$$

или

$$E(z, t) = 2E_{00} \cos kz \cos \omega t. \quad (3.2)$$

Из этой формулы видно, что вектор напряженности электрического поля результирующей волны в каждой точке совершает гармонические колебания с частотой ω и амплитудой

$$E_0(z) = 2E_{00} |\cos kz|.$$

При z , удовлетворяющем условию $kz = m\pi$, где $m = 0, \pm 1, \pm 2, \dots$, т.е. в плоскостях

$$z = z_m = m (\pi / k) = m(\lambda / 2),$$

в которых $|\cos kz| = 1$, амплитуда результирующей волны имеет максимальное значение, равное $2E_{00}$. В плоскостях же

$$z = z_m = (m + 1 / 2) (\pi / k) = (m + 1 / 2) (\lambda / 2),$$

в которых $|\cos kz| = 0$, амплитуда результирующей волны имеет минимальное значение, равное нулю.

Таким образом, при наложении двух когерентных световых волн одинаковой амплитуды и распространяющихся в противоположных направлениях образуется волна, у которой амплитуда в одних плоскостях принимает максимальное значение, равное удвоенному значению амплитуды складываемых волн, а в других – минимальное значение, равное нулю; фаза колебаний вектора электрического поля \mathbf{E} во всех точках между соседними плоскостями максимума или минимума одинакова, а колебания по разные стороны от этих плоскостей происходят в противофазе. Такая волна называется стоячей электромагнитной волной. В волновом процессе, описываемом выражением (3.2), отсутствуют перемещающиеся волновые поверхности (отсутствует аргумент вида $t \pm z/c$), поэтому волна вида (3.2) называется стоячей волной. Плоскости, в которых амплитуда имеет максимальное значение, называются пучностями стоячей волны, а плоскости, в которых амплитуда минимальна, называются узлами стоячей волны. Расстояние между соседними пучностями или узлами одинаково и равно $z_{m+1} - z_m = \lambda/2$ – половине длины волны. Расстояние между соседним узлом и пучностью равно $m\lambda/2 - (m+1/2)\lambda/2 = \lambda/4$ – четверти длины волны. Расстояние же между соседними узлами или пучностями

$$h = z_{m+1} - z_m = \lambda/2,$$

где z_i ($i = m, m+1$) – координата i -го узла или i -ой пучности, равно половине длины волны.

Практически стоячие волны возникают в результате сложения падающих и отраженных от преград волн. Поэтому необходимым условием образования стоячих волн является наличие границ, отражающих падающие на них волны. Большой интерес представляют волны, которые образуются при многократных отражениях. Этот случай реализуется при распространении волн в ограниченном и замкнутом объеме. Стоячая волна возникает в результате наложения двух волн, бегущих в противоположных направлениях, отражающихся от границ и снова бегущих

навстречу друг другу. Если отражение происходит без потери энергии, то отраженная волна будет иметь ту же амплитуду и ту же частоту, что и падающая. Поэтому комплексная амплитуда должна иметь вид

$$E_-(z) = E_{00} \exp(-ikz + i\theta).$$

Наличие постоянной θ учитывает возможное изменение фазы при отражении волны от границы. Ее значение будет установлено ниже. Результирующая комплексная амплитуда

$$E_0 = E_{00} \exp(ikz) + E_{00} \exp(-ikz + i\theta).$$

Отсюда находим

$$|E_0|^2 = 4E_{00}^2 \cos^2(kz - \theta/2).$$

Амплитуда стоячей волны

$$E(z) = 2E_{00} |\cos(kz - \theta/2)|.$$

Будем считать, что отражающие границы расположены в плоскостях $z = 0$ и $z = l$. В точках отражения колебания поля не происходит, поэтому эти точки являются узлами стоячей волны, а значит, амплитуды на отражающих границах равны нулю. Тогда из условия

$$E(0) = 2E_{00} |\cos(\theta/2)| = 0$$

находим, что $\theta = \pi$ – фаза волны при отражении меняется на противоположную. С учетом этого амплитуда стоячей волны примет вид

$$E(z) = 2E_{00} |\sin kz|.$$

Комплексная амплитуда

$$E_0 = E_{00} [\exp(ikz) + \exp(-ikz + i\pi)] = E_{00} [\exp(ikz) - \exp(-ikz)] = i2E_{00} \sin kz.$$

Для точек, координата z которых удовлетворяет условию $0 \leq kz \leq \pi$, т.е. $0 \leq z \leq \lambda / 2$, $\sin kz \geq 0$, и поэтому в этой области

$$E_0 = i2E_{00} \sin kz = \exp(i\pi/2) 2E_{00} \sin kz.$$

Колебания поля в этом интервале

$$\begin{aligned} E(z, t) &= 2E_{00} \sin kz \exp(i\pi/2) \exp(-i\omega t) = \\ &= 2E_{00} \sin kz \exp[-i(\omega t - \pi/2)], \end{aligned}$$

или

$$E(z, t) = 2E_{00} \sin kz \sin \omega t.$$

Для точек из интервала $\lambda / 2 \leq z \leq \lambda$ имеем $\sin kz \leq 0$, и тогда

$$E_0 = -i2E_{00} \sin kz = \exp(-i\pi/2) 2E_{00} \sin kz.$$

В этом промежутке

$$E(z, t) = -2E_{00} \sin kz \sin \omega t = 2E_{00} \sin kz \sin(\omega t - \pi).$$

Сравнивая формулы колебаний поля в двух рассмотренных интервалах, видим, что разность фаз колебаний постоянна и по абсолютному значению равна π . Это означает, что колебания в одном из интервалов совершаются навстречу колебаниям в другом.

Условие равенства нулю амплитуды на границе $z = l$ приводит к уравнению

$$\sin kl = 0.$$

Отсюда получаем, что

$$kl = m\pi,$$

где m – произвольное положительное число, исключая нуль. Отрицательные m исключаются по смыслу волнового числа. Последнее равенство можно переписать в виде

$$l = m \frac{\lambda}{2}.$$

Из этого соотношения следует, что в ограниченном пространстве возникают только такие стоячие волны, половина длины волны которых укладывается на линейном размере l этого пространства целое число раз. При $m = 1$ на указанном размере укладывается всего одна полуволна.

Рассмотренные колебания в ограниченном пространстве называют собственными колебаниями или модами. Частоты собственных колебаний

$$\omega = m(\pi v / l), \quad m = 1, 2, 3, \dots$$

Число собственных частот бесконечно. Интервалы между ними одинаковы и равны $\pi v / l$. Наименьшая частота $\omega_1 = \pi v / l$ называется основной частотой. Отметим, что возникновение собственных колебаний связано с выполнением определенных условий на границах области (граничных условий).

Стоячие волны образуются, в частности, в резонаторах оптических квантовых источников света – лазерах.

Стоячие волны можно зарегистрировать при помощи соответствующей светочувствительной среды, например фотопластинки, на которой пучности регистрируются в виде почернений, а узлы – в виде прозрачных неэкспонированных участков.

Стоячие электромагнитные волны обладают важным свойством: в них пучности электрического вектора **E** совпадают с узлами магнитного вектора **B**, а узлы вектора **E** – с пучностями вектора **B**. Таким образом, в стоячей волне оказывается возможным пространственно разделить электрическое и магнитное поля.

Рассмотрим теперь случай, когда две плоские волны сходятся под углом 2θ друг к другу (рис. 3.1); начальные фазы волн будем считать равными нулю. Проведем координатную ось Z вдоль

биссектрисы угла между волновыми векторами \mathbf{k}_1 и \mathbf{k}_2 этих волн. Тогда плоскость XY будет плоскостью симметрии волновых фронтов 11 и 22, векторы \mathbf{k} волн образуют с осью Z углы, равные $\pm\theta$. Комплексные амплитуды этих волн будут

$$E_1(x, z) = E_{01} \exp [ik(x\sin\theta + z\cos\theta)],$$

$$E_2(x, z) = E_{02} \exp [ik(-x\sin\theta + z\cos\theta)],$$

где учтено, что направляющий косинус вдоль оси Z первой волны равен $\cos\theta$, а вдоль оси X он равен $\sin\theta$, для второй волны соответствующие направляющие косинусы равны $\cos(-\theta) = \cos\theta$ и $\cos(\pi/2 + \theta) = -\sin\theta$.

В дальнейшем нас будет интересовать распределение интенсивности поля в плоскости $z = 0$. Комплексная амплитуда поля в этой плоскости

$$E(x) = E_{01} \exp(ikx\sin\theta) + E_{02} \exp(-ikx\sin\theta),$$

а интенсивность

$$I(x) = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos k\Delta(x), \quad (3.3)$$

где

$$\Delta(x) = 2x\sin\theta$$

– разность хода суммируемых волн. Заметим, что разность фаз $k\Delta(x)$ интерферирующих волн изменяется пропорционально координате x . Точно такое же выражение для интенсивности получается и для любой плоскости $z = \text{const} \neq 0$, так как в разности фаз складываемых волн член $kz\cos\theta$ исчезает. Из выражения

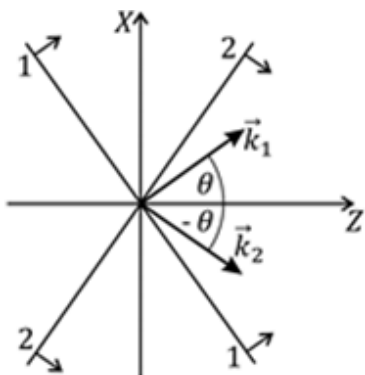


Рис. 3.1

(3.3) видно, что при интерференции двух направленных под углом друг к другу плоских волн возникает интерференционная картина с косинусоидальным распределением интенсивности. График зависимости $I(x)$ показан на рис. 3.2. Интенсивность изменяется по закону косинуса между значениями $I_{\min} = |E_{01} - E_{02}|^2$

и $I_{\max} = |E_{01} + E_{02}|^2$. В любой

плоскости, перпендикулярной оси Z , эта интерференционная картина представляет собой систему параллельных друг другу темных и светлых полос с непрерывным переходом от темной полосы к светлой полосе. При этом, поскольку интенсивность не зависит от координаты y , интерференционные полосы располагаются параллельно оси Y . Положение максимумов интенсивности определяется соотношением

$$\Delta(x) = (2\pi / \lambda) 2x_m \sin \theta = 2m\pi,$$

где m – целое число, включая нуль. Откуда x -координата максимумов

$$x_m = m\lambda / 2\sin \theta.$$

Минимумы интенсивности имеют x -координаты

$$x_m = (2m + 1)\lambda / 4\sin \theta.$$

Расстояние между соседними интерференционными полосами (пространственный период распределения)

$$h = x_{m+1} - x_m = \lambda / 2\sin \theta.$$

Это соотношение можно переписать в виде

$$2h \sin \theta = \lambda.$$

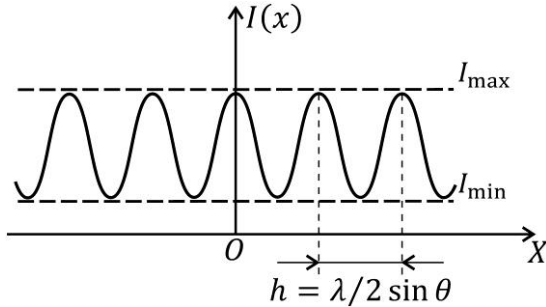


Рис. 3.2

Количество полос, приходящихся на единицу длины вдоль оси X ,

$$N_x = 1 / h = 2 \sin \theta / \lambda,$$

равно удвоенной пространственной частоте u волны. Заметим, что если угол θ возрастает, то период h уменьшается; число полос на единицу длины N_x при этом возрастает. Положив $\sin \theta = 1$ (что имеет место при угле схождения волн $2\theta = \pi$), получим минимальную ширину интерференционных полос: $h_{\min} = \lambda / 2$. Именно таким является расстояние между соседними узлами или пучностями в стоячей волне (угол схождения волн 2θ , равный π , как раз и соответствует образованию стоячей волны). Число интерференционных полос, приходящихся на единицу длины, в этом случае будет максимальным: $N_{x \max} = 2 / \lambda$.

Если интенсивности волн равны, $I_1 = I_2 \equiv I_0$, а угол θ мал ($\sin \theta \approx \theta$), то

$$I(x) = 2I_0(1 + \cos 2kx\theta) = 4I_0 \cos^2(kx\theta).$$

Это выражение можно переписать в виде

$$I(x) = 2I_0 + I_0 \exp(ikx2\theta) + I_0 \exp(-ikx2\theta). \quad (3.4)$$

В случае, когда одна плоская волна распространяется вдоль оси Z , а другая под углом θ к ней, результирующее поле в плоскости $z = 0$, создаваемое этими волнами, будет

$$E(x) = E_{01} + E_{02} \exp(ikx \sin \theta).$$

Распределение интенсивности в любой плоскости, перпендикулярной оси Z , будет определяться тем же выражением (3.3), в котором $\Delta(x) = x \sin \theta$. Положение максимумов интенсивности определяется соотношением $\Delta(x) = x_m \sin \theta = m\lambda$, $m = 0, 1, 2, \dots$. Из этого соотношения находим расстояние между соседними полосами (ширину полосы):

$$h = \lambda / \sin \theta. \quad (3.5)$$

Рассмотрим теперь распределение интенсивности светового поля вдоль оси Z по толщине некоторого слоя пространства. Указанное распределение интенсивности будет иметь вид

$$\begin{aligned} I(x, z) &= |E_{01} \exp(ikz) + E_{02} \exp[ik(x \sin \theta + z \cos \theta)]|^2 = \\ &= E_{01}^2 + E_{02}^2 + 2E_{01}E_{02} \cos\{k[x \sin \theta - z(1 - \cos \theta)]\}. \end{aligned}$$

Максимумы интенсивности в этом случае располагаются вдоль эквидистантных плоскостей

$$x \sin \theta - z(1 - \cos \theta) = m\lambda,$$

где m – целое число. Эти плоскости составляют с осью Z угол α , определяемый соотношением

$$\operatorname{tg} \alpha = dx/dz = (1 - \cos \theta) / \sin \theta = \operatorname{tg}(\theta/2).$$

Следовательно, $\alpha = \theta/2$, т.е. плоскости интерференционных максимумов располагаются по биссектрисе между направлениями интерферирующих волн (рис. 3.3). Для любой фиксированной

плоскости $z = \text{const}$ угол θ одинаков, поэтому ширина интерференционной полосы $h = \Delta x$ в любой плоскости $z = \text{const}$ также одинакова и равна $h = \lambda / \sin \theta$ (формула (3.5)).

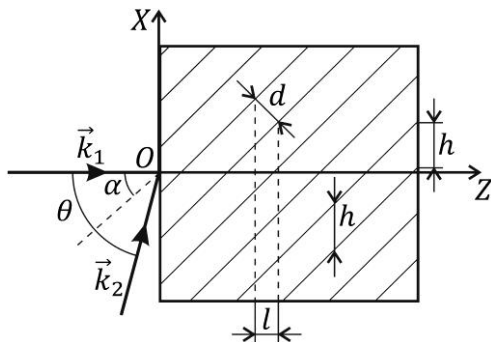


Рис.3.3

Расстояние d между плоскостями интерференционных максимумов, как видно из рис. 3.3,

$$d = h \cos \alpha = \lambda / [2 \sin(\theta / 2)].$$

Откуда

$$2d \sin \alpha = \lambda. \quad (3.6)$$

При переходе от одной плоскости $z = \text{const}$ к другой происходит сдвиг интерференционных полос. Сдвиг на ширину полосы, как видно на том же рис. 3.3, происходит при расстоянии $l = \Delta z$ между плоскостями $z = \text{const}$, равном

$$l = h / \tan \alpha = \lambda / (1 - \cos \theta).$$

При смещении вдоль оси Z на расстояние $l/2$, максимум интерференционной картины сменяется минимумом, и наоборот.

Представляет также интерес рассмотреть интерференцию плоских монохроматических волн, распространяющихся под углом друг к другу, и в векторном виде. Пусть угол между волновыми векторами \mathbf{k}_1 и \mathbf{k}_2 волн равен 2θ , а положение точки

наблюдения определяется радиус-вектором \mathbf{r} . Тогда комплексная амплитуда результирующего светового поля в этой точке определится как

$$E = E_{01} \exp[i(\mathbf{k}_1 \cdot \mathbf{r} - \alpha_1)] + E_{02} \exp[i(\mathbf{k}_2 \cdot \mathbf{r} - \alpha_2)],$$

а интенсивность – как

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\mathbf{k}_{21} \cdot \mathbf{r} + \delta),$$

где $\delta = \alpha_1 - \alpha_2$, $\mathbf{k}_{21} = \mathbf{k}_2 - \mathbf{k}_1$. Поверхности равных интенсивностей определяются уравнением $\mathbf{k}_{21} \cdot \mathbf{r} = \text{const}$ и представляют собой

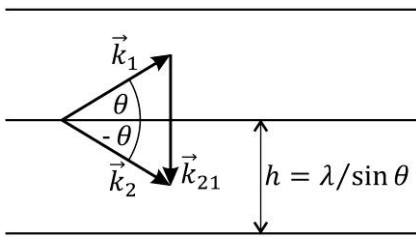


Рис. 3.4

эквиливантные плоскости, перпендикулярные вектору \mathbf{k}_{21} (рис. 3.4). При этом максимумы интенсивности лежат на плоскостях $\mathbf{k}_{21} \cdot \mathbf{r} + \delta = 2m\pi$, а минимумы – на плоскостях $\mathbf{k}_{21} \cdot \mathbf{r} + \delta = m\pi$, где m – целое

число. Расстояние h между соседними плоскостями максимальной (или минимальной) интенсивности определится условием $k_{21}h = 2\pi$, где k_{21} – модуль вектора \mathbf{k}_{21} . Поскольку модули векторов \mathbf{k}_1 и \mathbf{k}_2 одинаковы и равны $k = 2\pi / \lambda$, то модуль вектора \mathbf{k}_{21} , как видно на рис. 3.4,

$$k_{21} = 2k \sin \theta = 2 \frac{2\pi}{\lambda} \sin \theta.$$

С учетом этого находим ширину интерференционной полосы: $h = \lambda / \sin \theta$, т.е. снова приходим к формуле (3.5).

Сложение плоских волн, распространяющихся под углом друг к другу, сводится к сложению двух плоских волн, распространяющихся в одном направлении, и двух плоских волн, распространяющихся навстречу друг другу. Действительно, в

области перекрытия волн их волновые векторы можно разложить на две взаимно перпендикулярные составляющие: на составляющую вдоль биссектрисы угла между налагающимися волнами и на составляющую, перпендикулярную этой биссектрисе. Первые составляющие распространяются в одном направлении, а вторые – в противоположном.

3.2.2. Интерференция сферической и цилиндрической волн

Рассмотрим суперпозицию волн, излучаемых двумя точечными монохроматическими источниками S_1 и S_2 . В точке наблюдения P расходящиеся сферические волны будут иметь комплексные амплитуды

$$E_1(r_1) = \frac{E_{01}}{r_1} \exp(ikr_1), \quad E_2(r_2) = \frac{E_{02}}{r_2} \exp(ikr_2),$$

где r_1 и r_2 – расстояния от точки наблюдения P до источников S_1 и S_2 соответственно, E_{01} и E_{02} – амплитуды излучаемых источниками S_1 и S_2 волн. Интенсивность света в точке наблюдения, создаваемая этими волнами,

$$\begin{aligned} I(P) &= |E_1(r_1) + E_2(r_2)|^2 = \\ &= \frac{E_{01}^2}{r_1^2} + \frac{E_{02}^2}{r_2^2} + 2 \frac{E_{01}E_{02}}{r_1 r_2} \cos k\Delta, \end{aligned} \quad (3.7)$$

где $\Delta = r_2 - r_1$ – разность хода волн, приходящих в точку наблюдения. Начальные фазы волн полагаем равными нулю. Это не приводит к ограничению общности и сводится только к подходящему выбору начала координат. При разности хода волн $\Delta = m\lambda$ ($m = 0, \pm 1, \pm 2, \pm 3, \dots$) амплитуда суммарной волны максимальна и равна $E_0(P) = E_{01} / r_1 + E_{02} / r_2$, а при

разности хода $\Delta = (m + 1 / 2)\lambda$ – минимальна и равна $E_0(P) = E_{01} / r_1 - E_{02} / r_2$.

Две сферические волны, распространяющиеся от взаимно когерентных источников, интерферируют во всем пространстве. Поскольку интерференционные полосы удовлетворяют условию $\Delta = r_2 - r_1 = \text{const}$, то в пространстве они образуют гиперболоиды вращения. При наблюдении интерференции на каком-либо плоском экране вид интерференционных полос будет сильно зависеть от ориентации этого экрана относительно прямой, соединяющей источники S_1 и S_2 .

Интерференционное поле обладает круговой симметрией с осью симметрии, совпадающей с прямой, проходящей через оба источника. Поэтому для изучения данного поля достаточно рассмотреть интерференционную картину в двух плоскостях симметрии: в плоскости, параллельной оси симметрии, и в плоскости, перпендикулярной этой оси.

Рассмотрим сначала интерференционную картину в плоскости, содержащей ось симметрии S_1S_2 (поперечное наблюдение). Направим координатную ось Y вдоль оси симметрии, а в качестве поперечной оси координат X возьмем перпендикулярную ей ось, проведенную через середину отрезка S_1S_2 , соединяющего источники S_1 и S_2 (рис. 3.5). Пусть расстояние между источниками равно $2a$. Тогда координаты источников S_1 и S_2 в выбранной системе координат будут $(0, a, 0)$ и $(0, -a, 0)$ соответственно. Разность хода волн от этих источников до произвольной точки наблюдения $P(x, y, 0)$ будет равна

$$\Delta = r_2 - r_1 = \sqrt{x^2 + (y + a)^2} - \sqrt{x^2 + (y - a)^2}. \quad (3.8)$$

Интенсивность в этой точке определится по той же формуле (3.7), в которой разность хода определяется выражением (3.8). Условие интерференционного максимума приводит к следующему равенству

$$\sqrt{x^2 + (y + a)^2} - \sqrt{x^2 + (y - a)^2} = m\lambda,$$

где $m = 1, 2, 3, \dots$. Кривая, разность расстояний до каждой точки которой от двух заданных точек есть величина постоянная ($r_2 - r_1 = \text{const}$), представляет собой гиперболу. Произведя очевидные алгебраические преобразования, мы действительно получаем аналитическое выражение, описывающее в плоскости $z = 0$ семейство гипербол:

$$\frac{y^2}{\left(\frac{1}{2}m\lambda\right)^2} - \frac{x^2}{a^2 - \left(\frac{1}{2}m\lambda\right)^2} = 1. \quad (3.9)$$

Фокусы этих гипербол находятся в точках S_1 и S_2 , т.е. совпадают с

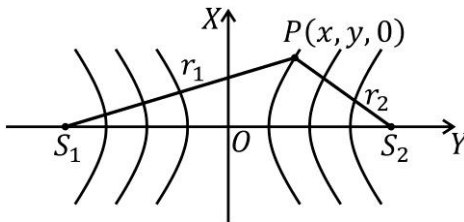


Рис. 3.5

источниками сферических волн. (Гиперболы отличаются друг от друга значением числа m .) Семейство этих гипербол показано на рис. 3.5. Если ограничиться достаточно малой

областью вблизи начала координат (вблизи оси X на рис. 3.5), то семейство гипербол можно приближенно заменить семейством параллельных прямых.

Почти во всем пространстве, кроме сравнительно небольшого участка между источниками и плоскостью симметрии, семейство гипербол можно заменить семейством их асимптот. Угловым коэффициентом асимптоты гиперболы равен отношению ее полуосей:

$$k = \text{tg}\alpha = \frac{\sqrt{a^2 - (m\lambda/2)^2}}{m\lambda/2} \approx \frac{2a}{m\lambda}.$$

На основе этого выражения можно построить асимптоты геометрическим путем.

В пространстве поверхности максимумов представляют собой двухполосные гиперболоиды вращения. Точку пересечения поверхностей максимумов с осью Y можно найти, если положить в (3.9) $x = 0$. В результате получим

$$y = m\lambda / 2. \quad (3.10)$$

Расстояние между максимумами вдоль оси Y одинаково и равно половине длины волны ($\Delta y = \lambda / 2$). Но так и должно быть, поскольку волны от источников S_1 и S_2 распространяются навстречу друг другу. Положив в равенстве (3.10) $y = a$, получим общее число интерференционных полос на полуоси гиперболы вдоль координатной оси Y :

$$m_{\max} = 2a / \lambda.$$

Сферическая волна, как уже отмечалось, может не только исходить из точечного источника (выпуклый фронт волны), но и сходить в одну точку (вогнутый фронт волны). В случае, когда одна из интерферирующих волн сходится в точку, интерференционное поле будет иметь другую картину. В этом случае происходит сложение волн

$$E_1(r_1) = \frac{E_{01}}{r_1} \exp(ikr_1), \quad E_2(r_2) = \frac{E_{02}}{r_2} \exp(-ikr_2).$$

В аргументе косинуса в третьем (интерференционном) члене в выражении (3.9) вместо разности $\Delta = r_2 - r_1$ появится сумма $r_1 + r_2$. Линии максимальной интенсивности теперь будут определяться условием $r_1 + r_2 = m\lambda = \text{const}$. Кривая, удовлетворяющая этому

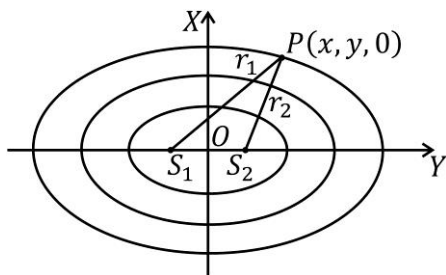


Рис. 3.6

условию, представляет собой эллипс. Поэтому интерференционная картина на экране в рассматриваемом случае будет представлять собой семейство эллипсов (полуоси которых различаются значением m) с фокусами в точках расположения источников S_1 и S_2 (рис. 3.6). Уравнения этих эллипсов будут иметь вид

$$\frac{y^2}{\left(\frac{1}{2}m\lambda\right)^2} + \frac{x^2}{\left(\frac{1}{2}m\lambda\right)^2 - a^2} = 1$$

Но при этом должно выполняться условие

$$\left(\frac{1}{2}m\lambda\right)^2 - a^2 > 0,$$

или, иначе, $m > 2a/\lambda$. В противном случае кривыми максимальной интенсивности будут гиперболы. В пространстве интерференционная картина имеет вид семейства эллипсоидов вращения.

Таким образом, интерференционные полосы будут иметь вид гипербол, когда обе волны выпуклые или вогнутые, и вид эллипсов — когда одна из волн выпуклая, а другая вогнутая.

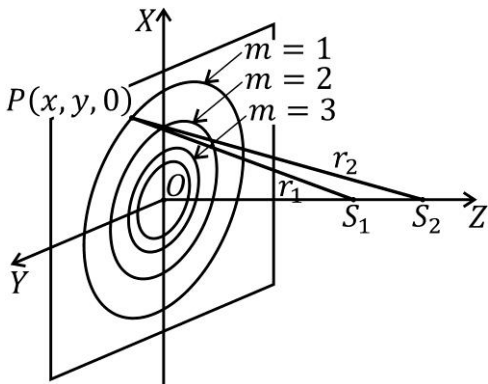


Рис. 3.7

Рассмотрим интерференционную картину в плоскости, перпендикулярной оси симметрии (продольное наблюдение). Направим ось Z вдоль линии, соединяющей источники S_1 и S_2 , а в

качестве плоскости наблюдения примем плоскость $z = 0$. Координаты источников S_1 и S_2 пусть будут $(0, 0, a)$ и $(0, 0, b)$ соответственно. Тогда разность хода волн от этих источников до точки наблюдения $P(x, y, 0)$ будет

$$\Delta = r_2 - r_1 = \sqrt{x^2 + y^2 + b^2} - \sqrt{x^2 + y^2 + a^2} .$$

Условие интерференционного максимума запишется в виде

$$\sqrt{x^2 + y^2 + b^2} - \sqrt{x^2 + y^2 + a^2} = m\lambda. \quad (3.11)$$

После преобразований, аналогичных предыдущему, получим аналитическое выражение, описывающее семейство окружностей:

$$x^2 + y^2 = \frac{[(b-a)^2 - (m\lambda)^2][(b+a)^2 - (m\lambda)^2]}{(2m\lambda)^2}. \quad (3.12)$$

Следовательно, интерференционные полосы в рассматриваемом случае имеют вид чередующихся светлых и темных колец, центры которых расположены в начале координат, а радиусы колец равны

$$R_m = \frac{\sqrt{[(b-a)^2 - (m\lambda)^2][(b+a)^2 - (m\lambda)^2]}}{2m\lambda}. \quad (3.13)$$

Из этой формулы видно, что интерференционные кольца расположены не на равных расстояниях друг от друга (неэквидистантно); с ростом номера m интерференционного кольца его радиус убывает. Наибольший радиус имеет кольцо под номером $m = 1$:

$$R_{\max} = \frac{\sqrt{[(b-a)^2 - \lambda^2][(b+a)^2 - \lambda^2]}}{2\lambda}.$$

Максимальное число интерференционных колец можно получить из условия неотрицательности подкоренного выражения в формуле (3.13):

$$m_{\max} = \frac{(b-a)^2}{\lambda^2}. \quad (3.14)$$

Распределение интенсивности для этого случая показано на рис. 3.7.

В предельном случае расстояние между точечными источниками сферических волн может быть числом, кратным $\lambda / 2$. При этом если это расстояние содержит четное число $2m$ полуволн, то в середине интерференционной картины получается максимум порядка m , а если нечетное $(2m + 1)$ – минимум того же порядка m .

Предположим теперь, что в рассмотренном выше втором случае вместо точечных источников сферических волн располагаются линейные источники цилиндрических волн, ориентированные параллельно оси Y . Чтобы найти вид интерференционной картины в плоскости $z = 0$ при наложении цилиндрических волн, следует на основании правила перехода от фазы сферической волны к фазе цилиндрической волны в формулах (3.11) и (3.12) положить $y = 0$. Тогда получим

$$x_m = \frac{\sqrt{[(b-a)^2 - (m\lambda)^2][(b+a)^2 - (m\lambda)^2]}}{2m\lambda}.$$

Это означает, что интерференционная картина в плоскости, параллельной линейным источникам цилиндрических волн, представляет собой семейство параллельных неэквидистантных темных и светлых полос. Положение наиболее удаленной от центра интерференционной картины светлой полосы определяется координатой

$$x_1 = \frac{\sqrt{[(b-a)^2 - \lambda^2][(b+a)^2 - \lambda^2]}}{2\lambda}.$$

Максимальное число интерференционных полос определится той же формулой (3.14).

Рассмотрим теперь вид интерференционной картины в той же плоскости $z = 0$ в случае, когда источники S_1 и S_2 сферических волн расположены не на оси Z , а в произвольных точках (x_1, y_1, a) и (x_1, y_1, b) , используя параксиальное приближение. Поля, создаваемые этими волнами, в параксиальном приближении будут

$$E_1(x, y) = \frac{E_{01}}{r_1} \exp(ikr_1) = \\ = \frac{E_{01} \exp(ika)}{a} \exp \left\{ i \frac{k}{2a} [(x-x_1)^2 + (y-y_1)^2] \right\}$$

и

$$E_2(x, y) = \frac{E_{02}}{r_2} \exp(ikr_2) = \\ = \frac{E_{02} \exp(ikb)}{b} \exp \left\{ i \frac{k}{2b} [(x-x_2)^2 + (y-y_2)^2] \right\},$$

так как в параксиальном приближении

$$r_1 = \sqrt{(x-x_1)^2 + (y-y_1)^2 + a^2} \approx a + \frac{(x-x_1)^2 + (y-y_1)^2}{2a}$$

в предположении, что $(x-x_1)^2 + (y-y_1)^2 \ll a^2$. В таком же виде можно представить и r_2 . Распределение интенсивности света в этой плоскости определится как

$$I(x, y) = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos \left\{ k \left[\frac{1}{2} \left(\frac{1}{a} - \frac{1}{b} \right) (x^2 + y^2) - \right. \right. \\ \left. \left. - \left(\frac{x_1}{a} - \frac{x_2}{b} \right) x - \left(\frac{y_1}{a} - \frac{y_2}{b} \right) y + \Delta \right] \right\},$$

где $I_1 = (E_{01}/a)^2$, $I_2 = (E_{02}/b)^2$, а величина

$$\Delta = a + \frac{x_1^2 + y_1^2}{2a} - b - \frac{x_2^2 + y_2^2}{2b}.$$

Подобно тому как циклическую частоту ω можно определить как производную по времени t от фазы $\varphi = \omega t + \alpha$, соответствующие циклические пространственные частоты можно найти взяв производные фазы по координатам:

$$k_x = k \left[\left(\frac{1}{a} - \frac{1}{b} \right) x - \left(\frac{x_1}{a} - \frac{x_2}{b} \right) \right]; \quad k_y = k \left[\left(\frac{1}{a} - \frac{1}{b} \right) y - \left(\frac{y_1}{a} - \frac{y_2}{b} \right) \right].$$

Циклическая пространственная частота в точке (x, y) будет равна

$$k_{\perp} = \sqrt{k_x^2 + k_y^2} = k \left\{ \left(\frac{1}{a} - \frac{1}{b} \right)^2 (x^2 + y^2) - 2 \left(\frac{1}{a} - \frac{1}{b} \right) \left[\left(\frac{x_1}{a} - \frac{x_2}{b} \right) x + \left(\frac{y_1}{a} - \frac{y_2}{b} \right) y + \left(\frac{x_1}{a} - \frac{x_2}{b} \right)^2 + \left(\frac{y_1}{a} - \frac{y_2}{b} \right)^2 \right] \right\}^{\frac{1}{2}}.$$

Как видно из этого равенства, коэффициенты при квадратичных слагаемых одинаковы, поэтому геометрическим местом точек с одинаковыми значениями пространственных частот являются концентрические окружности. В общем центре этих окружностей пространственная частота равна нулю, а значит, и $k_x = 0$, $k_y = 0$. Из вытекающей из этого условия системы уравнений находим координаты общего центра:

$$x_0 = \frac{x_1/a - x_2/b}{1/a - 1/b}; \quad y_0 = \frac{y_1/a - y_2/b}{1/a - 1/b}. \quad (3.15)$$

Таким образом, интерференционные полосы и в рассмотренном случае произвольного расположения источников являются концентрическими окружностями, которые сгущаются по мере увеличения радиуса.

3.2.3. Интерференция плоской и сферической волн

Предположим сначала, что поле в плоскости $z = 0$ создается сферической волной, источник которой находится на оси Z в точке S на расстоянии a от начала координат, и плоской волной, распространяющейся вдоль положительного направления оси Z (рис. 3.8). Комплексные амплитуды этих волн в плоскости $z = 0$ будут иметь вид

$$E_1 = E_{01} \exp(-i\alpha) \exp(ika);$$
$$E_2(x, y) = \frac{E_{02}}{r} \exp(ik \sqrt{x^2 + y^2 + a^2}),$$

где α – начальная фаза плоской волны, которую мы в дальнейшем будем полагать равной нулю, множитель $\exp(ika)$ в формуле плоской волны обусловлен набегом фазы при ее прохождении расстояния a от источника сферической волны до плоскости $z = 0$. Комплексная амплитуда суммарного поля в плоскости $z = 0$

$$E(x, y) = E_{01} \exp(ika) + \frac{E_{02}}{r} \exp(ik \sqrt{x^2 + y^2 + a^2}).$$

Распределение интенсивности в интерференционной картине будет определяться функцией

$$I(x, y) = I_1 + I_2 + 2 \sqrt{I_1 I_2} \cos [k (\sqrt{x^2 + y^2 + a^2} - a)], \quad (3.16)$$

где $I_1 = E_{01}^2$, $I_2 = (E_{02} / r)^2$. Как видно из (3.16), интенсивность зависит только от расстояния $\rho = \sqrt{x^2 + y^2 + a^2}$ от начала коор-

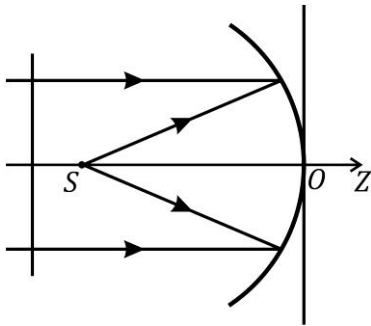


Рис. 3.8

динат. Поэтому интерференционная картина имеет вид чередующихся светлых и темных концентрических колец.

Рассмотрим некоторую достаточно малую область вблизи начала координат, такую что

$$x^2 + y^2 \ll a^2.$$

Тогда справедливо параксиальное приближение

$\sqrt{x^2 + y^2 + a^2} \approx a + \frac{x^2 + y^2}{2a}$. Разность хода волн в параксиальном приближении

$$\sqrt{x^2 + y^2 + a^2} - a \approx \frac{x^2 + y^2}{2a} = \frac{\rho^2}{2a}.$$

С учетом этого выражение (3.15) приводится к виду

$$I(\rho) = \alpha + \beta \cos k \frac{\rho^2}{2a}, \quad (3.17)$$

где коэффициенты $\alpha = I_1 + I_2$, $\beta = 2\sqrt{I_1 I_2}$. График зависимости интенсивности I от расстояния ρ показан на рис. 3.9. Расстояние между соседними минимумами функции $I(\rho)$ определяет ширину светлого кольца. Из графика видно, что с удалением от центра интерференционной картины ширина колец уменьшается.

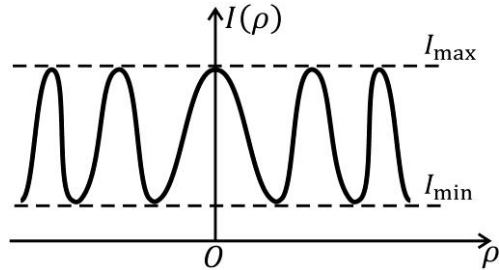


Рис. 3.9

Светлые интерференционные кольца (соответствующие максимальной интенсивности $I_{\max} = \alpha + \beta$) образуются при выполнении условия

$$k\rho^2 / 2a = 2\pi\rho^2 / 2a\lambda = 2m\pi.$$

Отсюда получаем радиусы светлых колец

$$\rho_m = \sqrt{2m\lambda a},$$

Аналогично можно показать, что темные интерференционные кольца (соответствующие минимальной интенсивности ($I_{\min} = \alpha - \beta$)) имеют радиус

$$\rho_m = \sqrt{(2m+1)\lambda a},$$

где $m = 0, 1, 2, \dots$ – номер кольца. Как видим, с увеличением номера кольца m радиус светлого кольца увеличивается пропорционально \sqrt{m} , а темного – пропорционально $\sqrt{(2m+1)}$. Первый интерференционный максимум образуется в центре интерференционной картины ($m = 0$), поэтому центр картины будет освещенным. Первое темное интерференционное кольцо имеет радиус $\rho_0 = \sqrt{\lambda a}$. Можно записать общее выражение для радиуса кольца:

$$\rho_m = \sqrt{m\lambda a},$$

где светлым кольцам соответствует $m = 0$ или четное число m , а темным – нечетное m . Используя эту формулу, условие выполнения паракиального приближения $\rho_m = \sqrt{m\lambda a} \ll a$ можно записать в виде $m \ll a/\lambda$.

Площадь светлого кольца

$$\Delta\sigma = \pi\rho_m^2 - \pi\rho_{m-1}^2 = 2\pi\lambda a. \quad (3.18)$$

Такую же площадь имеет и темное кольцо. Заметим, что площадь кольца не зависит от его номера – одинакова для всех колец. Это приводит к тому, что ширина колец уменьшается с удалением их от центра интерференционной картины. Действительно, площадь кольца, средний радиус которого ρ и ширина $\Delta\rho$, определяется как $\Delta\sigma = 2\pi\rho\Delta\rho$. Сравнивая это выражение с выражением (3.18), получим $\rho\Delta\rho = \lambda a$. Откуда ширина кольца $\Delta\rho = \lambda a/\rho$, т.е. обратно пропорциональна его радиусу ρ . Вид интерференционной картины показан на рис. 3.10.

В данном примере все четные кольца – светлые, а все нечетные – темные. Следует, однако, отметить, что при изменении фазовых соотношений между плоской и сферической волнами интерференционные кольца изменяют свой размер. В частности, если начальную фазу α плоской волны выбрать равной $\pi/2$, то светлые и темные кольца поменяются местами. При этом светлым кольцам будет соответствовать нечетное значение m , а темным – четное m . Такое распределение возникает в картине рассматриваемых ниже колец Ньютона, в описанном выше опыте по интерференции света при отражении от двух сторон тонкого плосковогнутого воздушного промежутка, образованного при контакте плосковыпуклой линзы с плоской стеклянной пластинкой.

Если бы источник сферической волны располагался не на оси

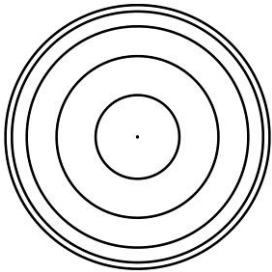


Рис. 3.10

Z , а в произвольной точке (x_1, y_1, a) , то и в этом случае интерференционная картина в плоскости $z = 0$ представляла бы систему концентрических колец. Координаты общего центра колец можно найти по общей формуле (3.15), учтя, что радиус кривизны плоской волны равен бесконечности. Положив в (3.15) $b \rightarrow \infty$, получим $x_0 = x_1, y_0 = y_1$. В этом случае центр находится в точке пересечения с плоскостью $z = 0$ луча от источника сферической волны, идущего перпендикулярно этой плоскости.

3.2.4. Интерференция плоской и цилиндрической волн

Предположим теперь, что поле в плоскости $z = 0$ создается плоской волной, распространяющейся вдоль положительного направления оси Z , и цилиндрической волной, линейный источник которой находится там, где располагался точечный источник S и который ориентирован параллельно оси Y . Распределение интенсивности в интерференционной картине при суперпозиции плоской и цилиндрической волн можно найти, положив в выражении (3.16) $y = 0$. Тогда получим

$$I(x) = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos [k(\sqrt{x^2 + a^2} - a)].$$

При этом $I_1 = E_{01}^2$, $I_2 = E_{02}^2 / r$. В параксиальном приближении, когда $x \ll a$, $\sqrt{x^2 + a^2} \approx a + \frac{x^2}{2a}$. Разность хода

плоской и цилиндрической волн в этом приближении будет

$$\Delta(x) = \sqrt{x^2 + a^2} - a \approx \frac{x^2}{2a}.$$

С учетом этого будем иметь

$$I(x) = \alpha + \beta \cos k \frac{x^2}{2a}. \quad (3.19)$$

График этого распределения имеет такой же вид, как и график распределения $I(\rho)$, показанный на рис. 3.9. Коэффициенты α и β в формуле (3.19) определяются теми же выражениями, что и в формуле (3.16). Начальная фаза плоской волны полагается равной нулю. Интерференционная картина, соответствующая такому распределению интенсивности, представляет собой чередующиеся светлые и темные параллельные полосы. Распределение их вдоль оси X подчиняется параболическому закону. Ширина полос при удалении от центра интерференционной картины уменьшается. Середины светлых полос располагаются вдоль прямых $x = x_m = \pm \sqrt{2m\lambda a}$, а темных – вдоль прямых $x = x_m = \pm \sqrt{(2m+1)\lambda a}$. В центре интерференционной картины ($m = 0$) располагается светлая полоса, а первая темная полоса располагается на расстоянии $x_0 = \sqrt{\lambda a}$ от центра. Знаки \pm в выражениях для x_m означают, что интерференционная картина симметрична относительно ее середины (относительно оси Y). Темные и светлые полосы поменяются местами, если начальная фаза плоской волны будет равна $\pi/2$.

3.3. Методы получения когерентных световых пучков

Мы рассмотрели интерференцию монохроматических волн. Однако реальные световые волны, испускаемые естественными

(не лазерными) источниками света, не являются монохроматическими. Это обусловлено самим механизмом излучения света такими источниками. Световая волна, излучаемая отдельным атомом, не является строго монохроматической (бесконечной синусоидой), а, как уже отмечалось, представляет собой затухающий волновой цуг длительности $\tau \sim 10^{-8}$ с и протяженности $l = c\tau \sim \sim 3$ м. Естественными (не лазерными) источниками света являются в основном сильно нагретые тела (Солнце, нить накала электрической лампочки и др.) и газоразрядные трубки. В этих источниках возбуждение атомов и молекул происходит либо за счет энергии их теплового движения при столкновении атомов или молекул, либо за счет энергии газового разряда. Возбужденные атомы и молекулы отдают свою энергию, испуская излучение в виде цугов волн. Излучение атомов происходит независимо друг от друга. Излучив цуг, атом через какой-то случайный промежуток времени может снова возбудиться и излучить новый цуг. При этом фаза этого нового цуга никак не будет связана (не скоррелирована) с фазой предыдущего цуга того же атома и тем более никак не будет связана с фазой цуга, излученного другим атомом. Одновременно излучает огромное число атомов. Амплитуды и фазы излучаемых ими цугов никак не скоррелированы, являясь случайными величинами. Вследствие различной степени возбужденности атомов частоты цугов также несколько отличаются друг от друга. Совокупность таких цугов и образует в любой момент времени создаваемое источником света излучение. В такой волне фаза φ случайным образом изменяется во времени. Так же случайным образом будет изменяться и разность фаз δ волн, излучаемых двумя независимыми источниками. Поэтому за время наблюдения множитель $\langle \cos \delta \rangle$ в интерференционном члене обращается в нуль, а интенсивность $I(P)$ в каждой точке P оказывается равной сумме интенсивностей складываемых волн – интерференция наблюдаться не будет. Следует, однако, отметить, что

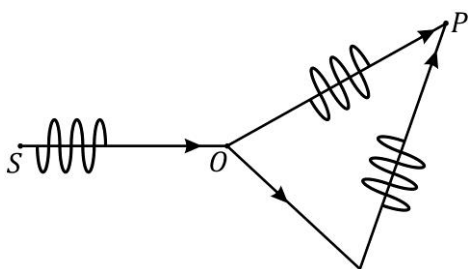


Рис. 3.11

если разность фаз меняется медленно, т.е. так, что за время наблюдения, равное времени разрешения приемника $t_{пр}$, она остается практически постоянной, то прибор зарегистрирует интерференцию, а такие

волны следует признать когерентными. Как видим, понятие когерентности относительно: две волны ведут себя как когерентные при наблюдении с одним прибором и как некогерентные при наблюдении с другим прибором.

Возникает вопрос: можно ли для света создать условия, при которых наблюдалась бы интерференция? Другими словами, как, пользуясь некогерентными (несогласованными) источниками, получить когерентные световые пучки? Это оказывается возможным, если обе, накладывающиеся друг на друга волны будут создаваться не разными, а одним и тем же источником. Для этого, как предложил Френель, световой пучок от источника следует искусственно разделить на два пучка, заставить их пройти различные расстояния для приобретения разности хода и затем наложить друг на друга подходящим образом (рис. 3.11). Тогда случайные слагаемые в фазах φ_1 и φ_2 , имея общее происхождение, сократятся, и останется только приобретенная разность фаз $\delta = \varphi_1 - \varphi_2$, связанная с оптической разностью хода $\Delta = n_2 r_2 - n_1 r_1$: $\delta = k\Delta = (2\pi / \lambda_0)\Delta$, где r_1 и r_2 – расстояния, проходимые световыми пучками от точки деления до точки наложения P , а n_1 и n_2 – показатели преломления сред, в которых распространяются пучки. Имея постоянную во времени разность фаз, эти волны будут интерферировать.

Лучи, образовавшиеся в результате деления исходного (первичного) луча, называют вторичными лучами, а интерференцию этих вторичных лучей – двухлучевой

интерференцией. Угол ω между лучами, исходящими из источника, которые каким-либо способом сводятся далее в точку наблюдения, называется апертурой интерференции. Апертура интерференции является важной характеристикой любой интерференционной схемы. Другой важной характеристикой является угол схождения лучей β в точке наблюдения P . Пути, по которым распространяются интерферирующие волны от точки деления до точки наложения P , называют плечами интерференционной схемы.

Обычно разность хода Δ мала, т.е. $r_1 \approx r_2$, поэтому интенсивности $I_1(P)$ и $I_2(P)$ можно считать равными. Тогда, обозначив интенсивность, создаваемую в точке P одним пучком, через $I_0(P)$, для результирующей интенсивности в точке P будем иметь:

$$I(P) = 2I_0(P) (1 + \cos k\Delta) \quad (3.20)$$

или

$$I(P) = 4I_0(P) \cos^2(k\Delta / 2). \quad (3.21)$$

Формулы (3.20) и (3.21) являются универсальными. Они справедливы для любой двухлучевой интерференции волн одинаковой интенсивности.

В точке раздвоения пучка происходит деление каждого цуга на две части одинаковой протяженности, равной протяженности исходного цуга. В этой точке обе части цуга имеют одинаковую фазу φ и, следовательно, разность фаз $\delta = 0$. Но в точку встречи P , благодаря разности хода Δ , они приходят с разностью фаз $\delta = (2\pi / \lambda_0)\Delta$. Накладываясь друг на друга, они создают интерференционную картину. Указанная разность фаз остается неизменной для каждого цуга, излучаемого источником, после его раздвоения и прихода в точку встречи. Поэтому интерференционная картина каждый раз будет в точности повторяться, а прибор регистрирует стационарную интерференционную картину.

Существуют два метода получения когерентных световых пучков: метод деления волнового фронта и метод деления амплитуды. В интерференционных схемах, работающих по методу деления волнового фронта, исходящий из источника пучок делится на два либо проходя через два близко расположенных отверстия в непрозрачном экране, либо отражаясь от зеркальных поверхностей, либо преломляясь в двух призмах или линзах. В интерференционных схемах, работающих по методу деления амплитуды, исходный пучок делится на одной или нескольких полупрозрачных (частично пропускающих, частично отражающих) поверхностях. При этом волновой цуг падающего света частично отражается на первой границе раздела сред; уменьшив амплитуду, он передается дальше, частично отражается на следующей границе и т. д. Интерференция возникает, если два отраженных пучка накладываются друг на друга, а ее результат зависит от разности пути, которая появляется между ними из-за разности поверхностей. Метод деления волнового фронта пригоден только при достаточно узких (точечных) отверстиях. Метод деления амплитуды может применяться с протяженными источниками и при малом расстоянии между отражающими поверхностями обеспечивает большую контрастность интерференционной картины, чем метод деления волнового фронта.

Эти два метода получения интерференционных пучков из одного светового пучка лежат в основе классификации устройств, применяемых в интерферометрии.

Четкость (контрастность) получающейся на экране интерференционной картины можно количественно оценить с помощью введенной Майкельсоном величины V , называемой функцией видности (или просто видностью) интерференционных полос. Эта функция определяется как отношение разности интенсивностей в соседнем максимуме и минимуме к удвоенному среднему значению интенсивности:

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \quad (3.22)$$

где I_{\max} и I_{\min} – соответственно максимальное и минимальное значения интенсивности. Заметим, что видность интерференционных полос определяется аналогично глубине модуляции в радиотехнике. Функция видности V имеет максимальное значение, равное единице, когда интенсивность в минимуме $I_{\min} = 0$, и минимальное значение, равное нулю, когда $I_{\max} = I_{\min}$ (в отсутствие интерференции). В общем случае видность принимает значения, заключенные между нулем и единицей ($0 \leq V \leq 1$). Подставив в формулу (3.22) выражения для максимальной и минимальной интенсивностей, соответствующих $\cos\delta = \pm 1$ в (3.1), выразим функцию видности через интенсивности I_1 и I_2 , создаваемые складываемыми волнами:

$$V = 2 \frac{\sqrt{I_1 I_2}}{I_1 + I_2}. \quad (3.23)$$

Из этой формулы следует, что при интерференции монохроматических световых волн видность интерференционной картины будет максимальной ($V = 1$) при равенстве интенсивностей $I_1 = I_2$.

3.4. Интерференция по методу деления волнового фронта. Опыт Юнга

Простейшей двухлучевой интерференционной схемой, в которой используется метод деления волнового фронта, является схема Юнга, показанная на рис. 3.12.

Пучок света падает на экран \mathcal{E}_1 с узким отверстием S или щелью. Прошедший через отверстие S свет вследствие дифракции образует расходящийся пучок, который падает на второй экран \mathcal{E}_2 с двумя узкими отверстиями или щелями S_1 и S_2 , расположенными близко друг к другу и на равных расстояниях от

S . Когда фронт волны из S достигнет этих отверстий, они согласно принципу Гюйгенса становятся источниками вторичных волн. Эти вторичные источники являются когерентными, так как колебания в них возбуждаются одной и той же падающей на экран волной. Испускаемые ими расходящиеся цилиндрические волны, накладываются друг на друга и создают интерференционную картину, наблюдаемую на экране \mathcal{E} , находящемся на достаточно большом удалении от экрана \mathcal{E}_2 . Так как источники S_1 и S_2 колеблются в одинаковых фазах, начальная разность фаз испускаемых ими волн равна нулю. Поэтому разность фаз этих волн в точке их наложе-

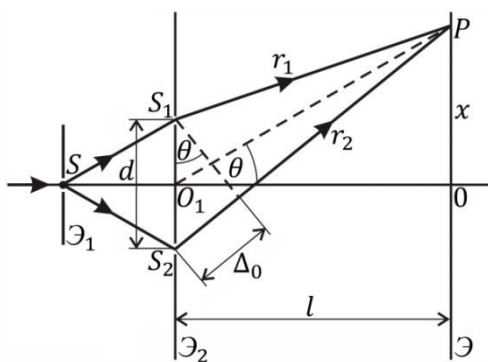


Рис. 3.12

ния P будет равна $\delta = (2\pi / \lambda_0)\Delta$, где $\Delta = n\Delta_0 = n(r_2 - r_1)$ – оптическая разность хода, а $\Delta_0 = r_2 - r_1$ – геометрическая разность хода волн, испускаемых источниками S_1 и S_2 ; n – показатель преломления среды, заключенной между

экранами \mathcal{E}_2 и \mathcal{E} ; r_1 и r_2 – расстояния от источников S_1 и S_2 до точки наблюдения P .

Пусть θ – угол, образуемый направлением на точку P с оптической осью O_1O схемы, т.е. с линией, перпендикулярной экранам \mathcal{E}_2 и \mathcal{E} и проходящей через середину расстояния между щелями S_1 и S_2 . Этот угол мал ($\theta \ll 1$), так как расстояние d между источниками S_1 и S_2 много меньше расстояния l от этих источников до экрана \mathcal{E} ($d \ll l$), поэтому, как видно из рис. 3.12, $\Delta_0 \approx d \cdot \theta$. Но $\theta \approx x / l$, где x – расстояние от точки P на экране \mathcal{E} до плоскости, перпендикулярной S_1S_2 и содержащей

оптическую ось O_1O , т.е. вертикальная координата точки P . Следовательно,

$$\Delta_0 \approx x d / l. \quad (3.24)$$

Так как отверстия S_1 и S_2 одинаковы, а $r_1 \approx r_2$ (поскольку $\Delta_0 \ll \ll l$), то интенсивности $I_1(P)$ и $I_2(P)$, создаваемые источниками S_1 и S_2 в точке P , можно считать одинаковыми. Поэтому результирующую интенсивность $I(P)$ в точке P можно рассчитать по формуле (3.20) или (3.21). Из этих формул видно, что максимальное значение интенсивности $I_{\max} = 4I_0$ достигается, когда

$$k\Delta = (2\pi / \lambda_0) \Delta = 2m \cdot \pi,$$

где $m = 0, \pm 1, \pm 2, \dots$, т.е. при разности хода Δ , равной целому числу длин волн: $\Delta = m\lambda_0$. Минимальное значение интенсивности $I_{\min} = 0$ имеет место, когда

$$(2\pi / \lambda_0) \Delta = (2m + 1)\pi,$$

т.е. при разности хода, равной полуцелому числу длин волн: $\Delta = (m + 1/2)\lambda_0$. Целое число m называется порядком интерференции. На экране \mathcal{E} наблюдается распределение максимумов и минимумов интенсивности, называемых интерференционными максимумами и минимумами.

Подставив в (3.20) значение $\Delta = n\Delta_0 = nxd / l$, получим зависимость интенсивности света на экране от координаты x :

$$I(x) = 2I_0 \left(1 + \cos \frac{2\pi d}{\lambda l} x \right) = 4I_0 \cos^2 \frac{\pi d}{\lambda l} x, \quad (3.25)$$

где $\lambda = \lambda_0 / n$ – длина волны света в среде. Из формулы (3.25) видно, что интенсивность имеет одно и то же значение при $x = \text{const}$, т.е. вдоль прямой, перпендикулярной плоскости,

содержащей оптическую ось O_1O . Иными словами, прямые $x = \text{const}$ представляют собой линии одинаковой интенсивности. При этом максимум интенсивности наблюдается на прямых

$$x = x_m = m\lambda l / d,$$

а минимум – на прямых

$$x = x_m = (m + 1/2)\lambda l / d.$$

В интервале между указанными прямыми интенсивность света принимает некоторые промежуточные значения между I_{\min} и I_{\max} .

Угловое расстояние φ_m максимума m -го порядка от плоскости симметрии (плоскости $x = 0$) при малых m , а значит, и при малых углах φ_m , когда $x_m / l = \text{tg } \varphi_m \approx \sin \varphi$, определится из соотношения

$$d \sin \varphi_m = m\lambda,$$

а минимума – из соотношения

$$d \sin \varphi_m = (m + 1/2)\lambda.$$

Таким образом, распределение интенсивности света на экране Э имеет вид плавно чередующихся темных и светлых полос, которые, как уже отмечалось, называются интерференционными полосами. Освещенность всех светлых полос одинакова. Вдоль прямой $x = 0$ расположен максимум, соответствующий нулевой разности хода ($\Delta = 0$). Порядок этого максимума $m = 0$. Этот максимум называется нулевым или центральным максимумом. Вверх и вниз от центрального максимума на равных расстояниях друг от друга располагаются максимумы и минимумы первого ($m = \pm 1$), второго ($m = \pm 2$) и т.д. порядков.

Ширина интерференционной полосы

$$h = x_m - x_{m-1} = \lambda l / d, \quad (3.26)$$

где x_i ($i = m, m - 1$) – координата i -го максимума или минимума,

или

$$h = \lambda / \beta,$$

где $\beta \approx d / l$ – угол схождения интерферирующих лучей. Выражение (3.25) можно получить также из условия, что при переходе от одной светлой или темной полосы к другой такой же полосе аргумент косинуса в (3.25) изменяется на 2π , т.е. из условия

$$[2\pi d / (\lambda l)] \Delta x = 2\pi, \Delta x = h.$$

Из выражения (3.26) видно, что ширина интерференционной полосы не зависит от ее номера m и является постоянной при заданных значениях l , d и λ . Она представляет собой пространственный период интерференционной картины. При постоянных значениях l и λ уменьшение расстояния d между источниками S_1 и S_2 приводит к уширению интерференционных полос, интерференционная картина становится более отчетливой (контрастной). К этому же результату приводит и увеличение расстояния от источников S_1 и S_2 до экрана Э при неизменном расстоянии d между источниками. Чтобы интерференционная картина наблюдалась

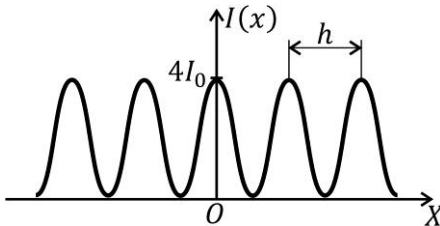


Рис. 3.13

невооруженным глазом, ширина полосы h должна быть не меньше 1 мм. Тогда для $\lambda \approx 0,5$ мкм будем иметь $d / l \approx 0,5 \cdot 10^{-3}$, так что условие $d \ll l$ выполняется.

$$\text{Формула} \quad (3.26)$$

позволяет по измерению ширины интерференционной полосы h и известным параметрам d и l схемы определить длину волны света λ .

Используя выражение для ширины интерференционной полосы, формулу (3.26), описывающую распределение

интенсивности в интерференционной картине, можно записать в обобщенном виде:

$$I(x) = 2I_0 \left(1 + \cos \frac{2\pi}{h} x \right). \quad (3.27)$$

Отсюда видно, что величина h является периодом изменения интенсивности в интерференционной картине. График зависимости $I(x)$ изображен на том же рис. 3.13.

Поскольку в этом опыте $I_1(P) \approx I_2(P)$, то видность интерференционной картины V в этом опыте близка к единице. Отметим также, что при смещении источника S вдоль линии, параллельной оси X , интерференционные полосы тоже смещаются вдоль оси X : вниз при смещении источника вверх и вверх при смещении источника вниз.

Расчет интерференционных картин, получаемых с помощью других интерференционных схем по методу деления волнового фронта, ничем не отличается от рассмотренного выше для схемы Юнга (эти схемы фактически эквивалентны схеме Юнга). Во всех таких схемах ход лучей, пересекающихся в точке P , таков, как если бы эти лучи исходили не из одного реального источника S , а из двух когерентных источников (действительных или мнимых) S_1 и S_2 , расположенных на расстоянии d друг от друга и на расстоянии l от экрана, на котором наблюдается интерференционная картина. Задача заключается лишь в том, чтобы выразить расстояния d и l через параметры конкретной схемы.

3.5. Интерференция по методу деления амплитуды

3.5.1. Кольца Ньютона

В качестве примера интерференции света по методу деления амплитуды рассмотрим интерференцию, возникающую при

падении света на плосковыпуклую линзу малой кривизны, соприкасающуюся в воздухе с плоской поверхностью стекла (рис. 3.14). Падающий нормально на плоскую поверхность линзы и прошедший сквозь ее толщину пучок 1 монохроматического света, частично отражается от границы раздела линза – воздух (луч 1'), частично проходит в воздух (преломляется) и затем отражается от границы раздела воздух – стекло (луч 1''). Интерференция происходит между двумя вторичными лучами 1' и 1''.

Заметим, что, кроме лучей 1' и 1'', возникают еще лучи, отраженные от верхней и нижней поверхностей линзы. Однако эти лучи не интерферируют, так как из-за большой их оптической разности хода, обусловленной большой толщиной линзы, они оказываются некогерентными.

Разность хода лучей 1' и 1'' равна $2d$, где d – толщина воздушного слоя в месте падения луча, зависящая от расстояния ρ до точки касания линзы со стеклом и радиуса кривизны линзы R . С помощью рис. 3.14 и условия $\rho \ll R$, т.е. при $(\rho/R)^2 \ll 1$, находим толщину воздушного промежутка:

$$d = R - \sqrt{R^2 - \rho^2} = R(1 - \sqrt{1 - \rho^2/R^2}) \approx \rho^2 / 2R.$$

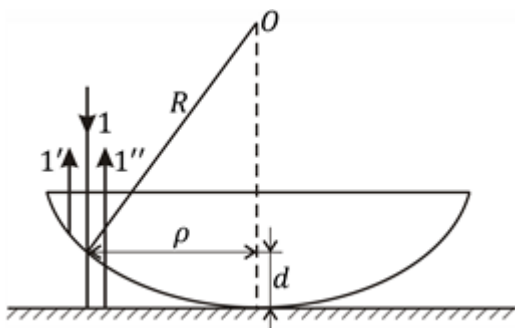


Рис. 3.14

Разность фаз рассматриваемых лучей в отраженном свете равна

$$\delta = (2\pi / \lambda_0) 2d + \pi = \frac{2\pi}{\lambda_0} \frac{\rho^2}{R} + \pi,$$

где добавочное слагаемое π

учитывает изменение фазы на π при отражении от стеклянной пластинки как от оптически более плотной среды, чем воздух.

При освещении установки монохроматическим пучком света в отраженном свете будет наблюдаться интерференционная картина в виде чередующихся светлых и темных колец. Эти кольца называют кольцами Ньютона.

Условием образования светлых интерференционных колец является соотношение

$$\frac{2\pi}{\lambda_0} \frac{\rho^2}{R} + \pi = 2m\pi,$$

где m – целое положительное число за исключением нуля. С учетом этого находим радиусы светлых колец Ньютона:

$$\rho_m = \sqrt{(m - 1/2)\lambda_0 R}. \quad (3.28)$$

Темные кольца, возникающие при выполнении условия

$$\frac{2\pi}{\lambda_0} \frac{\rho^2}{R} + \pi = (2m + 1)\pi,$$

имеют радиусы

$$\rho_m = \sqrt{m\lambda_0 R}, \quad (3.29)$$

где m – целое положительное число, включая нуль.

С помощью формулы (3.28) или (3.29) можно определить длину волны используемого света. Для этого следует измерить ра-

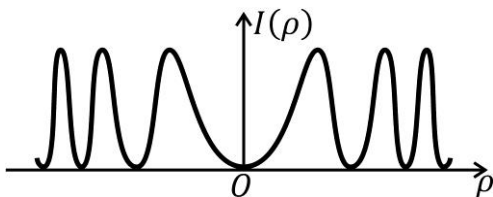


Рис. 3.15

диус m -го светлого или темного интерференционного кольца и радиус кривизны R линзы. График радиального распределения интенсивности $I(\rho)$ в

отраженном свете показан на рис. 3.15.

Таким образом, радиусы светлых колец Ньютона пропорциональны квадратному корню из полуцелых чисел, а темных – квадратному корню из натуральных чисел. Центр интерференционной картины (в отраженном свете) будет темным, так как в окрестности точки соприкосновения линзы с пластинкой толщина воздушной прослойки $d \approx 0$, а разность фаз δ интерферирующих лучей примерно равна π . Центру интерференционной картины соответствует радиус кольца $r_m = 0$, что имеет место при $m = 0$.

Интерференционная картина в проходящем свете будет дополняющей к интерференционной картине в отраженном свете. Радиусы темных колец будут определяться формулой (3.28), а светлых – формулой (3.29).

3.5.2. Интерферометр Майкельсона

На явлении интерференции по методу деления амплитуды основано действие интерферометров – приборов, служащих для создания когерентных световых пучков, формирования интерференционной картины и ее интерпретации. Наиболее известным интерферометром является интерферометр Майкельсона. Майкельсон впервые использовал его для проверки возможности движения Земли относительно «светоносного эфира», а точнее, для решения вопроса о том, зависит ли скорость света от скорости движения Земли (классический опыт Майкельсона – Морли). В настоящее время интерферометр Майкельсона используется в качестве диспергирующей системы для преобразования сложного немонахроматического излучения и анализа его спектрального состава. Спектральный прибор, построенный на базе интерферометра Майкельсона, называют фурье-спектрометром.

Оптическая схема интерферометра Майкельсона показана на рис. 3.16. Свет от источника S падает на полупрозрачную посеребренную пластинку P , на которой происходит деление

световой волны по амплитуде. После частичного отражения от пластинки и частичного преломления в ней первичный пучок делится на два луча 1 и 2, перпендикулярны друг другу и идущих к зеркалам M_1 и M_2 . После отражения от зеркал лучи попадают в линзу L , в фокальной плоскости которой и наблюдается интерференционная картина. При нестрогой перпендикулярности зеркал M_1 и M_2 в интерферометре будет иметь место такая же интерференция, как в клине; при строгой перпендикулярности – будет наблюдаться интерференция, аналогичная той, которую мы имели бы в плоскопараллельной пластине.

В общем случае зеркала M_1 и M_2 находятся на различных расстояниях d_1 и d_2 от пластинки P , поэтому между этими лучами создается разность хода $\Delta = 2(d_2 - d_1)$. Результирующую интенсивность можно определить с помощью формулы (3.20). При этом следует учесть, что волны 1 и 2 возникают после прохождения через полупрозрачную пластинку P , на которой исходная волна частично отражается, а частично преломляется, поэтому в формуле (3.20) вместо интенсивности I_0 падающей на пластинку волны следует взять произведение RTI_0 , где R и T – коэффициенты отражения и пропускания света (по интенсивности) делительной пластинкой P ; при этом в пренебрежении поглощением света в пластинке $R + T = 1$. Поскольку пластинка полупрозрачная, то $T = 1/2$, а значит, и $R = 1 - T = 1/2$. С учетом этого формулу (3.20) запишем в виде

$$I = \frac{1}{2} I_0 (1 + \cos k\Delta). \quad (3.30)$$

Из этой формулы видно, что в максимуме интенсивность света I становится равной интенсивности падающего света I_0 . Иначе

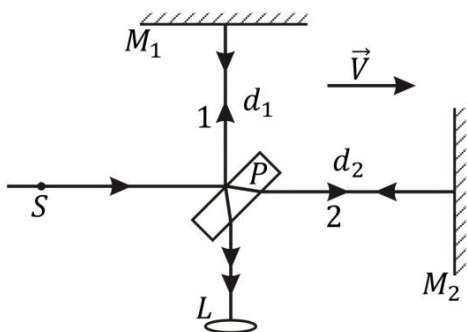


Рис. 3.16

говоря, при разности хода $\Delta = m\lambda$

интерферометр Майкельсона полностью пропускает падающую на него плоскую монохроматическую световую волну. В минимуме при $\Delta = m(1 + 1/2)\lambda$,

напротив, интенсивность прошедшего света оказывается равной нулю. Это означает, что падающая плоская волна полностью отражается назад в источник. Следовательно, интерферометр Майкельсона может служить фильтром, пропускание которого зависит от длины волны.

Если одно из зеркал (например, зеркало M_2) сместить еще на расстояние h , то возникнет дополнительная разность хода, равная $2h$. Это скажется на интерференционной картине – интерференционные полосы претерпят некоторое смещение, доступное измерению. Таким способом можно измерять очень малые изменения оптических путей обеих лучей. По величине смещения полос можно определить, например, длину волны λ , излучаемую источником S (если он монохроматический). Действительно, при дополнительной разности хода, равной λ , смещение полос $\Delta m = 1$, а при дополнительной разности хода $2h$ смещение полос составит $\Delta m = 2h/\lambda$, откуда $\lambda = 2h/\Delta m$. По величине смещения полос предполагалось определить и скорость движения Земли относительно «светоносного эфира».

Рассмотрим действие интерферометра Майкельсона как диспергирующей системы. Пусть одно из зеркал интерферометра (например, M_2) перемещается параллельно самому себе с постоянной скоростью v . Тогда разность хода Δ будет изменяться, имея в каждый момент времени t значение $\Delta = 2vt$. При монохроматическом освещении интерферометра распределение

интенсивности полос, в соответствии с (3.30), будет изменяться по закону $I(t) = (1/2)I_0(1 + \cos 2kvt)$ или

$$I(t) = (1/2)I_0(1 + \cos 2\pi f t). \quad (3.31)$$

Интенсивность света, регистрируемого фотоприемником, расположенным в задней фокальной плоскости линзы L , оказывается периодической функцией времени. Как видим, при таком способе наблюдения происходит модуляция интенсивности, а значит, и сигнала фотоприемника с частотой модуляции $f = (2v/c)v$. При этом частота модуляции f зависит от частоты ν монохроматического излучения источника. Поэтому, измерив частоту модуляции f , можно найти и частоту ν и тем самым получить информацию о спектре источника. При скорости перемещения зеркала $v = 1$ см / с и частоте источника $\nu = 4 \cdot 10^{15}$ Гц частота модуляции $f \approx 2,4 \cdot 10^5$ Гц, что лежит в радиочастотной области.

Обсуждение этого вопроса будет продолжено в п. 4.4, где будет рассмотрен случай освещения интерферометра Майкельсона сложным некогерентным излучением.

Глава 4

КОГЕРЕНТНОСТЬ

4.1. Частичная когерентность. Корреляция световых волн

В предыдущей главе речь шла об интерференционных явлениях, возникающих при наложении двух монохроматических волн. Рассмотрим теперь возникновение интерференции в некоторой точке P в случае, когда складываемые в ней волны от точечных источников S_1 и S_2 не являются строго монохроматическими. Будем считать, что источники S_1 и S_2

излучают квазимонохроматические волны с одинаковой центральной частотой ω_0 :

$$E_1(t) = E_{01}(t)\exp\{-i[\omega_0 t + \varphi_1(t)]\},$$

$$E_2(t) = E_{02}(t)\exp\{-i[\omega_0 t + \varphi_2(t)]\}.$$

Учитывая, что колебания светового поля, возбуждаемые источниками S_1 и S_2 в точке P , запаздывают по сравнению с колебаниями самих источников, будем обозначать эти колебания как $E_1(t - \tau_1)$ и $E_2(t - \tau_2)$, где $\tau_1 = r_1/c$ и $\tau_2 = r_2/c$ – времена распространения первой и второй волн от источников S_1 и S_2 соответственно до точки наблюдения P , r_1 и r_2 – расстояния от источников S_1 и S_2 до точки наблюдения P . Введение времен τ_1 и τ_2 означает, что колебания в точке P в момент времени t такие же, как в точках S_1 и S_2 в моменты времени $t - \tau_1$ и $t - \tau_2$. Результирующее поле в точке P , согласно принципу суперпозиции,

$$E(P) = E_1(t - \tau_1) + E_2(t - \tau_2),$$

а интенсивность в этой точке

$$I(P) = \langle |E(P)|^2 \rangle = \langle |E_1(t - \tau_1)|^2 \rangle + \langle |E_2(t - \tau_2)|^2 \rangle + 2\text{Re} \langle E_1(t - \tau_1) \cdot E_2^*(t - \tau_2) \rangle. \quad (4.1)$$

Смещение начального момента усреднения обычно не влияет на величину среднего², поэтому первые два слагаемых в (4.1) равны интенсивностям интерферирующих световых волн, так что

$$\langle |E(t)|^2 \rangle = \langle |E(t - \tau)|^2 \rangle = I.$$

² Оптические поля, удовлетворяющие этому условию, называют стационарными.

Сместим начальный момент усреднения t в точку $t' = t - \tau_1$, т.е. заменим t на t' . Тогда, опустив штрих в символе t' , третье слагаемое в (4.1) можно записать в виде $2\text{Re}\langle E_1(t) \cdot E_2^*(t - \theta) \rangle$.

Здесь $\theta = \tau_2 - \tau_1$ – относительное запаздывание колебаний $E_2(t)$ и $E_1(t)$ в точке наблюдения P . Если разность хода волн от этих источников до точки наблюдения P равна $\Delta = r_2 - r_1$, то колебания светового поля, создаваемые волной от одного источника (в нашем случае S_2), будут запаздывать по сравнению с колебаниями, создаваемыми волной от другого источника (S_1), на промежуток времени $\theta = \Delta / c$. Среднее по времени наблюдения произведения комплексной амплитуды $E_1(t)$ на комплексно сопряженную комплексной амплитуды $E_2(t - \theta)$, взятую в другой момент времени, отделенный от первого на промежуток θ , т.е. величину

$$\Gamma_{12}(\theta) = \langle E_1(t) \cdot E_2^*(t - \theta) \rangle,$$

называют комплексной функцией корреляции колебаний E_1 и E_2 (создаваемых источниками S_1 и S_2 в точке P). Ее также называют комплексной функцией взаимной когерентности колебаний или функцией взаимной когерентности. Она является количественной мерой коррелированности (согласованности) двух случайных процессов в два момента времени, разделенных промежутком θ . Следует отметить, что если сам волновой процесс $E(t)$ является случайной функцией времени, то функция когерентности $\Gamma_{12}(\theta)$ детерминирована. Она определяет закон изменения (потери) статистической памяти с ростом θ . Функцию взаимной когерентности широко используют при анализе интерференционных явлений. Выразив усреднение по времени через интегрирование, для этой функции можно записать

$$\Gamma_{12}(\theta) = \frac{1}{\tau} \int_0^{\tau} E_1(t) \cdot E_2^*(t - \theta) dt,$$

где τ – ширина временного интервала, по которому производится усреднение.

Функцию $\Gamma_{12}(\theta)$ нормируют, полагая

$$\gamma_{12}(\theta) = \frac{\langle E_1(t) \cdot E_2^*(t - \theta) \rangle}{\sqrt{\langle |E_1(t)|^2 \rangle} \cdot \sqrt{\langle |E_2(t)|^2 \rangle}} \quad (4.2)$$

или

$$\gamma_{12}(\theta) = \frac{\Gamma_{12}(\theta)}{\sqrt{I_1} \cdot \sqrt{I_2}}. \quad (4.3)$$

Нормированную функцию $\gamma_{12}(\theta)$ называют комплексной степенью взаимной когерентности (или корреляции колебаний), а ее модуль $|\gamma_{12}(\theta)|$ – степенью взаимной когерентности колебаний (в точке P). Комплексная корреляционная функция $\Gamma_{12}(\theta)$, а следовательно, и степень взаимной когерентности $|\gamma_{12}(\theta)|$ отражают статистические свойства излучения. Заметим, что модуль числителя в (4.2) не превышает знаменателя, поэтому степень взаимной когерентности $|\gamma_{12}(\theta)| \leq 1$. При $\theta = 0$ степень взаимной когерентности имеет максимальное значение, равное единице. С увеличением времени задержки θ степень взаимной когерентности убывает и при каком-то значении $\theta \geq t_{\text{ког}}$ она обращается в нуль. Максимальное значение времени задержки $\theta = t_{\text{ког}}$, при котором когерентность еще сохраняется, называется временем когерентности. Время когерентности можно определить и как промежуток времени, за который случайные изменения фазы волны достигают значения, равного π . В результате волна

как бы «забывает» свою первоначальную фазу и становится некогерентной по отношению к самой себе.

Расстояние $l_{\text{ког}} = ct_{\text{ког}}$, проходимое светом за время когерентности, называется длиной когерентности. Таким образом, степень взаимной когерентности имеет значение, заключенное в интервале $0 \leq |\gamma_{12}(\theta)| \leq 1$.

Величина $|\gamma_{12}(\theta)|$ доступна экспериментальному определению. Зная ее, можно проводить количественное изучение когерентных свойств оптических полей, определить возможность наблюдения интерференции в том или ином конкретном опыте.

С учетом соотношения (4.3) выражение (4.1) для интенсивности в точке P примет вид

$$I(P) = I_1(P) + I_2(P) + 2\sqrt{I_1(P)I_2(P)} \operatorname{Re} \gamma_{12}(\theta). \quad (4.4)$$

Эта формула может быть использована для определения $\operatorname{Re} \gamma_{12}(\theta)$. Для этого в опыте Юнга достаточно измерить в некоторой точке P интерферограммы значение полной интенсивности $I(P)$ и интенсивностей $I_1(P)$ и $I_2(P)$ света в этой же точке соответственно от каждого источника (отверстий S_1 и S_2) при закрытом другом источнике (отверстии).

При равных интенсивностях $I_1(P) = I_2(P) = I_0(P)$ формула (4.4) упрощается:

$$I(P) = 2 I_0(P) [1 + \operatorname{Re} \gamma_{12}(\theta)]. \quad (4.5)$$

Представим комплексную функцию $\gamma_{12}(\theta)$ в виде

$$\gamma_{12}(\theta) = |\gamma_{12}(\theta)| \exp[-i\varphi_{12}(\theta)],$$

где $\varphi_{12}(\theta)$ – аргумент комплексной функции $\gamma_{12}(\theta)$. Величину $\varphi_{12}(\theta)$ можно рассматривать как обобщенную разность фаз между колебаниями световых полей в точке P . Представим ее в виде суммы $\varphi_{12}(\theta) = \delta + \alpha(\theta)$, где δ – разность фаз,

обусловленная разностью хода Δ , а $\alpha(\theta)$ – дополнительная разность фаз, зависящая от времени задержки θ и обусловленная нестрогой когерентностью волн (при этом $\alpha(0) = 0$). Записав реальную часть в выражении (4.4) как $|\gamma_{12}(\theta)|\cos[\delta + \alpha(\theta)]$, это выражение приведем к виду

$$I(P) = I_1(P) + I_2(P) + 2\sqrt{I_1(P)I_2(P)} |\gamma_{12}(\theta)|\cos[\delta + \alpha(\theta)]. \quad (4.6)$$

При равных интенсивностях обеих волн формула (4.6) принимает вид

$$I(P) = 2I_0 [1 + |\gamma_{12}(\theta)|\cos[\delta + \alpha(\theta)]].$$

Формула (4.6) известна как общий закон интерференции стационарных оптических полей. Она представляет собой наиболее общее выражение, определяющее результат интерференции двух квазимонохроматических волн и справедлива для любой двулучевой интерференционной схемы. От формулы (3.1) для строго монохроматических световых волн формула (4.6) отличается дополнительным множителем $|\gamma_{12}(\theta)|$ в интерференционном члене и добавочным слагаемым $\alpha(\theta)$ в разности фаз. Добавочное слагаемое $\alpha(\theta)$ определяет фазовый сдвиг интерференционного максимума (или минимума) по сравнению с его положением в интерференционной картине при интерференции строго монохроматических волн. Фаза $\alpha(\theta)$ измеряет сдвиг полос вдоль оси X в плоскости наблюдения. Если величина $|\gamma_{12}(\theta)|$ равна своему максимальному значению, единице, а величина $\alpha(\theta) = 0$, то интенсивность в точке P будет такой же, как в случае монохроматического света с разностью фаз $\delta = (2\pi/\lambda_0)\Delta$ между волнами. При таких условиях волны, излучаемые источниками S_1 и S_2 , являются полностью

когерентными. Если же степень взаимной когерентности $|\gamma_{12}(\theta)|$ равна другому своему предельному значению, нулю, то интерференционный член в (4.6) будет отсутствовать; интерференция не происходит, волны, излучаемые источниками S_1 и S_2 , полностью не когерентны, а наблюдаемая интенсивность является простой суммой независимых интенсивностей, создаваемых источниками S_1 и S_2 . Когда величина $|\gamma_{12}(\theta)|$ принимает свое промежуточное значение $0 < |\gamma_{12}(\theta)| < 1$, говорят о частичной когерентности световых волн. В соответствии с этим формулу (4.6) называют также общим законом интерференции для частично когерентного света. На практике световые волны считают близкими к полной когерентности, если $0,88 \leq |\gamma_{12}(\theta)| \leq 1$ и частично когерентными, если $0 \leq |\gamma_{12}(\theta)| < 0,88$. Интерференционные полосы наблюдаются достаточно четко и при видности $V \geq 2/3$. В этом случае $I_{\max} \geq 5I_{\min}$.

Частично когерентный свет можно рассматривать как смесь когерентного и некогерентного света. Величину $|\gamma_{12}(\theta)|$ можно трактовать как долю когерентного света, присутствующего в излучении. В этом случае интенсивность каждого интерферирующего пучка, например первого, можно представить состоящей из двух слагаемых, $I_1 = \gamma I_1 + (1 - \gamma)I_1$, первое из которых (γI_1) определяет интенсивность когерентного света, а второе ($(1 - \gamma)I_1$) – интенсивность некогерентного света.

Степень взаимной когерентности $|\gamma_{12}(\theta)|$ имеет те же пределы изменения ($0 \leq |\gamma_{12}(\theta)| \leq 1$), что и функция видности $V(\Delta)$, определяемая соотношением (3.22). Покажем, что величины $V(\Delta)$ и $|\gamma_{12}(\theta)|$ связаны друг с другом. Действительно, ввиду медленности изменения функции $|\gamma_{12}(\theta)|$ максимальное и минимальное значения интенсивности получаются из соотношения (4.6) при $\cos[\delta + \alpha(\theta)] = \pm 1$ соответственно. С учетом этого для видности интерференционных полос получим

$$V(\Delta) = 2 \frac{\sqrt{I_1 I_2}}{I_1 + I_2} |\gamma_{12}(\theta)|. \quad (4.7)$$

Это и есть формула, выражающая связь видности полос $V(\Delta)$ со степенью взаимной когерентности $|\gamma_{12}(\theta)|$. С помощью этой формулы можно экспериментально определить степень взаимной когерентности $|\gamma_{12}(\theta)|$ исследуемого излучения по измерениям видности $V(\Delta)$ интерференционных полос при различных значениях разности хода $\Delta = c\theta$. По расположению полос можно судить об аргументе комплексной функции $\gamma_{12}(\theta)$. Действительно, согласно (4.6) максимум порядка m наблюдается при условии $\delta + \alpha(\theta) = 2\pi m$, где $\delta = \frac{2\pi}{\lambda} \Delta = \frac{2\pi}{\lambda} \frac{xd}{l}$, d и l – параметры схемы Юнга, x – расстояние от центра интерференционной картины до m -го максимума. Измеряя эти расстояния, находим

$$\alpha(\theta) = 2\pi \left(m - \frac{xd}{\lambda l} \right).$$

Подобные измерения, однако, выполнить довольно сложно.

При одинаковых значениях интенсивностей интерферирующих волн ($I_1 = I_2$) функция видности $V(\Delta)$ и степень взаимной когерентности $|\gamma_{12}(\theta)|$ будут равны друг другу, $V(\Delta) = |\gamma_{12}(\theta)|$, при всех значениях $\Delta = c\theta$. В этом случае при $|\gamma_{12}(\theta)| = 0$ будет и $V(\Delta) = 0$, т.е. в случае полностью некогерентных световых волн интерференционные полосы не наблюдаются. Если же $|\gamma_{12}(\theta)| = 1$, то и $V(\Delta) = 1$. В этом случае имеет место полная когерентность световых волн с резко выраженной интерференционной картиной. Интерференционная картина получается такой, какой она имеет место в модели

монохроматических волн. Случай $0 < V(\Delta) < 1$ соответствует, очевидно, частичной когерентности. При $I_1 \neq I_2$ функции $V(\Delta)$ и $|\gamma_{12}(\theta)|$ также одновременно достигают своего максимального и минимального значения, но эти значения для функции $V(\Delta)$ не равны соответствующим значениям для функции $|\gamma_{12}(\theta)|$.

Из формулы (4.7) видно, что видность интерференционной картины $V(\Delta)$ при любой разности хода Δ зависит также от соотношения интенсивностей I_1 и I_2 . Положив в этой формуле $|\gamma_{12}(\theta)| = 1$, и введя отношение интенсивностей $p = I_1 / I_2$, получим

$$V(\Delta) = 2 \frac{\sqrt{p}}{p+1}. \quad (4.8)$$

Из расчета по этой формуле величины $V(\Delta)$ следует, что при $p = 1$ (равных интенсивностях) $V(\Delta) = 1$, при $p = 2$ $V(\Delta) = 0,94$, при $p = 3$ $V(\Delta) = 0,87$ и т.д. Как видим, даже при трехкратном различии интенсивностей величина V близка к единице, когда γ_{12} в точности равна единице. Поэтому неравенство интенсивностей интерферирующих волн не оказывает решающего влияния на контраст интерференционной картины.

Сравнивая формулу (4.7) с формулой (3.23), определяющей видность интерференционных полос при полной когерентности, видим, что степень взаимной когерентности колебаний $|\gamma_{12}(\theta)|$ показывает, во сколько раз контраст интерференционной картины при частичной когерентности меньше ее контраста при полной когерентности. Следовательно, функцию $|\gamma_{12}(\theta)|$ можно трактовать и как меру уменьшения контраста при интерференции немонохроматических волн.

4.2. Временная когерентность

Рассмотрим случай, когда точки S_1 и S_2 совпадают. В этом случае говорят о когерентности колебаний полей, создаваемых в точке P одним и тем же источником в разные моменты времени t и $t - \theta$. Такая ситуация возникает, например, когда обе волны, испущенные одним и тем же источником, приходят в точку P разными путями. Способность световых волн к интерференции после разделения пучка света на два пучка и последующего их наложения с некоторой разностью хода называется временной когерентностью. Поэтому рассматриваемый случай соответствует временной когерентности. Величину

$$\Gamma_{11}(\theta) = \langle E_1(t) \cdot E_1^*(t - \theta) \rangle \quad (4.9)$$

называют комплексной функцией самокогерентности (или автокорреляционной функцией), а нормированную величину

$$\gamma_{11}(\theta) = \frac{\langle E_1(t) \cdot E_1^*(t - \theta) \rangle}{\langle |E_1(t)|^2 \rangle} \quad (4.10)$$

– комплексной степенью временной когерентности. Функция $\gamma_{11}(\theta)$ характеризует зависимость видности полос V от разности хода $\Delta = c\theta$ двух интерферирующих волн, возбуждаемых колебаниями в точке S_1 . При $\theta = 0$ величина

$$\Gamma_{11}(0) = \langle E_1(t) \cdot E_1^*(t) \rangle = \langle |E_1(t)|^2 \rangle$$

будет характеризовать когерентность процессов в один и тот же момент времени t . Она определяет интенсивность света I_1 , излучаемого источником S_1 , а величина $\gamma_{11}(0)$, как видно из (4.10), равна единице. С учетом этого выражение для комплексной степени временной когерентности можно переписать в виде

$$\gamma_{11}(\theta) = \frac{\Gamma_{11}(\theta)}{\Gamma_{11}(0)} = \frac{\Gamma_{11}(\theta)}{I_1}. \quad (4.11)$$

Вычислим степень временной когерентности в предположении, что излучение источника света представляет собой беспорядочно следующие друг за другом незатухающие пуги волн. Принимая длительность пуга τ за время усреднения, получим

$$\langle |E_1(t)|^2 \rangle = \frac{1}{\tau} \int_0^\tau E_1(t) E_1^*(t) dt = |E_{10}|^2.$$

Учитывая далее, что произведение $E_1(t)E_1^*(t-\theta)$ отлично от нуля только в промежутке $[\theta, \tau]$, получим

$$\begin{aligned} \langle E_1(t) \cdot E_1^*(t-\theta) \rangle &= |E_{10}|^2 \frac{1}{\tau} \int_\theta^\tau \exp(-i\omega_0\theta) e^{-i\omega_0(t-\theta)} dt = \\ &= |E_{10}|^2 e^{-i\omega_0\theta} \frac{\tau-\theta}{\tau}. \end{aligned}$$

Тогда, в соответствии с (4.10), будем иметь для комплексной степени временной когерентности

$$\gamma_{11}(\theta) = \begin{cases} (1-\theta/\tau) e^{-i\omega_0\theta}, & \theta < \tau; \\ 0, & \theta \geq \tau. \end{cases}$$

Откуда находим степень временной когерентности:

$$|\gamma_{11}(\theta)| = \sqrt{\gamma_{11}(\theta)\gamma_{11}^*(\theta)} = \begin{cases} 1-\theta/\tau, & \theta < \tau; \\ 0, & \theta \geq \tau. \end{cases}$$

Используя треугольную функцию

$$\Lambda(x) = \begin{cases} 1 - |x|, & |x| < 1; \\ 0, & |x| \geq 1, \end{cases}$$

выражение для степени временной когерентности $|\gamma_{11}(\theta)|$ можно записать короче: $|\gamma_{11}(\theta)| = \Lambda(\theta/\tau)$. Такой же вид имеет и видность полос V :

$$V(\Delta) = |\gamma_{11}(\theta)| = \Lambda(\theta/\tau).$$

Степень когерентности и видность полос, определяемые функцией $y(\theta) = 1 - \theta/\tau$, равны единице при $\theta = 0$, монотонно убывают с ростом θ и обращаются в нуль при значении времени задержки θ , равном длительности цуга τ . Такой ход зависимости $|\gamma_{11}(\theta)|$ и $V(\Delta)$ физически очевиден. При $\theta = 0$ ($\Delta = 0$) обе части каждого цуга, испускаемого источником и раздвоенного в точке деления, приходят в точку встречи P одновременно и полностью накладываются друг на друга. Интерференционная картина при этом имеет наивысшую четкость (видность). С увеличением времени задержки θ часть цуга, идущая по более длинному пути, будет отставать от части цуга, идущей по более короткому пути, вследствие чего в точке P будет происходить их частичное наложение. Это приведет к ухудшению контрастности полос. Если время задержки θ превысит длительность цуга τ , то часть цуга, идущая по короткому пути, минует точку встречи раньше, чем другая часть цуга достигнет ее. Произойдет наложение частей цугов, порожденных цугами, испущенными разными атомами либо одним и тем же атомом, но в разные моменты времени, а эти цуги не когерентны. Поэтому, когда время задержки θ станет не меньше длительности цуга τ , т.е. когда $\theta \geq \tau$, интерференционная картина не возникает – интерференция не наблюдается.

Из проведенного рассмотрения следует, что в указанной модели излучения источника света за время когерентности $t_{\text{ког}}$ можно принять промежуток времени, примерно равный длительности цуга τ , т.е. считать, что $t_{\text{ког}} = \tau$. Тогда за длину когерентности $l_{\text{ког}}$ можно принять длину цуга $c\tau$: $l_{\text{ког}} = c\tau$.

Длительность цуга τ связана с шириной его спектрального интервала $\Delta\omega$ соотношением неопределенностей $\tau \approx 2\pi/\Delta\omega$. Перейдя от круговой частоты ω к линейной частоте ν , это соотношение можно записать в виде $\tau \approx 1/\Delta\nu$. Следовательно, время когерентности

$$t_{\text{ког}} = \tau \approx 1/\Delta\nu.$$

Ширина спектрального интервала цуга определяет ширину спектрального интервала всего источника. Поэтому такое же соотношение получается и при другом определении времени когерентности, указанном на стр. 167. Действительно, в этом случае $t_{\text{ког}}\Delta\omega = \pi$. Опуская несущественный множитель $1/2$, приходим к соотношению $t_{\text{ког}} \approx 1/\Delta\nu$.

На практике обычно имеют дело не со временем, а с длиной когерентности. Эта величина связана с шириной спектрального интервала соотношением $l_{\text{ког}} = c\tau = 1/\Delta\nu$. В этом выражении, как и в предыдущем, под шириной спектрального интервала следует понимать ширину спектрального интервала излучения источника квазимонохроматического света, так как спектральная ширина излучения реального квазимонохроматического источника определяется значением спектральной ширины отдельного цуга, испускаемого источником.

Длину когерентности удобно выразить через ширину интервала длин волн излучения источника $\Delta\lambda$. Так как $\Delta\nu = \Delta(c/\lambda) = -(c/\lambda^2)\Delta\lambda$ (знак минус опущен, так как он только показывает, что с ростом длины волны λ частота ν убывает), то получим

$$l_{\text{ког}} = \frac{\lambda^2}{\Delta\lambda}, \quad (4.12)$$

где λ – среднее значение длины волны из спектрального интервала $\Delta\lambda$. Длине когерентности соответствует максимальное число интерференционных полос

$$m_{\text{max}} = l_{\text{ког}} / \lambda = \lambda / \Delta\lambda.$$

Подставив в эту формулу $l_{\text{ког}} = c t_{\text{ког}} = c\tau$, $\lambda = cT_0$, где $T_0 = 2\pi/\omega_0$ – период колебания поля, получим

$$m_{\text{max}} = \tau / T_0.$$

Следовательно, максимальный порядок интерференции равен числу периодов колебаний, содержащихся в цуге излучения. При этом поскольку $\tau \gg T_0$, то когерентность сохраняется на протяжении большого числа периодов колебаний поля.

Формула (4.12) показывает, что некогерентность световых волн приводит к ограничению разности их хода для наблюдения интерференции. Чем больше ширина спектрального интервала $\Delta\lambda$, тем меньше эта разность хода и тем, следовательно, труднее наблюдать интерференцию. С другой стороны, если бы световая волна была строго монохроматической ($\Delta\lambda = 0$), то длина когерентности $l_{\text{ког}} = \infty$, и интерференция наблюдалась бы при любой разности хода. Для обычных (не лазерных) источников света относительная ширина спектрального интервала $\Delta\lambda / \lambda \sim \sim 10^{-3} - 10^{-5}$, так что длина когерентности имеет порядок 0,1 мм – 1 см. Для белого света средняя длина волны $\lambda \approx 0,55$ мкм, а $\Delta\lambda = 0,76 - 0,38 = 0,38$ мкм, поэтому длина когерентности $l_{\text{ког}} \approx 1$ мкм, а $m_{\text{max}} \approx 1$. Вот почему интерференцию белого света получить очень трудно. Для лазерного излучения относительная ширина спектрального интервала $\Delta\lambda/\lambda$ может быть 10^{-11} и более, чему соответствует длина когерентности $l_{\text{ког}} \sim 10^5$ и более. Поэтому интерференция света от лазерного источника

наблюдается практически при любой разности хода. Более того, в экспериментах с лазерами для получения двулучевой интерференции можно использовать не один источник, а два отдельных лазерных источника.

Мы видим, что временная когерентность связана с некогерентностью света. Интерференция раздвоенного, а затем сведенного в одну точку пучка возможна, если разность хода не будет превышать длину когерентности:

$$\Delta \leq l_{\text{ког.}}$$

Это неравенство выражает собой так называемое условие временной когерентности.

4.3. Спектральный подход к рассмотрению временной когерентности

Изложенный выше подход к рассмотрению временной когерентности можно назвать временным. Возможен и иной подход, основанный на представлении излучения источника света в виде совокупности бесконечно большого числа гармонических составляющих, т.е. на спектральном разложении излучения. При этом, как и выше, мы будем считать волны квазимонохроматическими с шириной частотного спектрального интервала $\Delta\nu$.

Рассмотрим сначала случай, когда частотный спектр излучения источника состоит всего из двух частот $\nu_1 = \nu_0 - \Delta\nu/2$ и $\nu_2 = \nu_0 + \Delta\nu/2$, т.е. будем считать, что источник света состоит из двух элементарных излучателей на частотах ν_1 и ν_2 . В двулучевой интерференции каждая такая спектральная составляющая (элементарный излучатель) создает свою интерференционную картину, распределение интенсивности в которой описывается формулой

$$I(\Delta) = 2I_0 (1 + \cos k\Delta) = 2I_0 \left(1 + \cos \frac{2\pi\nu}{c} \Delta \right)$$

с соответствующими значениями I_0 и ν , где I_0 – интенсивность вторичной волны в дулучевой интерференционной схеме. Разность хода Δ вторичных волн для двух совпадающих в пространстве источников излучения одинакова. Будем считать также, что интенсивности вторичных волн для обоих элементарных источников одинаковы и равны I_0 . Так как источники излучают некогерентно ($\nu_1 \neq \nu_2$), то, в соответствии с правилом сложения интенсивностей при наложении волн от двух некогерентных источников, полная интенсивность при наложении двух интерференционных картин составит

$$\begin{aligned} I(\Delta) &= 2I_0 \left(1 + \cos \frac{2\pi\nu_1}{c} \Delta \right) + 2I_0 \left(1 + \cos \frac{2\pi\nu_2}{c} \Delta \right) = \\ &= 4I_0 \left[1 + \cos \frac{\pi(\nu_2 - \nu_1)}{c} \Delta \cos \frac{\pi(\nu_2 + \nu_1)}{c} \Delta \right] = \\ &= 4I_0 \left(1 + \cos \frac{\pi \Delta \nu}{c} \Delta \cos \frac{2\pi\nu_0}{c} \Delta \right) \end{aligned}$$

или

$$I(\Delta) = 4I_0 \left(1 + \cos \frac{\pi \Delta \nu}{c} \Delta \cos k \Delta \right), \quad (4.13)$$

где $\nu_0 = (\nu_1 + \nu_2) / 2$ – среднее значение частоты, а $\Delta \nu = \nu_2 - \nu_1$ – разность частот излучений источников, $k = (2\pi\nu_0 / c)$ – среднее значение волнового числа.

Второй множитель ($\cos k\Delta$) в скобках в выражении (4.13) изменяется быстро и определяет ширину интерференционных полос. Первый множитель ($\cos[(\pi\Delta\nu/c)\Delta]$) ввиду малости $\Delta\nu$ меняется медленно (по сравнению со вторым множителем),

оставаясь практически постоянным на протяжении многих интерференционных полос. Этот множитель и определяет видность, т.е. контрастность полос. Вследствие сказанного максимальное и минимальное значение интенсивности определяется соотношением $\cos k\Delta = \pm 1$ (если первый множитель брать по модулю). Тогда экстремальные значения интенсивности

$$I_{\max} = 4I_0 \left(1 + \left| \cos \frac{\pi \Delta \nu}{c} \Delta \right| \right), \quad I_{\min} = 4I_0 \left(1 - \left| \cos \frac{\pi \Delta \nu}{c} \Delta \right| \right).$$

Подставив эти значения в формулу (4.13), получим выражение для видности интерференционных полос:

$$V(\Delta) = \left| \cos \frac{\pi \Delta \nu}{c} \Delta \right|$$

или, используя формулу времени задержки $\theta = \Delta / c$,

$$V(\theta) = \left| \cos \pi \Delta \nu \theta \right|.$$

Как видим, видность интерференционных полос в этом случае является периодической функцией времени задержки θ и разности частот $\Delta \nu$

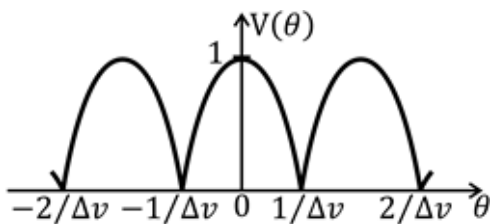


Рис. 4.1

вторичных волн. Зависимость видности полос от времени задержки показана на рис. 4.1.

Функция видности $V(\theta)$ принимает свое максимальное

значение, равное единице, при $\theta = 0$ ($\Delta = 0$), когда $I_{\max} = 8I_0$, а $I_{\min} = 0$. С увеличением времени задержки видность уменьшается и при $\pi \Delta \nu \theta = \pi/2$, т.е. при $\theta = 1 / 2\Delta \nu$, обращается в нуль. Это

соответствует случаю, когда светлые полосы интерференционной картины от одного источника совмещаются с темными полосами от другого источника (так как при указанном условии $I_{\max} = I_{\min}$). При дальнейшем увеличении времени задержки θ видность полос возрастает и при $\theta = 1 / \Delta \nu$ (или при разности хода $\Delta = c / \Delta \nu$) она снова равна единице. Полный период изменения видности полос охватывает $N = \nu_0 / \Delta \nu$ интерференционных полос.

Перейдем теперь к рассмотрению квазимонохроматического источника света с шириной спектрального интервала $\Delta \nu$. Такой источник можно рассматривать как бесконечное число элементарных излучателей, частоты излучения которых заключены в интервале частот от $\nu_0 - \Delta \nu / 2$ до $\nu_0 + \Delta \nu / 2$.

Любой спектральный прибор измеряет интенсивность света в некотором достаточно малом спектральном интервале, шириной $d\nu$, которая может быть записана в виде $I(\nu)d\nu$. Здесь $I(\nu)$ – спектральная плотность интенсивности. Эта функция характеризует распределение интенсивности по спектру частот. Поскольку различные спектральные составляющие не когерентны между собой, то на основании закона аддитивного сложения интенсивностей для интенсивности света, излучаемого источником, можно записать

$$I = \int_0^{\infty} I(\nu) d\nu.$$

В случае квазимонохроматического источника интегрирование следует проводить в пределах его интервала частот ($\nu_0 - \Delta \nu / 2, \nu_0 + \Delta \nu / 2$), где ν_0 – средняя частота излучения источника.

Распределение интенсивности, возникающее в любой интерференционной схеме, есть результат наложения независимых интерференционных картин, полученных в монохроматическом свете от каждой спектральной составляющей.

Предполагая, что квазимонохроматическое излучение в бесконечно узком спектральном интервале с хорошим приближением можно считать монохроматическим, распределение интенсивности в элементарной интерференционной картине можно получить с помощью соотношения $I = 2I_0 (1 + \cos k\Delta)$ (формула (3.20)). Подставив в эту формулу dI вместо I , $I(\nu)d\nu$ вместо I_0 и $2\pi\nu / c$ вместо k , получим

$$dI = 2I(\nu) \left(1 + \cos \frac{2\pi\nu}{c} \Delta \right) d\nu, \quad (4.14)$$

где dI представляет собой интенсивность, возникающую в результате наложения двух монохроматических компонент частоты ν и интенсивности $I(\nu)d\nu$ (интенсивность бесконечно узкого спектрального интервала $d\nu$ данного квазимонохроматического излучения). Проинтегрировав это выражение по всем спектральным компонентам, получим распределение интенсивности в результирующей интерференционной картине:

$$I = 2 \int_0^{\infty} I(\nu) \left(1 + \cos \frac{2\pi\nu}{c} \Delta \right) d\nu. \quad (4.15)$$

На рис. 4.2 представлено распределение интенсивности $I(\nu)$ в квазимонохроматическом излучении (спектральной линии) с мак-

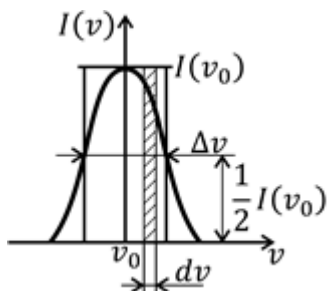


Рис. 4.2

симумом на частоте ν_0 . Ширина спектральной линии на уровне, соответствующем половине максимальной интенсивности $I(\nu_0)$, равна $\Delta\nu$.

Аппроксимируем это распределение прямоугольником (прямоугольным контуром) высотой $I(\nu_0)$ и шириной $\Delta\nu$, который дает по спектральной плотности интенсивности такое же излучение, как и распределение $I(\nu)$.

Эта аппроксимация упрощает теоретический анализ, не меняя заметно физической сущности явления. Для определения интенсивности от всего квазимонохроматического источника выражение (4.14) следует проинтегрировать по частоте в пределах его полосы частот излучения (от $\nu_0 - \Delta\nu/2$ до $\nu_0 + \Delta\nu/2$):

$$I = 2I(\nu_0) \int_{\nu_0 - \Delta\nu/2}^{\nu_0 + \Delta\nu/2} \left(1 + \cos \frac{2\pi\nu\Delta}{c} \right) d\nu.$$

После интегрирования, получим

$$I = 2I(\nu_0)\Delta\nu + 2I(\nu_0)\frac{c}{2\pi\Delta} \left[\sin \frac{2\pi(\nu_0 + \Delta\nu/2)\Delta}{c} - \sin \frac{2\pi(\nu_0 - \Delta\nu/2)\Delta}{c} \right],$$

или после преобразований

$$I = 2I(\nu_0)\Delta\nu \left(1 + \frac{\sin \pi(\Delta/c)\Delta\nu}{\pi(\Delta/c)\Delta\nu} \cos \frac{2\pi\nu_0\Delta}{c} \right).$$

Это выражение представляет собой формулу интерференции от квазимонохроматического источника с прямоугольным контуром распределения интенсивности по спектру частот (с постоянным распределением в интервале частот от $\nu_0 - \Delta\nu/2$ до $\nu_0 + \Delta\nu/2$). Величина $2I(\nu_0)\Delta\nu$ представляет собой удвоенное

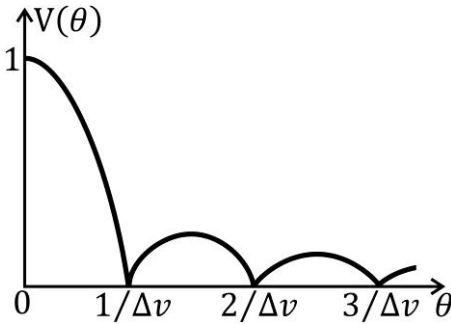


Рис. 4.3

значение интенсивности источника в рассматриваемой модели. Ввиду медленности изменения первого множителя во втором члене в скобках максимальное и минимальное значения интенсивности

Достигаются при $\cos(2\pi\nu \Delta / c) = \pm 1$

соответственно. Учитывая это и то, что $\Delta / c = \theta$, для функции видности полос, а значит, и для степени временной когерентности получим

$$V(\theta) = |\gamma_{11}(\theta)| = \left| \frac{\sin \pi\theta \Delta \nu}{\pi\theta \Delta \nu} \right| = |\text{sinc}(\theta \Delta \nu)|. \quad (4.16)$$

График функции $V(\theta)$ показан на рис. 4.3. Из графика видно, что видность интерференционных полос $V(\theta)$, а с ней и степень временной когерентности $|\gamma_{11}(\theta)|$ имеют максимальное значение, равное единице, при $\theta = 0$. С увеличением времени задержки θ эти характеристики убывают и при $\theta \Delta \nu = 1$, т.е. при $\theta = 1 / \Delta \nu$, $V(\theta)$ и $|\gamma_{11}(\theta)|$ обращаются в нуль. Это означает, что при времени задержки $\theta = 1 / \Delta \nu$ интерференционные полосы наблюдаться не будут. При дальнейшем увеличении времени задержки θ функции $V(\theta)$ и $|\gamma_{11}(\theta)|$ снова возрастают, оставаясь при этом значительно меньше единицы, затем убывают, обращаясь в нуль при $\theta = 2 / \Delta \nu$ и т.д. Значение времени задержки, при котором функции $V(\theta)$ и $|\gamma_{11}(\theta)|$ первый раз обращаются в нуль, когда когерентность исчезает, и принимается

за время когерентности $t_{\text{ког}}$. Тем самым мы снова приходим к формуле времени когерентности

$$t_{\text{ког}} = 1 / \Delta \nu.$$

С точки зрения спектрального подхода исчезновение когерентности при значении времени задержки $\theta = 1 / \Delta \nu$ можно объяснить как результат наложения большого числа интерференционных картин вида (4.14), полученных в монохроматическом свете. Существенную роль играет то обстоятельство, что интерференционные картины, образованные различными спектральными компонентами, имеют различный пространственный период (ширину полос).

4.4. Теорема Винера – Хинчина

Эта теорема устанавливает связь между степенью временной когерентности и обратным преобразованием Фурье от спектральной плотности интенсивности излучения источника. Для ее доказательства запишем выражения (4.5) и (4.15) для распределения интенсивности в интерференционной картине (с учетом того, что $\Delta / c = \theta$):

$$I(\theta) = 2 I_0 [1 + \text{Re } \gamma_{11}(\theta)],$$
$$I(\theta) = 2 \int_0^{\infty} I(\nu)(1 + \cos 2\pi\nu\theta) d\nu.$$

Первое из этих выражений, содержащее функцию временной когерентности, получено на основе временного рассмотрения. Второе выражение получено в результате спектрального подхода; оно выражает распределение интенсивности в

интерференционной картине через спектр излучения источника. Сравнивая эти выражения и учитывая при этом, что

$$\int_0^{\infty} I(\nu) d\nu = I_0,$$

получим

$$I_0 \operatorname{Re} \gamma_{11}(\theta) = \int_0^{\infty} I(\nu) \cos 2\pi\nu\theta d\nu.$$

Поскольку спектральная плотность интенсивности $I(\nu)$ – действительная функция частоты, а $\gamma_{11}(-\theta) = \gamma_{11}^*(\theta)$, то это соотношение можно переписать в виде

$$I_0 \operatorname{Re} \gamma_{11}(\theta) = \operatorname{Re} \int_{-\infty}^{\infty} I(\nu) e^{-i2\pi\nu\theta} d\nu,$$

откуда

$$\gamma_{11}(\theta) = \frac{\int_{-\infty}^{\infty} I(\nu) e^{-i2\pi\nu\theta} d\nu}{\int_{-\infty}^{\infty} I(\nu) d\nu}. \quad (4.17)$$

Таким образом, комплексная степень временной когерентности равна нормированному обратному преобразованию Фурье от спектральной плотности интенсивности излучения источника света. Это утверждение и составляет содержание теоремы Винера – Хинчина.

Применим полученную формулу для определения степени временной когерентности в рассмотренном выше случае, когда спектральная плотность интенсивности источника аппроксимируется прямоугольной функцией

$$I(\nu) = \begin{cases} I(\nu_0), & |\nu - \nu_0| \leq \Delta\nu/2; \\ 0, & |\nu - \nu_0| > \Delta\nu/2. \end{cases}$$

Имеем

$$\begin{aligned} \gamma_{11}(\theta) &= \frac{I(\nu_0) \int_{\nu_0 - \Delta\nu/2}^{\nu_0 + \Delta\nu/2} e^{-i2\pi\nu\theta} d\nu}{I(\nu_0) \int_{\nu_0 - \Delta\nu/2}^{\nu_0 + \Delta\nu/2} d\nu} = \frac{\sin \pi\theta \Delta\nu}{\pi\theta \Delta\nu} e^{-i2\pi\nu_0\theta} = \\ &= \text{sinc}(\theta \Delta\nu) e^{-i2\pi\nu_0\theta}, \end{aligned}$$

а $|\gamma_{11}(\theta)| = |\text{sinc}(\theta \Delta\nu)|$. Мы снова пришли к выражению (4.16). Таким образом, нормированное обратное преобразование Фурье от спектральной плотности интенсивности излучения источника приводит к тому же результату для степени временной когерентности, что и вычисление ее с помощью функции видности.

Поскольку, в соответствии с формулой (4.11), $\Gamma_{11}(\theta) = I_0 \gamma_{11}(\theta)$, то

$$\Gamma_{11}(\theta) = \int_{-\infty}^{\infty} I(\nu) e^{-i2\pi\nu\theta} d\nu. \quad (4.18)$$

Следовательно, автокорреляционная функция является обратным фурье-преобразованием спектральной интенсивности $I(\nu)$ источника света. Выполнив фурье-преобразование, получим спектральную интенсивность:

$$I(\nu) = \int_{-\infty}^{\infty} \Gamma_{11}(\theta) e^{i2\pi\nu\theta} d\theta. \quad (4.19)$$

Формулы (4.18) и (4.19) можно рассматривать как математические выражения, связывающие два метода исследования излучения – спектральный и интерференционный. При спектральных измерениях можно получить информацию о спектральной интенсивности $I(\nu)$ и, применив обратное преобразование Фурье от этой функции, найти $\Gamma_{11}(\theta)$. С помощью двухлучевой интерференционной схемы можно определить автокорреляционную функцию $\Gamma_{11}(\theta)$ и затем, взяв от нее преобразование Фурье, найти спектральную плотность интенсивности излучения источника. Функции $I(\nu)$ и $\Gamma_{11}(\theta)$, связанные между собой фурье-преобразованием, несут одинаково полную информацию об излучении.

Вернемся теперь снова к интерферометру Майкельсона как к фурье-спектрометру и рассмотрим случай освещения его сложным некогерентным излучением со спектральной плотностью интенсивности $I(\nu)$. Каждая монохроматическая составляющая дает в сигнал на выходе вклад, определяемый выражением (3.31). Полный выходной сигнал получим, проинтегрировав это выражение по всем частотам:

$$I(t) = \int_0^{\infty} I(f)[1 + \cos 2\pi ft] df = \int_0^{\infty} I(f) df + \int_0^{\infty} I(f) \cos 2\pi ftdf .$$

Первое слагаемое в правой части этого выражения представляет собой постоянную составляющую выходного сигнала, зависящую от интенсивности излучения источника. Сигнал, соответствующий этому слагаемому, отфильтровывается фильтром, установленным на выходе прибора. Поэтому сигнал на выходе $I(t)$ будет определяться вторым слагаемым, представляющим собой обратный фурье-образ функции $I(f)$:

$$I(t) = \int_{-\infty}^{\infty} I(f) \exp(-i2\pi f t) df .$$

Вид получаемой сложной интерферограммы на выходе прибора будет зависеть от спектрального распределения излучения источника света. Поэтому, измерив функцию $I(t)$, с помощью преобразования Фурье можно определить спектральную плотность интенсивности излучения источника на частотах $f = (2\nu/c)\nu$:

$$I(f) = \int_{-\infty}^{\infty} I(t) \exp(i2\pi f t) dt.$$

Получив распределение интенсивности $I(f)$ на частотах f , можно рассчитать спектральную плотность интенсивности на частотах $\nu = (c/2\nu)f$ излучения источника, т.е. функцию $I(\nu)$. Все операции осуществляются с помощью компьютера.

Зависимость интерференционной картины от спектрального состава излучения при периодическом изменении разности хода и лежит в основе использования интерферометра Майкельсона в качестве фурье-спектрометра.

4.5. Пространственная когерентность

Рассмотрим теперь случай, когда точки S_1 и S_2 , о которых шла речь в п. 4.2, не совпадают, но время задержки $\theta = 0$. Будем считать, что эти точки расположены на псевдодолновой поверхности какого-либо протяженного источника (поверхности, которая является огибающей волновых поверхностей волн, излучаемых всеми точками протяженного источника) и в соответствии с принципом Гюйгенса являются вторичными источниками. В этом случае величина

$$\gamma_{12}(0) = \frac{\langle E_1(t) \cdot E_2^*(t) \rangle}{\sqrt{\langle |E_1(t)|^2 \rangle} \cdot \sqrt{\langle |E_2(t)|^2 \rangle}} = \frac{\langle E_1(t) \cdot E_2^*(t) \rangle}{\sqrt{I_1} \cdot \sqrt{I_2}}$$

будет характеризовать корреляцию колебаний в разных пространственных точках S_1 и S_2 в один и тот же момент времени. Этот случай соответствует пространственной когерентности. Под пространственной когерентностью понимают способность к интерференции световых волн, исходящих из двух пространственно разделенных точек, взятых в поперечном сечении пучка (на псевдодолговой поверхности). Величину $\gamma_{12}(0)$ называют комплексной степенью пространственной когерентности световых колебаний в двух различных точках, а ее модуль $|\gamma_{12}(0)|$ – степенью пространственной когерентности. Степень пространственной когерентности поля в двух точках характеризует видность интерференционной картины, получаемой в экспериментах, где эти точки являются вторичными источниками. При $|\gamma_{12}(0)| = 1$ говорят о полной пространственной когерентности, при $|\gamma_{12}(0)| = 0$ – об ее отсутствии, а при $0 < |\gamma_{12}(0)| < 1$ – о частичной пространственной когерентности.

Комплексную величину

$$I_{12} = \Gamma_{12}(0) = \langle E_1(t) \cdot E_2^*(t) \rangle,$$

определяющую корреляцию колебаний в двух разных точках в один и тот же момент времени, называют взаимной интенсивностью. Она составляется из двух функций $E(t)$, характеризующих световые колебания в точках S_1 и S_2 в моменты, когда колебания, одновременно (в момент t) доходящие до точки наблюдения P , только вышли из точек S_1 и S_2 . По конструкции функция $\Gamma_{12}(0)$ напоминает выражение интенсивности через комплексные амплитуды, но существенная

особенность выражения $\Gamma_{12}(0)$ состоит в том, что функции E_1 и E_2^* рассматриваются для разных точек. Это позволяет учесть степень взаимосвязи (корреляции) между световыми колебаниями в точках S_1 и S_2 .

Пространственная когерентность связана с неточечностью источника света (временная когерентность, как мы видели, связана с

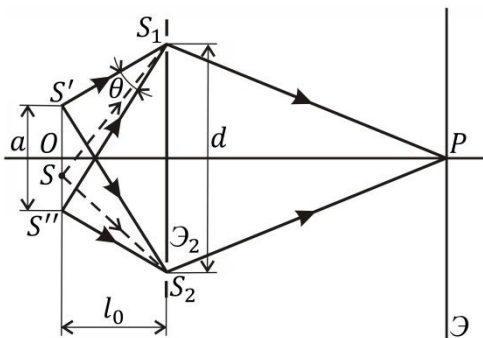


Рис. 4.4

ей точечный источник S на протяженный источник $S'S''$ (рис. 4.4). Излучение этого источника будем предполагать монохроматическим, т.е. имеющим неограниченную длину когерентности. Будем рассматривать протяженный источник, состоящий из множества точечных излучателей со случайными начальными фазами. Точечные же отверстия S_1 и S_2 в плоском экране \mathcal{E}_2 мы будем рассматривать в качестве произвольных точек S_1 и S_2 на псевдодолновой поверхности. Поскольку рассматривается случай, когда $\theta = 0$, т.е. $\Delta \approx 0$, то точка наблюдения P находится на одинаковом расстоянии от отверстий S_1 и S_2 , т.е. на оптической оси интерференционной схе мы. При этих предположениях результат интерференции в точке P будет зависеть только от разности фаз колебаний в точках S_1 и S_2 . Видность интерференционных полос в этом случае будет

его
немонохроматичность
ю). Здесь нас будет
интересовать влияние
пространственной
когерентности, т.е.
конечных размеров
источника света, на
интерференционную
картину. Обратимся
снова к
интерференционной
схеме Юнга, заменив в

определять пространственную когерентность колебаний поля в точках S_1 и S_2 : $V(\Delta) = |\gamma_{12}(0)|$.

Возьмем какую-либо точку протяженного источника SS'' , например точку S . Свет, исходящий из этой точки, освещает отверстия S_1 и S_2 и на экране \mathcal{E} в окрестности точки P возникают интерференционные полосы. В зависимости от расстояния ζ точки S от оптической оси схемы существует определенная разность хода $\Delta(\zeta)$, величина которой и определяет положение максимумов и минимумов интенсивности полос на экране \mathcal{E} . Если бы источник состоял из одной только этой точки (как в идеализированном опыте Юнга), то наблюдались бы полосы с максимальной видностью. Однако реальный источник имеет конечный размер, и полосы, созданные излучением из других точек, будут смещены относительно полос, созданных излучением из S , и вследствие некогерентности этих элементарных излучений просто накладываются друг на друга. В результате такого наложения интерференционные полосы либо вовсе исчезают (если максимумы одной картины наложатся на минимумы другой), либо оказываются более или менее размытыми: видность интерференционных полос будет уменьшаться.

Предположим сначала, что протяженный источник состоит всего из двух одинаковых некогерентных светящихся точек S' и S'' , расположенных симметрично относительно оптической оси схемы на расстоянии a друг от друга и на расстоянии l_0 до экрана с отверстиями S_1 и S_2 . Каждый точечный источник создает на экране свою интерференционную картину. Так как при смещении источника перпендикулярно оптической оси центр интерференционной картины смещается в противоположном направлении, то центр интерференционной картины от источника S' будет смещен в точку P' , расположенную ниже оптической оси на расстоянии x от точки P . Центру интерференционной картины соответствует разность хода $\Delta = 0$, поэтому для источника S' будем иметь

$$\Delta = S'S_2P' - S'S_1P' = \Delta_1 + \Delta_2 = 0,$$

где

$$\Delta_1 = S'S_2 - S'S_1, \quad \Delta_2 = S_2P' - S_1P'.$$

Разности хода Δ_1 и Δ_2 можно найти, например, подобно тому, как была получена формула (3.23). Принимая, что $l_0 \gg a$, с помощью рис. 4.4 находим

$$\Delta_1 = (a / 2l_0)d, \quad \Delta_2 = -(x / l)d,$$

где l – расстояние от экрана с отверстиями S_1 и S_2 до экрана \mathcal{E} , где наблюдается интерференционная картина. Приравнявая их сумму нулю, получим величину смещения центра интерференционной картины от источника S' :

$$x = (l / 2l_0) a.$$

Центр интерференционной картины от источника S'' смещен вверх от точки P , причем на то же расстояние x , так как он расположен симметрично источнику S' . Если бы источник S' или S'' располагался на оптической оси, то он создавал бы интерференционную картину с центром в точке P и с распределением интенсивности, описываемым функцией (3.20). С учетом смещения центра интерференционной картины вниз, распределение интенсивности, даваемое источником S' , будет описываться функцией

$$I'(x) = 2I_0 \left\{ 1 + \cos \left[\frac{2\pi}{h} \left(x + \frac{l}{2l_0} a \right) \right] \right\}.$$

Распределение интенсивности в интерференционной картине от источника S'' с учетом смещения ее центра вверх будет описываться функцией

$$I''(x) = 2I_0 \left\{ 1 + \cos \left[\frac{2\pi}{h} \left(x - \frac{l}{2l_0} a \right) \right] \right\}.$$

В обеих формулах $h = \lambda l / d$ – ширина интерференционной полосы, I_0 – интенсивность вторичной волны, прошедшей через одно из отверстий. Для обоих источников величина I_0 предполагается одинаковой. Так как источники S' и S'' не являются когерентными, то интерференция не происходит. В каждой точке экрана происходит сложение не комплексных амплитуд полей, а их интенсивностей. Следовательно, результирующая интенсивность

$$I(x) = 2I_0 \left\{ 1 + \cos \left[\frac{2\pi}{h} \left(x + \frac{l}{2l_0} a \right) \right] \right\} + 2I_0 \left\{ 1 + \cos \left[\frac{2\pi}{h} \left(x - \frac{l}{2l_0} a \right) \right] \right\}. \quad (4.20)$$

После преобразований этого выражения находим распределение интенсивности в результирующей интерференционной картине:

$$I(x) = 4I_0 \left(1 + \cos \frac{\pi ad}{\lambda l_0} \cos \frac{2\pi}{h} x \right).$$

При перемещении точки наблюдения P вдоль экрана множитель $\cos(2\pi/h)x$ быстро осциллирует. Он определяет ширину интерференционных полос. Первый сомножитель при этом изменяется медленно. Этот сомножитель определяет видность интерференционных полос. Максимумы и минимумы интенсивности возникают в тех местах, где $\cos(2\pi/h)x = \pm 1$. Их значения соответственно равны

$$I_{\max} = 2I_0 \left(1 + 2 \left| \cos \frac{\pi a d}{\lambda l_0} \right| \right) \text{ и } I_{\min} = 2I_0 \left(1 - 2 \left| \cos \frac{\pi a d}{\lambda l_0} \right| \right).$$

С учетом этого для видности интерференционных полос получаем

$$V(\Delta) = \left| \cos \frac{\pi a d}{\lambda l_0} \right|.$$

Мы видим, что в зависимости от расстояния a между источниками видность периодически изменяется, достигая максимального значения, равного единице, при a , удовлетворяющем условию $\pi a d / \lambda l_0 = m\pi$, т.е. при $a = m\lambda (l_0 / d)$, и минимального значения, равного нулю, при $a = (m + 1 / 2)\lambda (l_0 / d)$, где m – любое целое положительное число, включая нуль. Видность достигает максимума, когда накладываются друг на друга светлые полосы, и минимума, когда светлые полосы накладываются на темные. Подобная зависимость видности интерференционных полос имеет место и от расстояния d между отверстиями S_1 и S_2 .

Вернемся теперь к рассмотрению интерференции света от протяженного источника $S' S''$ и найдем зависимость видности интерференционных полос $V(\Delta)$ от его размеров. Для простоты будем считать, что этот источник представляет собой бесконечно длинную равномерно светящуюся полосу (или светящуюся щель) шириной a . Совместим с плоскостью полоски (щели) координатную плоскость XU , направив ось U вдоль полоски и поместив начало координат в ее центр. Разделим мысленно весь источник на элементарные полоски шириной $d\xi$. Так как источник излучает равномерно, то интенсивность излучения элементарной полоски будет $I_0 d\xi / a$, где I_0 – интенсивность излучения всего протяженного источника. Интенсивность dI света в интерференционной картине, создаваемой одной элементарной полоской $d\xi$, можно определить с помощью соотношения (3.20),

если в нем I_0 заменить на $I_0 d\xi/a$ и подставить выражение для разности фаз $\delta(\xi)$ от элемента $d\xi$, находящегося на расстоянии ξ от середины источника, до точки P для двух интерферирующих лучей. Эта разность фаз может быть представлена в виде

$$\delta(\xi) = \frac{2\pi}{h} \left(x - \frac{l}{l_0} \xi \right),$$

где l_0 – расстояние от источника света до экрана с отверстиями S_1 и S_2 , остальные обозначения соответствуют обозначениям выражения (4.20). С учетом этого для элементарной интенсивности dI будем иметь

$$dI = \frac{2I_0}{a} \left\{ 1 + \cos \left[\frac{2\pi}{h} \left(x - \frac{l}{l_0} \xi \right) \right] \right\} d\xi.$$

Распределение интенсивности в интерференционной картине от всего протяженного источника найдется суммированием интенсивностей от отдельных элементарных полосок:

$$\begin{aligned} I(x) &= \frac{2I_0}{a} \int_{-a/2}^{a/2} \left\{ 1 + \cos \left[\frac{2\pi}{h} \left(x - \frac{l}{l_0} \xi \right) \right] \right\} d\xi = \\ &= 2I_0 \left\{ 1 + \frac{\sin(kad/2l_0)}{kad/2l_0} \cos \frac{2\pi}{h} x \right\}. \end{aligned}$$

В этом выражении второй множитель в скобках $(\cos(2\pi/h)x)$ при изменении координаты x быстро осциллирует, описывая изменение интенсивности при переходе от одной интерференционной полосы к другой. Первый же множитель изменяется медленно, давая плавную огибающую осцилляций второго множителя. Он и определяет видность интерференционных полос. Учитывая, что максимальное и

минимальные значения интенсивности достигаются при $\cos(2\pi/h)x = \pm 1$, находим

$$V(a) = \left| \frac{\sin(kad/2l_0)}{kad/2l_0} \right| = \left| \frac{\sin(\pi ad/\lambda l_0)}{\pi ad/\lambda l_0} \right| = \left| \operatorname{sinc}\left(\frac{ad}{\lambda l_0}\right) \right|,$$

или

$$V(\theta) = \left| \operatorname{sinc}\left(\frac{\theta d}{\lambda}\right) \right|.$$

Здесь $\theta \approx a / l_0$ – угол (в радианах), под которым виден поперечный размер светящейся полоски $S' S''$ из точек S_1 и S_2 (угловая ширина источника света). Этим же соотношением определяется и степень пространственной когерентности колебаний в точках S_1 и S_2 , находящихся на расстоянии d друг от друга:

$$|\gamma_{12}(0)| = V(a) = \left| \operatorname{sinc}\left(\frac{ad}{\lambda l_0}\right) \right|, \quad (4.21)$$

или

$$|\gamma_{12}(0)| = V(\theta) = \left| \operatorname{sinc}\left(\frac{\theta d}{\lambda}\right) \right|. \quad (4.22)$$

Графики зависимости степени пространственной когерентности $|\gamma_{12}(0)|$ и видности интерференционных полос $V(\theta)$ от величины $\theta d / \lambda$ имеют такой же вид, как и графики зависимости тех же характеристик от времени задержки θ . В соответствии с изменением модуля sinc -функции, видность интерференционных полос и степень пространственной когерентности имеют максимальное значение, близкое к единице, при очень малом угловом размере θ источника света, т.е. когда

источник точечный. С увеличением углового размера источника функции $|\gamma_{12}(0)|$ и $V(\theta)$ убывают и обращаются в нуль при значении $\theta d / \lambda = 1$, т.е. при угловом размере источника $\theta = \lambda / d$. (Значению $V = 2 / 3$ соответствует $\theta d / \lambda = 1 / 2$ и угловой размер источника $\theta = \lambda / 2d$.)

Мы видим, что протяженный источник может создать когерентное световое поле, несмотря на то, что он состоит из многих миллионов некогерентных излучателей, если угловые размеры источника, рассматриваемого из плоскости, где расположены отверстия S_1 и S_2 , меньше, чем отношение длины волны света к расстоянию между отверстиями.

Такой же характер зависимости имеют степень пространственной когерентности колебаний в точках S_1 и S_2 , взятых в поперечном сечении светового пучка, и видность интерференционных полос от расстояния d между этими точками. При заданном угловом размере θ источника света колебания светового поля в точках S_1 и S_2 будут когерентными ($|\gamma_{12}(0)| \neq 0$), а исходящие из них волны при наложении будут интерферировать, если расстояние между этими точками $d < \lambda / \theta$. При $d = \lambda / \theta$ функции $|\gamma_{12}(0)|$ и $V(\theta)$ обращаются в нуль. Колебания в точках S_1 и S_2 в этом случае будут некогерентными, и при наложении волн из этих точек будет иметь место простое сложение интенсивностей – интерференция наблюдаться не будет. Выражение $d < \lambda / \theta$ можно переписать в виде $\omega < \lambda / a$, где $\omega \approx d / l_0$ – апертура интерференции (угол S_1OS_2) интерференционной схемы Юнга. Неравенство

$$\omega < \lambda / a \quad (4.23)$$

и является условием наблюдения интерференционных полос в схеме Юнга при протяженном источнике света.

Соотношение (4.23) можно получить и исходя из следующих простых соображений. Рассмотрим две светящиеся точки источника: крайнюю точку S' и точку O , расположенную

посередине. Как было показано выше, интерференционная картина, созданная крайней точкой S' источника, смещена относительно интерференционной картины от точки O , расположенной посередине источника, на расстояние, равное $x = (l/2l_0)a$. Если это смещение много меньше ширины интерференционной полосы, максимумы (и минимумы) от этих точек практически наложатся друг на друга и интерференционная картина будет наблюдаться. Если же это расстояние больше или равно ширине полосы, то максимумы от одной точки придется на минимумы от другой и интерференционная картина окажется полностью размытой. В качестве условия наблюдения интерференционной картины можно принять неравенство $x < (1/2)h$. Тогда учитывая, что $h = \lambda l / d$ и $d / l_0 \approx \omega$, приходим к соотношению (4.23).

Максимальное расстояние между точками S_1 и S_2 , взятыми в поперечном сечении светового пучка, при котором колебания в них еще когерентны, называют радиусом когерентности $\rho_{\text{ког}}$. Таким образом, принимают, что $\rho_{\text{ког}} \approx d_{\text{max}}$. С учетом этого для радиуса когерентности будем иметь

$$\rho_{\text{ког}} = \lambda / \theta.$$

Величина

$$S_{\text{ког}} = \pi \rho_{\text{ког}}^2$$

называется площадью когерентности, а величина

$$V_{\text{ког}} = \pi \rho_{\text{ког}}^2 l_{\text{ког}}$$

– объемом когерентности/

Угловой размер Солнца при наблюдении с Земли $\theta \approx 10^{-2}$ рад, средняя длина волны солнечного света $\lambda \approx 0,5$ мкм, следовательно, $\rho_{\text{ког}} = \lambda / \theta \approx 0,05$ мм. В п. 6.7 будет показано, что радиус когерентности круглого источника $\rho_{\text{ког}} = 1,22 \lambda / \theta$, а поскольку Солнце при наблюдении с Земли представляется

круглым светящимся диском, то более точное значение радиуса когерентности для солнечных лучей $\rho_{\text{ког}} = 0,06$ мм. Такое же расстояние d должно быть и между отверстиями S_1 и S_2 в схеме Юнга. Поэтому для уменьшения углового размера источника света и соответствующего увеличения d в опыте Юнга и ставится экран \mathcal{E}_1 с узким отверстием. В лазерных источниках света, в отличие от обычных тепловых или газоразрядных источников, пространственная когерентность сохраняется поперек всего пучка. Причем его ширина для практических задач (например, при голографировании) может быть увеличена без потерь когерентности с помощью линз – расширителя пучка. Таким образом, свет лазера обладает не только временной, но и пространственной когерентностью. Очень похожим на свет лазера с точки зрения временной и пространственной когерентности являются радиоволны.

Следует обратить внимание на следующее обстоятельство. Рабочей характеристикой для определения возможностей получения интерференционной картины методом деления амплитуды является время или длина когерентности, зависящие от величины спектрального интервала излучения. В опытах, осуществляемых по методу деления волнового фронта, аналогичной величиной служит радиус когерентности, зависящий от угловых размеров источника излучения.

4.6. Теорема Ван-Циттера – Цернике

Ван-Циттер и Цернике получили общую формулу для степени пространственной когерентности световых колебаний, возбуждаемых протяженным источником в точках S_1 и S_2 . Рассмотрим рис. 4.5, на котором изображен плоский протяженный источник S , находящийся в плоскости $\xi\eta$ и плоскость XU , расположенная параллельно плоскости $\xi\eta$ на расстоянии z от нее, $M(\xi, \eta)$ – произвольная точка источника.

Выберем в плоскости XU две близкие точки S_1 и S_2 с координатами

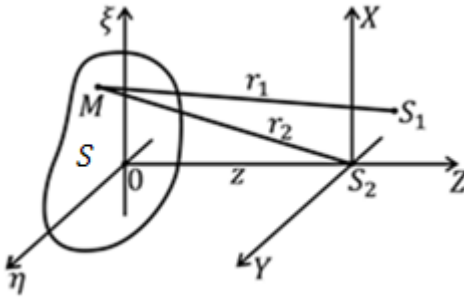


Рис. 4.5

тами (x, y, z) и $(0, 0, z)$ соответственно, причем точка S_2 расположена на одной оси с центром O протяженного источника. Точки S_1 и S_2 находятся на большом удалении от источника. Ставится задача, определить, в какой

мере когерентны поля в этих точках.

Излучение источника S будет складываться из излучений всех его атомов. Для решения поставленной задачи найдем напряженности полей, создаваемых в точках S_1 и S_2 каждым атомом источника.

Сферическая волна

$$E_j(t) = \frac{E_{0j}(t)}{r_j} \exp(-i\omega t),$$

излучаемая одним из таких атомов, создает колебание поля в точке S_1 , описываемое формулой

$$E_{1j}(t) = \frac{E_{0j}(t - r_{1j}/c)}{r_{1j}} \exp\left[-i\omega\left(t - \frac{r_{1j}}{c}\right)\right],$$

где r_{1j} – расстояние от j -го элементарного излучателя до точки S_1 . Такой же формулой будут описываться колебания поля и в точке S_2 с той лишь разницей, что в формуле для $E_{2j}(t)$ вместо r_{1j} будет стоять r_{2j} – расстояние от j -го излучателя до точки S_2 ,

Волновые поля в точках S_1 и S_2 , создаваемые всеми такими элементарными излучателями будут соответственно

$$E_1(t) = \sum_j \frac{E_{0j}(t - r_{1j}/c)}{r_{1j}} \exp \left[-i\omega \left(t - \frac{r_{1j}}{c} \right) \right],$$

$$E_2(t) = \sum_j \frac{E_{0j}(t - r_{2j}/c)}{r_{2j}} \exp \left[-i\omega \left(t - \frac{r_{2j}}{c} \right) \right].$$

Для вычисления взаимной корреляционной функции $\Gamma_{12}(\theta)$ при $\theta = 0$ следует перемножить величины $E_1(t)$ и $E_2^*(t)$ и полученное произведение усреднить по времени, т.е. найти $\langle E_1(t) E_2^*(t) \rangle$. При почленном перемножении выделим произведения, относящиеся к одному излучателю, и произведения, относящиеся к двум различным излучателям. Последние при усреднении обратятся в нуль, так как разные излучатели излучают не когерентно. В результате получим

$$\begin{aligned} \Gamma_{12}(0) &= \langle E_1(t) E_2^*(t) \rangle = \\ &= \sum_j \left\langle E_{0j} \left(t - \frac{r_{1j}}{c} \right) E_{0j}^* \left(t - \frac{r_{2j}}{c} \right) \right\rangle \frac{\exp[ik(r_{1j} - r_{2j})]}{r_{1j} r_{2j}}. \end{aligned}$$

Поскольку величины $(r_{1j} - r_{2j})/c$ очень малы, то при усреднении они исчезают, и мы получим

$$\begin{aligned} \Gamma_{12}(0) &= \sum_j \left\langle E_{0j}(t) E_{0j}^*(t) \right\rangle \frac{\exp[ik(r_{1j} - r_{2j})]}{r_{1j} r_{2j}} = \\ &= \sum_j \left\langle |E_{0j}(t)|^2 \right\rangle \frac{\exp[ik(r_{1j} - r_{2j})]}{r_{1j} r_{2j}}, \end{aligned}$$

где $\langle |E_{0j}(t)|^2 \rangle$ – интенсивность j -го излучателя.

Число элементарных излучателей в протяженном источнике велико (стремится к бесконечности). Их можно считать

непрерывно распределенными по источнику с определенной поверхностной плотностью. Поэтому, переходя к пределу, суммирование по отдельным излучающим атомам можно заменить интегрированием по площади источника S , положив

$$\langle |E_{0j}(t)|^2 \rangle = I_S(\xi, \eta) d\xi d\eta,$$

где $I_S(\xi, \eta)$ – интенсивность света, создаваемая единицей площади источника на единичном расстоянии от него. С учетом этого для функции взаимной когерентности будем иметь

$$\Gamma_{12}(0) = \iint_S I_S(\xi, \eta) \frac{\exp[ik(r_1 - r_2)]}{r_1 r_2} d\xi d\eta, \quad (4.24)$$

где r_1 и r_2 – расстояния от центра элемента $d\sigma = d\xi d\eta$ (на рис. 4.5 точка M) до точек S_1 и S_2 соответственно. Поскольку размеры источника малы по сравнению с расстоянием z , а точки S_1 и S_2 близки друг к другу (а значит, $(x - \xi)^2 + (y - \eta)^2 \ll z^2$ и $\xi^2 + \eta^2 \ll z^2$), то можно воспользоваться параксиальным приближением. В этом случае

$$r_1 = \sqrt{(x - \xi)^2 + (y - \eta)^2 + z^2} \approx z + \frac{(x - \xi)^2 + (y - \eta)^2}{2z};$$

$$r_2 = \sqrt{\xi^2 + \eta^2 + z^2} \approx z + \frac{\xi^2 + \eta^2}{2z};$$

$$r_1 - r_2 \approx \frac{x^2 + y^2}{2z} - \frac{x\xi + y\eta}{z};$$

$$r_1 r_2 \approx z^2.$$

Подставляя эти значения в подынтегральное выражение интеграла (4.24), для функции взаимной когерентности окончательно будем иметь

$$\Gamma(x, y) \equiv \Gamma_{12}(0) = e^{i\psi_{12}} \iint_S I_S(\xi, \eta) \exp[-i2\pi(u\xi + v\eta)] d\xi d\eta, \quad (4.25)$$

где $\psi_{12} = (k/2z)(x^2 + y^2)$, $u = x/\lambda z$, $v = y/\lambda z$.

Комплексную степень пространственной когерентности $\gamma_{12}(0)$ получим, разделив выражение (4.25) на произведение

$$\sqrt{\langle |E_1(t)|^2 \rangle} \cdot \sqrt{\langle |E_2(t)|^2 \rangle} = \sqrt{I_1} \cdot \sqrt{I_2},$$

где $I_i = \iint_S \frac{I_S(\xi, \eta)}{r_i^2} d\xi d\eta$ – интенсивность поля, создаваемого источником S в точке S_i , $i = 1, 2$. Учитывая, что

$$I_1 \approx I_2 \approx \frac{1}{z^2} \iint_S I_S(\xi, \eta) d\xi d\eta,$$

для комплексной степени пространственной когерентности получим

$$\gamma_{12}(0) = e^{i\psi_{12}} \frac{\iint_S I_S(\xi, \eta) \exp[-i2\pi(u\xi + v\eta)] d\xi d\eta}{\iint_S I_S(\xi, \eta) d\xi d\eta}. \quad (4.26)$$

Таким образом, комплексная степень пространственной когерентности колебаний светового поля в точках S_1 и S_2 , освещенных протяженным некогерентным источником, с точностью до несущественного фазового множителя $\exp(i\psi_{12})$ является нормированным фурье-образом распределения интенсивности $I_S(\xi, \eta)$ источника. Существенно при этом, что центр источника расположен на одной оси с точкой S_2 . Соотношение (4.26) и является

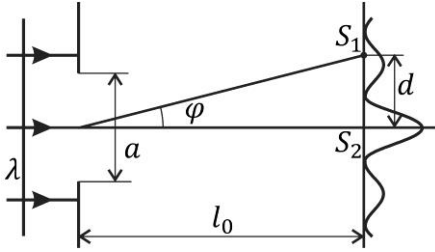


Рис. 4.6

математическим выражением теоремы Ван-Циттера – Цернике.

Теорема Ван-Циттера – Цернике для пространственной когерентности аналогична теореме Винера – Хинчина для временной когерентности. При этом

если первая связывает степень пространственной когерентности с преобразованием Фурье от распределения интенсивности по источнику, то вторая – степень временной когерентности с обратным преобразованием Фурье от спектральной плотности интенсивности излучения источника.

Покажем на примере источника света в виде бесконечно длинной светящейся полоски или щели, что теорема Ван-Циттера – Цернике приводит к тому же результату для степени пространственной когерентности, что и расчет ее с помощью основной формулы интерференции. Поскольку полоска (щель) бесконечно длинная, то распределение интенсивности источника будет функцией одной переменной ξ . Для одномерного случая формула (4.26) примет вид

$$\gamma_{12}(0) = e^{i\psi_{12}} \frac{\int_S I_S(\xi) \exp(-i2\pi u \xi) d\xi}{\int_S I_S(\xi) d\xi},$$

где интегрирование производится по ширине полоски (щели).

Если начало координатной оси ξ находится в середине источника и распределение интенсивности по источнику постоянно и равно I_0 , то

$$I_S(\xi) = \begin{cases} I_0, & |\xi| \leq a/2; \\ 0, & |\xi| > a/2. \end{cases}$$

Тогда

$$\begin{aligned} \gamma_{12}(0) &= e^{i\psi_{12}} \frac{I_0 \int_{-a/2}^{a/2} \exp(-i2\pi u \xi) d\xi}{I_0 \int_{-a/2}^{a/2} d\xi} = e^{i\psi_{12}} \frac{\sin \pi u a}{\pi u a} = \\ &= e^{i\psi_{12}} \operatorname{sinc}(ua). \end{aligned}$$

Поскольку пространственная частота для точки S_1 при условии, что точка S_2 , как и центр источника, находится на оси Z (рис. 4.6), равна $u = \sin \varphi / \lambda$, где φ – угол между направлением на точку S_1 , и осью Z . Так как $d \ll l_0$, то $\sin \varphi \approx \operatorname{tg} \varphi = d / l_0$ и $u = d / \lambda l_0$. Тогда

$$\begin{aligned} \gamma_{12}(0) &= e^{i\psi_{12}} \operatorname{sinc}(ua) = \\ &= e^{i\psi_{12}} \operatorname{sinc}\left(\frac{ad}{\lambda l_0}\right) = e^{i\psi_{12}} \operatorname{sinc}\left(\frac{\theta d}{\lambda}\right). \end{aligned}$$

Взяв модуль этого выражения, найдем и степень пространственной когерентности:

$$|\gamma_{12}(0)| = \left| \operatorname{sinc} \frac{\theta d}{\lambda} \right|.$$

Полученное выражение в точности совпадает с выражением (4.22).

В п.6.10 с помощью теоремы Ван-Циттера – Цернике будут получены выражения для степени пространственной когерентности и радиуса когерентности в случае источников прямоугольной и круглой формы.

Глава 5

ДИФРАКЦИЯ СВЕТА. ДИФРАКЦИЯ ФРЕНЕЛЯ

5.1. Определение дифракции. Виды дифракции

Под дифракцией света обычно понимают явления, связанные с отклонением света от закона прямолинейного распространения. Дифракцию можно наблюдать, например, когда на пути распространения света находятся препятствия, т.е. непрозрачные тела произвольной формы, или когда свет проходит сквозь отверстия в экранах, размеры которых (отверстий) соизмеримы с длиной волны света. Такие экраны называют дифракционными экранами, а также объектными экранами, объектными масками, или, иначе, апертурными масками, а сами отверстия – апертурами или диафрагмами. Закон прямолинейного распространения света предсказывает существование за экраном области тени, резко отграниченной от тех областей, куда свет попадает. Опыт, однако, показывает, что вместо резкой границы между светом и

тенью получается довольно сложная картина распределения освещенности (интенсивности), состоящая из темных и светлых участков (обычно из чередующихся темных и светлых полос или колец подобно интерференционным). Это означает, что световая волна проникает в область геометрической тени. Указанная картина называется дифракционной картиной, а чередующиеся темные и светлые полосы или кольца – дифракционными полосами или кольцами. В отличие от интерференционных полос или колец, дифракционные полосы или кольца имеют различную интенсивность, причем она убывает с увеличением расстояния от центра картины.

Отклонение света от прямолинейного распространения происходит и в тех случаях, когда свет проходит через прозрачные объекты, показатель преломления которых зависит от координат. Такими объектами являются, например, транспаранты – тонкие пластинки, толщина и прозрачность которых в общем случае зависят от координат точек их поверхности. Транспарантами являются фотопленки или фотопластинки с нанесенными на них изображениями предметов (с записанными на них сигналами). В этих случаях происходит нарушение одинаковости изменения амплитуды или фазы по всей поверхности волнового фронта.

Дифракционные явления, таким образом, наблюдаются всегда, при ограничении фронта волны и когда свет проходит через прозрачную среду или экран с оптическими неоднородностями. В последнем случае изменение амплитуды и фазы волны оказывается неодинаковым на всей поверхности волнового фронта, имеет место локальное амплитудное и (или) фазовое нарушение волнового фронта. Экраны, изменяющие только амплитуду оптического поля, называются амплитудными экранами, а только фазу – фазовыми экранами. Примером фазового экрана является тонкая линза.

В электромагнитной теории света явление дифракции объясняется тем, что распространение света подчиняется волновому уравнению.

В зависимости от расстояний a и b соответственно источника света S и точки наблюдения P (или экрана) от преграды, поставленной на пути распространения света, различают два вида дифракции. Если эти расстояния велики, так что лучи, падающие на преграду, и лучи, идущие в направлении точки наблюдения P , образуют практически параллельные пучки, говорят о дифракции Фраунгофера, дифракции в параллельных лучах или о дифракции в дальней зоне. В противоположном случае говорят о дифракции в непараллельных лучах, дифракции Френеля или о дифракции в ближней зоне.

Наиболее распространенным видом дифракции является дифракция Френеля. Дифракцией Френеля является, например, дифракция сферических волн на любых преградах. Дифракцию Фраунгофера можно, очевидно, рассматривать как некий частный случай дифракции Френеля. Она возникает только в одном предельном случае, когда $a \rightarrow \infty$ и $b \rightarrow \infty$. Однако принципиального различия и резкой границы между этими двумя видами дифракции нет; один вид дифракции непрерывно переходит в другой. При теоретическом рассмотрении дифракционных явлений обычно считают, что дифракция Френеля имеет место на расстояниях от апертуры $z \approx R^2 / \lambda$, а дифракция Фраунгофера – на расстояниях $z \gg R^2 / \lambda$, где R – линейный размер апертуры.

На практике для наблюдения дифракции Фраунгофера используется схема, показанная на рис. 5.1. Точечный источник света S помещается в переднем фокусе собирающей линзы L_1 . Возника-

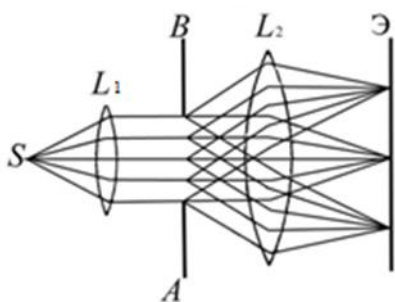


Рис. 5.1

ющий параллельный пучок света дифрагирует на преграде (например, на отверстии в экране AB). Дифракционная картина наблюдается на экране \mathcal{E} , помещенном в задней фокальной плоскости другой линзы L_2 , которая располагается за преградой. Линза L_2 собирает в

различных участках своей фокальной плоскости все испытавшие дифракцию на отверстии лучи (их называют дифрагировавшими лучами), в том числе и лучи, отклонившиеся от первоначального направления. Диаметры линз предполагаются настолько большими, что линзы не ограничивают фронт падающей на них волны и не вызывают ее дифракцию.

Во всех оптических системах угловая ширина (апертура) светового пучка ограничена и, следовательно, дифракция происходит во всех оптических приборах. Поэтому изучение дифракционных явлений имеет большое практическое значение для понимания многих процессов инструментальной оптики. Дифракционные представления служат основой для построения теории передачи и обработки оптических сигналов.

Оптическая фраунгоферовская дифракция является промежуточным шагом в формировании изображений с помощью линз. При этом линза выполняет задачу сведения дифрагировавшего света в плоскости изображения. Дифракция определяет также пределы возможного улучшения качества оптических изображений и точности измерений определенного типа.

Дифракционные явления важны и для понимания всей волновой теории и имеет ряд интересных применений.

Ниже излагается приближенная теория дифракции. Она использует два основных упрощения. Во-первых, в этой теории

рассматривается только скалярная величина – одна из составляющих вектора \mathbf{E} электрического поля. При этом предполагается, что другая составляющая может рассматриваться независимо таким же образом. (В этом смысле рассматриваемая теория является скалярной.) Во-вторых, в этой теории предполагается, что область, в которой происходят дифракционные явления, ограничена частью пространства, близко примыкающей к оптической оси системы. Это означает, что рассматриваются только те волны, которые дифрагируют под достаточно малыми углами к оптической оси. Указанную область называют областью параксиальных лучей. Однако эта теория вполне достаточна для большинства задач прикладной оптики.

5.2. Принцип Гюйгенса – Френеля

Дифракционные явления исследуются с помощью общего принципа, называемого принципом Гюйгенса – Френеля. Идея этого принципа основана на том, что световая волна, излученная источником, в дальнейшем ведет автономное существование, независимое от наличия этого источника. Принцип Гюйгенса – Френеля представляет собой усовершенствованный и дополненный Френелем принцип Гюйгенса, согласно которому каждая точка волнового фронта рассматривается как источник вторичных сферических волн; огибающая этих волн дает положение фронта волны в следующий момент времени. По Френелю, действие источника света S в какой-либо точке P можно заменить действием произвольной поверхности σ , окружающей источник. Каждый элемент $d\sigma$ этой поверхности рассматривается в качестве источника вторичных волн, распространяющихся по всем направлениям. Считается, что эти вторичные источники создаются одним и тем же первичным источником S , и поэтому являются когерентными. Световое поле $E(P)$ в точке P представляет собой результат интерференции волн, приходящих в эту точку от всех вторичных источников,

видимых из точки P . Что касается самого источника S , то предполагается, что он только создает вторичные источники на поверхности σ , но сам непосредственно не вносит вклада в поле в точке P . Таким образом, сущность принципа Гюйгенса – Френеля состоит в замене реального источника света действием произвольно выбранной поверхности, рассматриваемой как совокупность вторичных источников когерентных волн. Обычно в качестве указанной поверхности выбирается поверхность фронта волны, падающей на экран.

Пусть на пути распространения света от источника света S

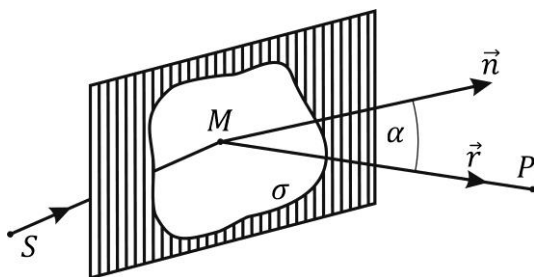


Рис. 5.2

находится непрозрачный экран с произвольным отверстием и требуется найти амплитуду светового поля $E(P)$ в некоторой точке P за экраном (рис. 5.2). Следуя Френелю, предположим, что

отверстие покрыто некоторой произвольной поверхностью σ , являющейся системой источников вторичных сферических волн. Эти вторичные волны распространяются за экраном в различных направлениях и, в частности, в направлении точки P . Разделим мысленно поверхность σ на элементарные участки площадью $d\sigma$ (точка M на рис. 5.2 – центр этого элемента). Поле, создаваемое элементом $d\sigma$, будет пропорционально полю E_S , создаваемому истинным источником S на элементе поверхности $d\sigma$ и площади этого элемента $d\sigma$ (т.е. числу когерентных точечных источников вторичных волн внутри элемента). Чтобы найти поле $dE(P)$, создаваемое элементом $d\sigma$ в точке P , находящейся на расстоянии r от него, следует учесть также изменение амплитуды вторичной сферической волны при

ее распространении от элемента $d\sigma$ к точке P , обратно пропорциональное r , и набег фазы kr при распространении ее на расстояние r , где $k = 2\pi/\lambda$ – волновое число, λ – длина волны, создаваемой источником S . Это приводит к появлению в выражении для $dE(P)$ множителя e^{ikr}/r . С учетом этого поле, порождаемое элементом $d\sigma$ в точке P , можно представить в виде

$$dE(P) = K(\alpha) E_s \frac{e^{ikr}}{r} d\sigma, \quad (5.1)$$

где $K(\alpha)$ – так называемый коэффициент наклона, учитывающий тот факт, что вклад элемента $d\sigma$ в результирующее поле зависит от его ориентации по отношению к направлению на точку наблюдения P , α – угол между нормалью \mathbf{n} к элементу поверхности $d\sigma$ и направлением на точку наблюдения. По Френелю, этот коэффициент принимает максимальное значение, когда нормаль \mathbf{n} совпадает с направлением на точку P ($\alpha = 0$), монотонно убывает и обращается в нуль при $\alpha = \pi/2$. Согласно Кирхгофу – Рэлею – Зоммерфельду, коэффициент наклона

$$K(\alpha) = -(i/\lambda) \cos \alpha$$

(множитель $-i = \exp(-i\pi/2)$ учитывает дополнительное отставание по фазе вторичной волны в точке P по сравнению с излучением в месте расположения элемента $d\sigma$ на $\pi/2$). Фазовый множитель $\exp(-i\omega t)$ в (5.1) опущен, так как он не влияет на распределение поля в пространстве. Следовательно, выражение (5.1) определяет элементарную комплексную амплитуду светового поля в точке P . Результирующая амплитуда в этой точке, в соответствии с принципом суперпозиции, определится как сумма амплитуд полей, создаваемых всеми такими элементарными участками. Действие самого экрана на падающую волну в общем случае учитывается с помощью

функции $t(M)$, зависящей от положения точки M на экране и называемой амплитудной функцией пропускания экрана или апертурной функцией.

Функция $t(M)$ является характеристикой только самого отверстия (апертуры) и не учитывает ни свойства материала, из которого изготовлен экран, ни его размеры. Эта функция будет действительной величиной, если экран изменяет только амплитуду волны, и комплексной – если еще и фазу. При вычислении амплитуды поля в точке P эту функцию следует умножить на E_S . С учетом этого для результирующей амплитуды поля $E(P)$ в точке P будем иметь:

$$E(P) = -\frac{i}{\lambda} \iint_{\sigma} E(M) \frac{\cos \alpha}{r} e^{ikr} d\sigma, \quad (5.2)$$

где $E(M) = E_S t(M)$ – амплитуда поля в точке M в плоскости непосредственно за экраном. Интегрирование в (5.2) производится по площади поверхности апертуры σ . В случае дифракции на простых экранах, размеры которых значительно больше длины волны, когда справедливо приближение геометрической оптики (называемое также приближением Кирхгофа), амплитуду поля $E(M)$ принимают равной E_S в пределах отверстия и равной нулю, где экран перекрывает свет. Это означает, что

$$t(M) = \begin{cases} 1 & \text{в пределах отверстия,} \\ 0 & \text{вне пределов отверстия.} \end{cases}$$

Указанное соотношение называют граничным условием Кирхгофа.

Выражение (5.2) является математической формулировкой принципа Гюйгенса – Френеля. Кирхгоф, Рэлей и Зоммерфельд обосновали принцип Гюйгенса – Френеля, показав, что этот принцип, по существу, является одним из методов решения

скалярного волнового уравнения Гельмгольца с заданными граничными условиями.

Из (5.2) видно, что поле за экраном можно рассматривать как совокупность сферических волн, амплитуда и фаза которых определяется амплитудой и фазой в плоскости непосредственно за экраном.

5.3. Дифракционный интеграл Кирхгофа – Френеля – Зоммерфельда

При решении задачи о дифракции на плоском экране в координатной области с плоскостью отверстия связывают координатную плоскость XU декартовой системы координат XUW , а с плоскостью, в которой расположена точка P (эта плоскость называется плоскостью наблюдения), параллельной плоскости XU и находящейся на расстоянии z от нее – координатную плоскость $X'U'$, начало O' которой расположим на оси W системы XUW (рис. 5.3). Координаты x' и y' вводятся только для удобства описания поля в плоскости наблюдения; при таком выборе начала координат ко-

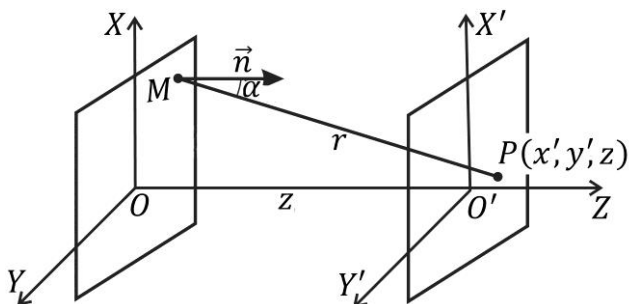


Рис. 5.3

ординаты x' и y' принимают те же значения, что и x и y . Учитывая, что координатами точки M являются x и y , заменим в

интеграле (5.2) $E(M)$ на $E(x, y)$, $E(P)$ на $E'(x', y')$, а $d\sigma$ на $dxdy$. Интеграл (5.2) примет вид

$$E'(x', y') = -\frac{i}{\lambda} \iint_{-\infty-\infty}^{\infty \infty} E(x, y) \frac{\cos \alpha}{r} e^{ikr} dx dy, \quad (5.3)$$

где $E(x, y) = E_s t(x, y)$, $t(x, y)$ – апертурная функция экрана, α – угол между нормалью \mathbf{n} к плоскости экрана (положительным направлением оси Z) и отрезком прямой, соединяющим точку $M(x, y, 0)$, являющуюся центром элемента $dxdy$, с точкой $P(x', y', z)$ в плоскости наблюдения XY' ,

$$r = \sqrt{(x' - x)^2 + (y' - y)^2 + z^2} \quad (5.4)$$

– расстояние между точками M и P . Интеграл (5.3) и определяет искомую связь амплитуд $E(x, y)$ и $E'(x', y')$. Предполагается, что амплитуда $E(x, y)$ тождественно равна нулю всюду кроме отверстия, поэтому интегрирование в (5.3) распространено на всю плоскость XY (производится по бесконечным пределам).

Интеграл (5.3) называют интегралом суперпозиции или дифракционным интегралом Кирхгофа – Френеля – Зоммерфельда (иногда Рэлея – Зоммерфельда). Он связывает распределение поля (оптический сигнал) $E(x, y)$ в плоскости непосредственно за экраном (в плоскости $z = 0$) с распределением поля (оптическим сигналом) $E'(x', y')$ в плоскости на расстоянии z от экрана. Роль оптической системы, преобразующей входной сигнал $E(x, y)$ в выходной сигнал $E'(x', y')$, играет здесь слой свободного пространства между двумя параллельными плоскостями. Соотношение (5.3) показывает, что на расстоянии $r \gg \lambda$ комплексная амплитуда светового поля в произвольной точке $P(x', y', z)$ плоскости наблюдения определяется как суперпозиция элементарных комплексных амплитуд $E(x, y) dx dy$

с комплексной весовой функцией $(e^{ikr} / r) \cos \alpha$. Физически это означает, что, как и в случае интеграла (5.2), каждая точка $M(x, y)$ поля $E(x, y)$ испускает сферическую волну e^{ikr} / r с амплитудным множителем $-(i / \lambda) \cos \alpha$.

Интеграл (5.3) является решением скалярного волнового уравнения Гельмгольца для свободного пространства.

Если экран освещается нормально падающей плоской монохроматической волной, что соответствует бесконечно удаленному от экрана источнику S или точечному источнику S , находящемуся в фокусе собирающей линзы, помещенной перед экраном, то $E_S = E_0$, где E_0 – комплексная амплитуда падающей на экран плоской волны. Если экран освещается сферической волной от точечного источника S , то $E_S = (E_0 / r_0) \exp(ikr_0)$, где E_0 – комплексная постоянная, r_0 – расстояние от источника S до центра элемента $d\sigma$. В этом случае интеграл (5.3) преобразуется к виду

$$E'(x', y') = \frac{i}{\lambda} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \frac{\cos \alpha}{r_0 r} e^{ik(r_0+r)} dx dy. \quad (5.5)$$

Если точечный источник S расположен на оси z в точке $(0, 0, -a)$, то расстояние

$$r_0 = \sqrt{a^2 + x^2 + y^2}. \quad (5.6)$$

5.4. Представление поля излучения в виде совокупности плоских волн

Представление поля оптического излучения в виде совокупности сферических волн не является единственным способом его описания. Другой подход к определению поля в плоскости $z = \text{const} > 0$ предложил Рэлей. Согласно этому представлению, поле за плоскостью $z = 0$ можно рассматривать

как совокупность плоских волн. Этот подход лежит в основе спектрального метода решения дифракционных задач.

Как было показано в п. 1.6 (формула (1.30)), амплитуду плоской волны можно записать в виде

$$E(x, y, z) = E_0 e^{i2\pi(u x + v y)} e^{i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}}, \quad (5.7)$$

где u и v – пространственные частоты волны вдоль координатных осей X и Y соответственно. Плоская волна является решением волнового уравнения Гельмгольца. Уравнение Гельмгольца представляет собой линейное дифференциальное уравнение; для него справедлив принцип суперпозиции. Поэтому решением этого уравнения будет и линейная комбинация полей вида (5.7), т.е. сумма плоских волн с амплитудами, определяемыми функцией $E(u, v)$, различными фазами и направлениями распространения (пространственными частотами). Такое решение можно записать в виде интеграла

$$E(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) e^{i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}} e^{i2\pi(u x + v y)} dudv, \quad (5.8)$$

где интегрирование производится по всем пространственным частотам, удовлетворяющим условию $u^2 + v^2 < 1/\lambda^2$.

Комплексная функция $E(u, v)$ в подынтегральном выражении интеграла (5.8) описывает как амплитуду, так и фазу отдельной плоской волны с направлением распространения, определяемым совокупностью пространственных частот u , v и w . Соотношение (5.8) называется представлением Рэлея. Его можно рассматривать как обобщение понятия комплексной амплитуды на случай неплоской монохроматической волны.

Представление Рэлея позволяет решить следующую задачу. Пусть задано распределение поля (сигнал) в плоскости $z = 0$, требуется найти распределение поля (сигнал) в плоскости $z = \text{const} > 0$, параллельной плоскости $z = 0$. Более строго эта

задача формулируется так: известно распределение амплитуды волнового поля в плоскости $z = 0$, требуется построить решение волнового уравнения в плоскости $z = \text{const} > 0$, обращающееся в заданную функцию на плоскости $z = 0$. Эту задачу можно трактовать как распространение поля (сигнала) от плоскости $z = 0$ к плоскости $z = \text{const} > 0$. При этом исходное распределение, благодаря процессу распространения, преобразуется в новое распределение. Возможность решения поставленной задачи обусловлена тем, что функция (5.8) представляет собой аналитическую функцию во всем пространстве, а такие функции имеют однозначное продолжение внутри замкнутой области, если известны их значения на границах области.

Обозначим функцию распределения комплексных амплитуд в плоскости $z = 0$ через $E(x, y, 0)$ и потребуем, чтобы общее решение (5.8) обращалось бы в заданное распределение комплексной амплитуды при $z = 0$. Тогда получим

$$E(x, y, 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) e^{i2\pi(ux+vy)} dudv. \quad (5.9)$$

Заметим, что, поскольку поле $E(x, y, 0)$ разлагается в непрерывный спектр по плоским волнам, можно рассматривать распространение каждой составляющей пространственного спектра плоских волн в отдельности, а затем, сложив их в плоскости $z = \text{const} > 0$, получить в этой плоскости новое распределение поля $E(x, y, z)$. Это и отражено в выражении (5.8).

Выражение (5.9) показывает, что функция $E(x, y, 0)$ связана с функцией $E(u, v)$ обратным преобразованием Фурье. Функцию $E(u, v)$, являющуюся фурье-спектром волнового поля $E(x, y)$, можно определить с помощью прямого преобразования Фурье:

$$E(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y, 0) e^{-i2\pi(ux+vy)} dx dy.$$

Подставляя это в выражение (5.8), получим

$$E(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y, 0) \exp(i 2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}) dudv. \quad (5.10)$$

Выражение (5.10) позволяет решить поставленную задачу определения распределения комплексной амплитуды поля в плоскости $z = \text{const} > 0$ по заданному распределению поля в плоскости $z = 0$. Видим, таким образом, что в основе метода Рэля лежит интегральное преобразование Фурье.

Таким образом, зная распределение комплексных амплитуд поля $E(x, y, 0)$ в плоскости $z = 0$, с помощью преобразования Фурье можно определить комплексную функцию $E(u, v)$, а затем с помощью интеграла (5.8) найти распределение амплитуды поля в любой плоскости $z = \text{const} > 0$.

Задачи в сформулированной выше постановке на практике встречаются редко. Чаще приходится иметь дело с дифракционными задачами, когда в поле световой волны помещено препятствие и требуется определить поле, которое получается на большом расстоянии от этого препятствия. К решению задач такого типа метод Рэля применяется следующим образом. Непосредственно за препятствием со стороны прошедшей волны строится плоскость $X'Y'$, принимаемая за плоскость $z = 0$. На этой плоскости, используя приближение геометрической оптики (приближение Кирхгофа), находится поле волны, искаженное препятствием. Если падающая на объектный экран волна является плоской с амплитудой в плоскости $z = 0$, равной E_0 , то это поле будет $E(x, y, 0) = E_0 t(x, y)$. Далее можно воспользоваться выражением (5.8), которое для данного случая примет вид

$$E(x, y, z) = E_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(u, v) e^{i 2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}} e^{i 2\pi (ux + vy)} dudv,$$

(5.11)

где $t(u, v)$ – фурье-образ амплитудной функции пропускания $t(x, y)$ экрана, найти поле в любой плоскости $z = \text{const} > 0$ и тем самым решить поставленную дифракционную задачу.

Соотношение (5.11) показывает, что если плоская волна, распространяющаяся в направлении оси Z и имеющая амплитуду в плоскости $z = 0$, равную E_0 , падает на объектный экран, имеющий амплитудную функцию пропускания $t(x, y)$, то спектр $E(u, v, z)$ комплексной амплитуды волны в плоскости $z = \text{const} > 0$ имеет вид

$$E(u, v, z) = E_0 t(u, v) \exp(i 2 \pi z \sqrt{1/\lambda^2 - u^2 - v^2}). \quad (5.12)$$

Если функция пропускания $t(x, y)$ периодическая, то интегрирование в (5.11) заменяется суммированием и тогда

$$E(x, y, z) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} t_{mn} \exp(i 2 \pi z \sqrt{1/\lambda^2 - u^2 - v^2}) \times \\ \times \exp[i 2 \pi (m u x + n v y)].$$

Коэффициенты разложения t_{mn} функции пропускания $t(x, y)$ в ряд Фурье при этом определяются как

$$t_{mn} = \frac{1}{d_1 d_2} \int_{-d_1/2}^{d_1/2} \int_{-d_2/2}^{d_2/2} t(x, y) \exp[-i 2 \pi (u x + v y)] dx dy,$$

где $d_1 = 1 / u$ и $d_2 = 1 / v$ – периоды функции $t(x, y)$ по переменным x и y соответственно.

При решении дифракционных задач метод Рэлея приводит к тем же результатам, что и метод Гюйгенса – Френеля. В этом смысле оба метода эквивалентны друг другу. Можно показать также, что интеграл (5.10), как и интеграл (5.3), является решением скалярного волнового уравнения Гельмгольца для

свободного пространства. Решение (5.10) является общим решением указанного уравнения.

Из выражений (5.3) и (5.10) видно, что распространение оптического сигнала в свободном пространстве описывается линейным преобразованием. Это обстоятельство делает возможным рассматривать большинство оптических явлений с точки зрения линейных систем. Такие оптические явления, как, например, формирование оптического изображения, голография, и интегральные преобразования (например, преобразование Френеля и преобразование Фурье) являются линейными по своему характеру. Линейные интегральные преобразования и линейные оптические системы подробно будут рассматриваться нами в восьмой и девятой главах.

Рассмотрим теперь случай представления Рэлея, когда z мало. Произведем разложение показателя степени первого фазового множителя в подынтегральном выражении интеграла (5.8) в ряд по степеням $(u^2 + v^2)/(1/\lambda^2)$. Учитывая, что разложение функции $\sqrt{1 + \alpha}$ в ряд по степеням α

$$\sqrt{1 + \alpha} = 1 + \frac{1}{2}\alpha - \frac{1}{8}\alpha^2 + \frac{1}{16}\alpha^3 - \dots$$

и ограничиваясь первыми двумя членами разложения, получим

$$\begin{aligned} 2\pi z \sqrt{1/\lambda^2 - u^2 - v^2} &= \frac{2\pi}{\lambda} z \sqrt{1 - \frac{u^2 + v^2}{1/\lambda^2}} \approx \\ &\approx \frac{2\pi}{\lambda} z \left(1 - \frac{u^2 + v^2}{2 \cdot 1/\lambda^2} + \dots \right) = kz - \pi z \lambda (u^2 + v^2), \end{aligned} \quad (5.13)$$

где учтено, что $2\pi/\lambda = k$. Поскольку фаза зависит от переменных u и v по параболическому закону, то приведенное разложение, как и подобное разложение в координатном пространстве, называют параболическим приближением. Оно справедливо, если третий член разложения

$$\frac{2\pi}{\lambda} z \frac{(u^2 + v^2)^2}{8 \cdot 1/\lambda^4} = \frac{1}{4} \pi \lambda^3 z (u^2 + v^2)^2 \ll \pi$$

или если $u^2 + v^2 \ll 2/\sqrt{\lambda^3 z}$. Заметим, что предельное значение суммы $u^2 + v^2$ определяется расстоянием z до плоскости наблюдения и длиной волны λ излучения. Приведем числовой пример. Пусть $z = 10$ см, $\lambda = 0,5$ мкм. Тогда предельное значение пространственной частоты, для которого справедливо параболическое приближение, будет

$$\chi_{\max} = \sqrt{(u^2 + v^2)_{\max}} = \sqrt{2/\sqrt{\lambda^3 z}} = 113 \text{ мм}^{-1}.$$

В параболическом приближении выражение (5.12) примет вид

$$E(u, v, z) = E_0 t(u, v) \exp[-i\pi\lambda z(u^2 + v^2)].$$

Фазовый множитель $\exp(ikz)$, постоянный в плоскости XY , здесь опущен.

Если $\pi\lambda z(u^2 + v^2)_{\max} \ll \pi$, то второй и последующие члены разложения (5.13) практически не приводят к изменению указанного множителя и их можно отбросить, ограничившись только первым слагаемым kz . В этом случае первый фазовый множитель принимает вид $\exp(ikz)$. Его можно вынести из-под знака интеграла (5.8); оставшийся интеграл, согласно (5.9), будет представлять собой распределение поля в плоскости $z = 0$. С учетом этого получаем

$$E(x, y, z) = E(x, y, 0) e^{ikz}. \quad (5.14)$$

Полученное соотношение называют приближением тени. Оно справедливо на расстояниях $z \ll [\lambda(u^2 + v^2)_{\max}]^{-1}$, т.е. в плоскостях, близко расположенных к плоскости $z = 0$.

Поскольку $(u^2 + v^2)_{\max} \approx (1/R)^2$, где R – максимальный размер отверстия, то условие применимости приближения тени можно записать в виде $z \ll R^2 / \lambda^2$ или в виде $p \ll 1$, где $p = \sqrt{\lambda z} / R$ – так называемый параметр дифракции. Это условие является условием применимости геометрической оптики, поэтому приближение тени можно назвать приближением геометрической оптики. Заметим, что условие $p \ll 1$ будет выполняться не только при малых z , но и при больших размерах отверстия R . При любых заданных z и R условие $p \ll 1$ будет выполняться, если $\lambda \rightarrow 0$ – общее условие применимости законов геометрической оптики.

Приближение тени показывает, что при малых значениях z поле в плоскости $z = \text{const} > 0$ распределено точно так же, как и в плоскости $z = 0$. Не зависящий от x и y множитель e^{ikz} , учитывает лишь набег фазы, происшедший при распространении световой волны до плоскости $z = \text{const} > 0$, и на распределение поля в плоскости $z = 0$ не влияет. При этом распределение интенсивности $I(x, y, z)$ в плоскости $z = \text{const} > 0$ тождественно распределению интенсивности $I(x, y, 0)$ в плоскости $z = 0$:

$$I(x, y, z) = I(x, y, 0).$$

5.5. Приближение Френеля

При исследовании дифракционных явлений с помощью формулы Кирхгофа – Френеля – Зоммерфельда (5.3) обычно рассматривают два предельных случая. Первый из них соответствует дифракции Френеля и описывает распределение поля в ближней зоне, второй – дифракции Фраунгофера и описывает распределение поля в дальней зоне. Рассмотрим сначала первый случай.

Будем считать, что линейные размеры отверстия (области, занятой полем $E(x, y)$ в плоскости XU) и зоны наблюдения дифракции в плоскости $X'U'$ малы по сравнению с расстоянием z

между этими плоскостями, т.е. будем считать, что справедливо параксиальное приближение, выражающееся условием

$$(x' - x)^2 + (y' - y)^2 \ll z^2. \quad (5.15)$$

В этом случае

$$\cos \alpha \approx 1, \quad (5.16)$$

а r мало отличается от z . Поэтому в знаменателе подынтегрального выражения в (5.3) расстояние r можно заменить на z и вынести множитель $1/z$ из-под знака интеграла, т.е. предполагается, что

$$1/r \approx 1/z = \text{const}. \quad (5.17)$$

В аргументе экспоненты замену r на z произвести нельзя, так как возникающая при этом погрешность умножается на очень большое число $k = 2\pi/\lambda \approx 10^6$ рад/м, вследствие чего фазовые искажения становятся значительными (много большими π). Для получения более точного значения r произведем разложение выражения (5.4) в степенной ряд:

$$\begin{aligned} r &= \sqrt{(x' - x)^2 + (y' - y)^2 + z^2} = \\ &= z \sqrt{1 + \frac{(x' - x)^2 + (y' - y)^2}{z^2}} \approx z + \\ &\quad + \frac{(x' - x)^2 + (y' - y)^2}{2z} - \\ &\quad - \frac{[(x' - x)^2 + (y' - y)^2]^2}{8z^3} + \dots \quad (5.18) \end{aligned}$$

Ограничиваясь первыми двумя членами разложения, будем иметь

$$r \approx z + \frac{(x' - x)^2 + (y' - y)^2}{2z}. \quad (5.19)$$

Соотношения (5.15) – (5.17) и (5.19) определяют условия дифракции Френеля и называются приближением Френеля. С физической точки зрения приближение Френеля означает замену сферических волновых фронтов вторичных волн Гюйгенса параболическими поверхностями.

В приближении Френеля интеграл суперпозиции (5.3) принимает вид

$$E'(x', y') = \frac{e^{ikz}}{i\lambda z} \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[ik \frac{(x' - x)^2 + (y' - y)^2}{2z} \right] dx dy \quad (5.20)$$

где учтено, что $-i/\lambda = 1/(i\lambda)$. Интеграл (5.20) называют интегралом суперпозиции в приближении Френеля.

Замена в фазовом множителе точного выражения для r на приближенное (5.19) справедливо, если отброшенные члены более высокого порядка в разложении (5.18) практически не изменяют величины фазового множителя e^{ikr} . Это возможно, если третий член разложения

$$k \frac{[(x' - x)^2 + (y' - y)^2]^2}{8z^3} \ll \pi.$$

Откуда

$$\sqrt[3]{\frac{[(x' - x)^2 + (y' - y)^2]^2}{4\lambda}} \ll z. \quad (5.21)$$

Соотношение (5.21) и определяет условие применимости приближения Френеля. Однако это условие не является необходимым, поскольку основной вклад в значение интеграла (5.20) дают только те значения x и y , которые близки к

соответствующим значениям x' и y' . Действительно, важно лишь, чтобы отброшенные в разложении (5.18) члены практически не изменяли величины интеграла (5.3). В этом смысле имеет значение, насколько быстро осциллирует множитель $\exp(ikr)$. При достаточно больших значениях $1/\lambda z$ экспоненциальный множитель в (5.20), который можно переписать в виде $\exp\{i\pi[(x' - x)^2 + (y' - y)^2]/\lambda z\}$, будет осциллировать быстро, так что отличающийся от нуля вклад в интеграл (5.20) будут давать точки, которые удовлетворяют условию $x' \approx x$, $y' \approx y$ (так называемая область стационарной фазы). Но при $x' \approx x$ и $y' \approx y$ отброшенные в разложении (5.18) члены оказываются пренебрежимо малыми по сравнению с первыми двумя даже при небольших значениях z . Поэтому интеграл (5.20) можно использовать для вычисления поля и на сколь угодно близких расстояниях z от экрана, независимо от размеров отверстия.

К соотношению (5.20) приводит и представление Рэлея. Используем формулу (5.8) для определения поля в плоскости $X'Y'$ (в плоскости $z = \text{const}$), подставив в нее координаты x' и y' в плоскости $X'Y'$, заменив символ E на символ E' и опустив символ z в левой части равенства. Тогда получим

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) e^{i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}} e^{i2\pi(ux' + vy')} dudv, \quad (5.22)$$

Проведя те же выкладки, что и в п. 5.4, вместо выражения (5.10) будем иметь выражение

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp(i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}) \times \\ \times \exp\{i2\pi[u(x' - x) + v(y' - y)]\} dudv dx dy, \quad (5.23)$$

где в обозначении поля в плоскости $z = 0$ опущен символ 0.

В параксиальном приближении, когда выполняется приближенное равенство (5.13), интеграл (5.23) можно представить в виде

$$E'(x', y') = e^{ikz} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[-i2\pi\lambda z(u^2 + v^2) + i2\pi u(x' - x) + i2\pi v(y' - y)] dudv \right\} dx dy.$$

Используя формулу

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[-\alpha(x^2 + y^2) \pm \beta_1 x \pm \beta_2 y] dx dy = \frac{\pi}{\alpha} \exp\left(\frac{\beta_1^2 + \beta_2^2}{4\alpha}\right) \quad (5.24)$$

и полагая $\alpha = i\pi\lambda z$, $\beta_1 = i2\pi(x' - x)$, $\beta_2 = i2\pi(y' - y)$, внутренний интеграл

$$I = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[-i2\pi\lambda z(u^2 + v^2) + i2\pi u(x' - x) + i2\pi v(y' - y)] dudv$$

можно вычислить. Получим

$$I = \frac{1}{i\lambda z} \exp\left[ik \frac{(x' - x)^2 + (y' - y)^2}{2z} \right].$$

Учитывая это, приходим к формуле (5.20)

При решении задачи дифракции на одномерных структурах, т.е. таких структурах, функция пропускания которых $t(x, y)$, а с ней и распределение поля $E(x, y)$, зависят только от одной переменной (например, x), т.е. когда $E(x, y) = E(x)$, интеграл

(5.20) упрощается. Подставив в этот интеграл указанное распределение поля и учтя, что

$$\begin{aligned} \int_{-\infty}^{\infty} \exp\left[ik \frac{(y' - y)^2}{2z}\right] dy &= \int_{-\infty}^{\infty} \exp\left[-k \frac{(y' - y)^2}{2iz}\right] dy = \\ &= \int_{-\infty}^{\infty} \exp\left(-k \frac{y^2}{2iz}\right) dy = \sqrt{i\lambda z}, \end{aligned}$$

так как

$$\int_{-\infty}^{\infty} \exp(-\alpha x^2 \pm \beta x) dx = \sqrt{\frac{\pi}{\alpha}} \exp\left(\frac{\beta^2}{4\alpha}\right), \quad (5.25)$$

а $\alpha = k/2iz$ и $\beta = 0$, приходим к соотношению

$$E'(x') = \frac{e^{ikz}}{\sqrt{i\lambda z}} \int_{-\infty}^{\infty} E(x) \exp\left[ik \frac{(x' - x)^2}{2z}\right] dx \quad (5.26)$$

Приближение Френеля описывает френелевскую дифракцию для любой формы волнового фронта. При рассмотрении дифракции Френеля сферических волн следует исходить из интеграла (5.5). Пусть экран освещается сферической волной от точечного источника S , расположенного в точке $(0, 0, -a)$. Тогда в параксиальном приближении для радиуса r_0 сферической волны находим

$$r_0 = \sqrt{a^2 + x^2 + y^2} \approx a + \frac{x^2 + y^2}{2a}$$

при условии, что

$$k \frac{(x^2 + y^2)^2}{8a^3} \ll \pi$$

или при

$$x^2 + y^2 \ll 2\sqrt{\lambda a^3}.$$

Это более сильное условие, чем условие $x^2 + y^2 \ll a^2$. Например, при $a = 1$ м, $\lambda = 0,5$ мкм максимальный размер области вокруг оси Z , где выполняется условие дифракции Френеля,

$$R_{\max} = \sqrt{(x^2 + y^2)_{\max}} = \sqrt{2\sqrt{\lambda a^3}} \approx 12 \text{ см.}$$

в то время как условие $x^2 + y^2 \ll a^2$ дает $R_{\max} = 1$ м. В знаменателе подынтегрального выражения интеграла (5.5) радиус r_0 можно заменить расстоянием a и вынести его из-под знака интеграла. Тогда с учетом (5.16), (5.17) и (5.19) интеграл суперпозиции в приближении Френеля для сферических волн приводится к виду

$$E'(x', y') \frac{e^{ik(a+z)}}{ia\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \times \\ \times \exp \left[ik \left(\frac{x^2 + y^2}{2a} + \frac{(x' - x)^2 + (y' - y)^2}{2z} \right) \right] dx dy. \quad (5.27)$$

Интегральные соотношения (5.20) и (5.27) позволяют значительно упростить решение френелевских дифракционных задач. Ниже мы рассмотрим несколько таких задач.

5.6. Дифракция Френеля на круглом отверстии

В качестве первого примера применения интеграла суперпозиции в приближении Френеля для решения дифракционных задач рассмотрим дифракцию Френеля плоских и сферических волн на круглом отверстии. Нас будет интересовать

распределение амплитуды светового поля вдоль оси отверстия, совпадающей с координатной осью Z . Рассмотрим сначала дифракцию плоских волн с амплитудой

$$E(z) = E_0 \exp(ikz),$$

падающих нормально плоскости отверстия. Схема опыта показана на рис. 5.4. Будем считать, что расстояние z от отверстия до точки наблюдения P велико (т.е. такое, что $z \gg R$), где R – радиус отверстия. Тогда условие приближения Френеля будет выполняться и, следовательно, для расчета поля в указанной точке $P(0, 0, z)$ можно использовать интеграл суперпозиции (5.20). Так как для рассматриваемой точки $x' = y' = 0$, то

$$E'(0, 0, z) = \frac{e^{ikz}}{i\lambda z} \iint_{\sigma} E(x, y) \exp\left(ik \frac{x^2 + y^2}{2z}\right) dx dy. \quad (5.28)$$

Интегрирование производится по площади отверстия σ . Форма отверстия обладает осевой симметрией, поэтому при вычислении этого интеграла в плоскости отверстия удобно перейти к полярным координатам ρ и θ , поместив начало координат в центре отверстия. Тогда

$$x = \rho \cos \theta, \quad y = \rho \sin \theta, \quad \rho = \sqrt{x^2 + y^2}, \quad dx dy = \rho d\rho d\theta.$$

Функцию пропускания $t(x, y) = t(\rho)$ принимаем равной единице в пределах отверстия и нулю вне этих пределов, т.е.

$$t(\rho) = \begin{cases} 1, & \rho \leq R, \\ 0, & \rho > R \end{cases}$$

Функцию такого вида называют круговой функцией аргумента ρ/R и обозначают как $\text{circ}(\rho/R)$. Таким образом,

$$t(\rho) = \text{circ}(\rho/R).$$

Амплитуда светового поля в плоскости XY , расположенной непосредственно за отверстием, определится как

$$E(x, y) = E(\rho) = \begin{cases} E_0, & \rho \leq R; \\ 0, & \rho > R, \end{cases}$$

где E_0 – амплитуда падающей на экран плоской волны в плоскости $z = 0$. Учитывая это, получим

$$E'(0, 0, z) = E_0 \frac{e^{ikz}}{i\lambda z} \int_0^R \int_0^{2\pi} \exp\left(ik \frac{\rho^2}{2z}\right) \rho d\rho d\theta. \quad (5.29)$$

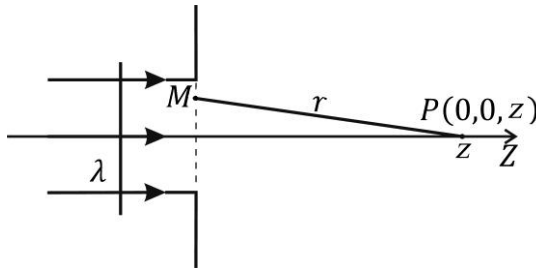


Рис. 5.4

Вычисление этого двойного интеграла приводит к выражению

$$E'(0, 0, z) = E_0 \frac{\exp(ikz)}{i\lambda z} \frac{z}{ik} 2\pi \left[\exp\left(ik \frac{R^2}{2z}\right) - 1 \right].$$

После преобразования получим распределение комплексной амплитуды светового поля вдоль оси Z :

$$E'(0, 0, z) = -2i E_0 e^{ikz} \exp\left(ik \frac{R^2}{4z}\right) \sin\left(k \frac{R^2}{4z}\right). \quad (5.30)$$

При получении этого выражения была использована формула

$$\exp(i\varphi) - \exp(-i\varphi) = 2i\sin\varphi.$$

Распределение интенсивности света вдоль оси Z определится как

$$I(0, 0, z) = I_0 \sin^2 \left(k \frac{R^2}{4z} \right) = I_0 \sin^2 \left(\frac{\pi R^2}{2\lambda} \frac{1}{z} \right), \quad (5.31)$$

где $I_0 \sim 4|E_0|^2$. Из (5.31) видно, что интенсивность света на оси отверстия изменяется по закону $\sin^2 \frac{1}{z}$. Видно также, что эта интенсивность определяется отношением площади отверстия к расстоянию до точки, в которой она определяется.

Зависимость интенсивности от расстояния z до отверстия показана на рис 5.5. Из графика $I(z)$ видно, что интенсивность света на оси пучка осциллирует в пределах от нуля до $4I_0$. Максимального значения она достигает во всех тех точках P_m , расстояние z_m до которых от центра отверстия удовлетворяет соотношению

$$\pi R^2 / 2\lambda z_m = (2m - 1)\pi / 2,$$

т.е. для расстояний

$$z_m = R^2 / [(2m - 1)\lambda].$$

Интенсивность равна нулю при

$$\pi R^2 / 2\lambda z_m = 2m\pi / 2,$$

т.е. на расстояниях

$$z_m = R^2 / (2m\lambda),$$

где $m = 1, 2, 3, \dots$ Таким образом, перемещаясь вдоль оси отверстия, мы будем последовательно переходить от максимумов интенсивности к минимумам и наоборот. При $z > R^2 / \lambda$ интенсивность света убывает, стремясь к нулю.

Дифракцию сферических волн можно рассмотреть с помощью интеграла (5.27), который при $x' = y' = 0$ принимает вид

$$E'(0, 0, z) = \frac{e^{ik(a+z)}}{i\lambda az} \iint_{\sigma} E(x, y) \exp \left[ik \left(\frac{1}{a} + \frac{1}{z} \right) \frac{x^2 + y^2}{2} \right] dx dy. \quad (5.32)$$

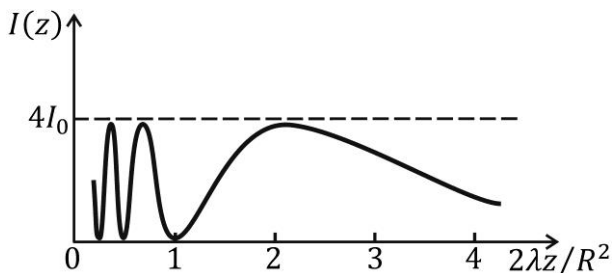


Рис. 5.5

Подынтегральные выражения интегралов (5.32) и (5.28) различаются только коэффициентом перед выражением $\frac{x^2 + y^2}{2}$ в показателе экспоненты, поэтому, не повторяя тех же выкладок, которые привели к выражению (5.30), можно сразу записать конечный результат, заменив в формулах (5.30) и (5.31) $\frac{1}{z}$ на $\frac{1}{a} + \frac{1}{z}$. Тогда для интенсивности в точке P на оси Z будем иметь

$$I(0, 0, z) = I_0 \sin^2 \left[k \frac{R^2}{4} \left(\frac{1}{a} + \frac{1}{z} \right) \right] = I_0 \sin^2 \left[\frac{\pi R^2}{2\lambda} \left(\frac{1}{a} + \frac{1}{z} \right) \right].$$

Расстояния z_m , соответствующие максимумам интенсивности $I(0, 0, z)$, находятся из условия

$$\frac{\pi R^2}{2\lambda} \left(\frac{1}{a} + \frac{1}{z_m} \right) = (2m - 1) \frac{\pi}{2},$$

а минимумы – из условия

$$\frac{\pi R^2}{2\lambda} \left(\frac{1}{a} + \frac{1}{z_m} \right) = 2m \frac{\pi}{2}.$$

При этом максимальное значение интенсивности равно I_0 , а минимальное – нулю.

Можно показать, что при удалении от точки P в направлении, перпендикулярном оси отверстия, интенсивность будет периодически изменяться, переходя от максимума к минимуму и наоборот. В результате на экране, расположенном перпендикулярно оси отверстия, будет наблюдаться дифракционная картина в виде чередующихся светлых и темных колец, центры которых будут располагаться в точке P на оси Z . При этом в центре картины будет наблюдаться светлое или темное пятно в зависимости от расстояния z экрана до отверстия.

5.7. Дифракция Френеля на круглом диске

В случае дифракции на круглом диске задача решается точно так же, как и задача дифракции на круглом отверстии, с использованием полярных координат в плоскости диска. Функция пропускания и в этом случае будет зависеть только от полярного радиуса ρ , она равна нулю в пределах диска и единице вне пределов диска, т.е.

$$t(\rho) = \begin{cases} 0, & \rho \leq R, \\ 1, & \rho > R, \end{cases}$$

где $\rho = \sqrt{x^2 + y^2}$ – расстояние от центра диска до его произвольной точки, R – радиус диска. Амплитуда светового поля в плоскости XY , расположенной непосредственно за диском, при освещении диска нормально падающей плоской волной будет

$$E(\rho) = \begin{cases} 0, & \rho \leq R, \\ E_0, & \rho > R, \end{cases}$$

где E_0 – амплитуда падающей на диск плоской волны в плоскости $z = 0$.

Повторяя те же выкладки, что и при определении поля, создаваемого отверстием в точке $P(0, 0, z)$, придем к интегралу вида (5.29) с той лишь разницей, что интегрирование по ρ теперь должно производиться в пределах от R до ∞ . Но при таком верхнем пределе интеграл по ρ будет расходиться. Чтобы избежать этого, квадратичный фазовый множитель в подынтегральном выражении запишем в виде $\exp\left(-k \frac{\rho^2}{2iz}\right)$.

Тогда получим

$$E'(0, 0, z) = E_0 \frac{e^{-ikz}}{i\lambda z} \int_R^\infty \int_0^{2\pi} \exp\left(-k \frac{\rho^2}{2iz}\right) \rho \, d\rho \, d\theta.$$

Вычислив этот интеграл и вернувшись к прежнему виду квадратичного фазового множителя, будем иметь

$$E'(0, 0, z) = E_0 \frac{e^{-ikz}}{i\lambda z} \frac{z}{-ik} 2\pi \exp\left(ik \frac{R^2}{2z}\right).$$

После преобразований получим распределение комплексной амплитуды вдоль оси диска

$$E'(0, 0, z) = E_0 e^{-ikz} \exp\left(ik \frac{R^2}{2z}\right).$$

Распределение интенсивности света

$$I(0, 0, z) = |E_0|^2.$$

Видим, что интенсивность света не зависит от координаты z , т.е. одинакова для всех точек P , расположенных на оси диска, и равна интенсивности падающей на диск волны. Дело обстоит так, как если бы волна распространялась свободно. Ни в какой точке на оси диска интенсивность не обращается в нуль, поэтому в окрестности любой такой точки будет наблюдаться светлое пятно (его называют пятном или диском Пуассона), которое, как и при дифракции на круглом отверстии, будет окружено чередующимися темными и светлыми кольцами.

Аналогично можно показать, что такой же результат получается и при дифракции на диске сферических волн.

5.8. Дифракция Френеля на крае полуплоскости

Решим теперь практически важную задачу определения интенсивности света вблизи границы тени, отбрасываемой краем протяженного экрана (полуплоскостью) при падении на него световой волны. Как и прежде, рассмотрим случаи плоской и сферической волн. Ось Y координатной плоскости XU , параллельной плоскости экрана и расположенной непосредственно за ним, направим параллельно краю экрана, а ось X – перпендикулярно ему (рис. 5.6).

Так как экран протяженный, то его апертурная функция будет зависеть только от координаты, перпендикулярной краю

экрана. При выбранном направлении координатных осей – от координаты x : $t(x, y) = t(x)$, причем

$$t(x) = \begin{cases} 0, & x < 0; \\ 1, & x \geq 0. \end{cases}$$

Функция такого вида в математике называется функцией скачка, или единичной функцией Хевисайда и обозначается как $1(x)$; следовательно, $t(x) = 1(x)$. Графически она изображается в виде ступеньки высотой, равной единице. Если скачок происходит не в начале координат, а в точке x_0 , то эта функция обозначается как $1(x - x_0)$. При $x < x_0$ она принимает значения, равные нулю, а при $x \geq x_0$ – единице.

Амплитуду поля будем искать в точке P с координатами $(x', 0, z)$.

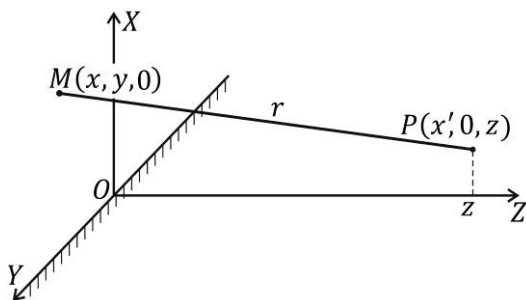


Рис. 5.6

Поскольку нас интересует распределение светового поля вблизи края геометрической тени, т.е. малые значения x' ($x' \ll z$), то применимость приближения Френеля будет обеспечена.

Поскольку, кроме того, функция пропускания экрана зависит только от одной коор-

динаты, задача является одномерной, и поэтому амплитуду светового поля в точке P в случае падения на экран плоской волны можно вычислить с помощью интегрального соотношения (5.24).

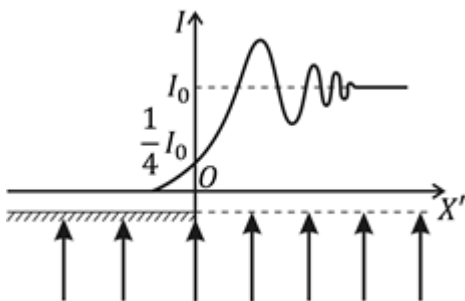


Рис. 5.7

Поскольку мы рассматриваем малые отклонения от геометрической оптики (т.е. от прямолинейного распространения света), то основной вклад в интеграл по x будут вносить лишь участки плоскости XY , близко расположенные к краю экрана. Поэтому в интеграле (5.26) верхний предел по переменной x можно оставить равным бесконечности, а нижний положить равным нулю. Учитывая при этом, что при указанной апертурной функции поле $E(x) = E_0$ при $x \geq 0$ и $E(x) = 0$ при $x < 0$, будем иметь

$$E'(x', 0) = C \int_0^{\infty} \exp \left[ik \frac{(x' - x)^2}{2z} \right] dx,$$

где постоянная $C = \frac{E_0 \exp(ikz)}{\sqrt{i\lambda z}}$.

Введя вместо x новую переменную интегрирования $\eta = \sqrt{k/\pi z} (x - x')$, этот интеграл приведем к виду

$$E'(x', 0) = \text{const} \int_{-\infty}^{\infty} e^{i\pi\eta^2/2} d\eta. \quad (5.33)$$

Здесь $s = \sqrt{k/\pi z} x' = \sqrt{2/\lambda z} x'$, а const – другая постоянная, равная первоначальной постоянной, умноженной на $\sqrt{\pi z/k} = \sqrt{\lambda z/2}$: $\text{const} = E_0 \exp(ikz)/\sqrt{2i}$. Интеграл в (5.33) называется интегралом Френеля.

Так как s содержит в виде множителя величину $\sqrt{k} \gg 1$, то $|s| \gg 1$. При этом следует различать три случая:

1) $x' > 0$, т.е. $s > 0$, 2) $x' < 0$, т.е. $s < 0$, 3) $x' = 0$, т.е. $s = 0$.

В первом случае точка наблюдения P расположена в освещенной области, во втором – в области геометрической тени, отбрасываемой экраном; в третьем случае указанная точка лежит как раз на краю геометрической тени. В первом случае

$$\int_{-s}^{\infty} e^{i\pi\eta^2/2} d\eta \approx \sqrt{2i} + \sqrt{\frac{2}{\pi}} \frac{e^{is^2}}{2is}, \quad s > 0,$$

во втором случае

$$\int_{-s}^{\infty} e^{i\pi\eta^2/2} d\eta \approx \sqrt{\frac{2}{\pi}} \frac{e^{is^2}}{2is}, \quad s < 0,$$

в третьем случае

$$\int_{-s}^{\infty} e^{i\pi\eta^2/2} d\eta \approx \sqrt{i/2}, \quad s = 0.$$

Амплитуды полей в указанных областях найдем, подставив в результаты вычисления интегралов выражение $s = \sqrt{2/\lambda z} x'$ и умножив их на $\text{const} = E_0 \exp(ikz)/\sqrt{2i}$. Будем иметь

$$E'(x', 0) = E_0 e^{ikz} \left(1 + \sqrt{\frac{\lambda z}{2i\pi}} \frac{\exp(2i/\lambda z x'^2)}{2ix'} \right), \quad x' > 0, \quad (5.34)$$

$$E'(x', 0) = E_0 e^{ikz} \sqrt{\frac{\lambda z}{2i\pi}} \frac{\exp(2i/\lambda z x'^2)}{2ix'}, \quad x' < 0, \quad (5.35)$$

$$E'(x', 0) = \frac{1}{2} E_0 e^{ikz}, \quad x' = 0. \quad (5.36)$$

Интенсивность света $I(x', 0)$ при дифракции плоских волн будет определяться квадратом модуля поля $E'(x', 0)$ с учетом (5.34), (5.35), (5.36). Графически зависимость интенсивности $I(x', 0)$ представлена на рис. 5.7, где I_0 – интенсивность света при полностью открытом фронте, т.е. при отсутствии экрана. Видно, что резкой границы между светом и тенью нет. В области геометрической тени ($x' < 0$) интенсивность не спадает резко до нуля, а убывает непрерывно и монотонно по закону $1/x'^2$. При $x' = 0$ интенсивность $I(0, 0) = I_0 / 4$. В освещенной же области ($x' > 0$) интенсивность света не всюду равна I_0 , а, осциллируя около этого значения (из-за наличия второго члена в выражении $I(x', 0)$ при $x' > 0$), постепенно приближается к I_0 . При этом высота первого максимума над уровнем I_0 более чем на одну треть превышает значение I_0 . Следовательно, вблизи края протяженного экрана в освещенной области возникает дифракционная картина в виде чередующихся светлых (убывающей освещенности) и темных полос, параллельных краю экрана. При этом с увеличением расстояния x' от края геометрической тени расстояние между полосами уменьшается, полосы сгущаются и становятся все менее контрастными.

К тем же интегралам Френеля приходим и при падении на экран сферической волны, исходящей из точки $S(0, 0, -a)$. В этом случае амплитуда светового поля в точке $P(0, 0, z)$ определится интегралом (5.27). Положив в этом интеграле $y' = 0$, получим

$$E'(x', 0) = \frac{e^{ik(a+z)}}{i\lambda a z} \int_0^{\infty} \int_{-\infty}^{\infty} \exp \left\{ ik \left[\frac{x^2}{2a} + \frac{(x' - x)^2}{2z} + \left(\frac{1}{a} + \frac{1}{z} \right) \frac{y^2}{2} \right] \right\} dx dy.$$

Проинтегрировав по y , как и в предыдущем случае, будем иметь

$$E'(x', 0) = \text{const} \cdot \int_0^{\infty} \exp \left\{ ik \left[\frac{x^2}{2a} + \frac{(x' - x)^2}{2z} \right] \right\} dx,$$

где const включает в себя как результат интегрирования по y –

выражение $\sqrt{\frac{i\lambda a z}{a+z}}$, так и множитель перед интегралом в

предыдущем выражении. Введем новую переменную

интегрирования $\eta = \sqrt{\frac{k(a+z)}{\pi a z}} \left(x - \frac{ax'}{a+z} \right)$, тогда после

некоторых преобразований придем к интегралу Френеля, в

котором нижний предел интегрирования $s = \sqrt{\frac{ka}{\pi z(a+z)}} x'$.

Дальнейший анализ этого интеграла аналогичен предыдущему.

5.9. Дифракция Френеля на щели

Решим теперь задачу о дифракции плоских волн на щели. Щелью называют прямоугольное отверстие, один из размеров которого (длина) много больше другого (ширины).

Пусть щель, ширина которой равна a , равномерно освещается нормально падающим монохроматическим пучком света с длиной волны λ и амплитудой E_0 . Расположим непосредственно за щелью координатную плоскость XU и ориентируем ее так, чтобы ось X

была направлена перпендикулярно краям щели (вдоль размера a); ось Z перпендикулярно плоскости щели, а начало координат поместим в точку, расположенную посередине щели (рис. 5.8, a).

Апертурная функция щели будет зависеть от одной только поперечной координаты x : $t(x, y) = t(x)$. В выбранной системе координат она будет иметь вид:

$$t(x) = \begin{cases} 1, & |x| \leq a/2; \\ 0, & |x| > a/2. \end{cases}$$

Функция такого вида называется прямоугольной функцией аргумента x/a и обозначается $\text{rect}(x/a)$. Следовательно, $t(x) = \text{rect}(x/a)$. Графически функция $t(x)$ изображается в виде прямоугольника с основанием, равным a , и высотой, равной единице (рис. 5.8, b). Такую функцию можно представить в виде разности двух скачков: одного в точке $x = -a/2$, другого в точке $x = a/2$:

$$\text{rect}(x/a) = 1(x + a/2) - 1(x - a/2).$$

При такой апертурной функции переменные в интеграле (5.20), как и в предыдущем случае, разделяются и интегрирование по координате y в бесконечных пределах (так как предполагается, что длина щели $b \rightarrow \infty$) даст постоянный множитель $\sqrt{i\lambda z}$. С учетом этого амплитуда поля в произвольной точке $P(x', 0, z)$ определится как

$$\begin{aligned} E'(x', 0, z) &= CE_0 \int_{-\infty}^{\infty} \exp \left[ik \frac{(x' - x)^2}{2z} \right] dx = \\ &= CE_0 \int_{-a/2}^{a/2} \exp \left[ik \frac{(x' - x)^2}{2z} \right] dx, \end{aligned} \quad (5.37)$$

где $C = \frac{\exp(ikz)}{\sqrt{i\lambda z}}$ – комплексная постоянная.

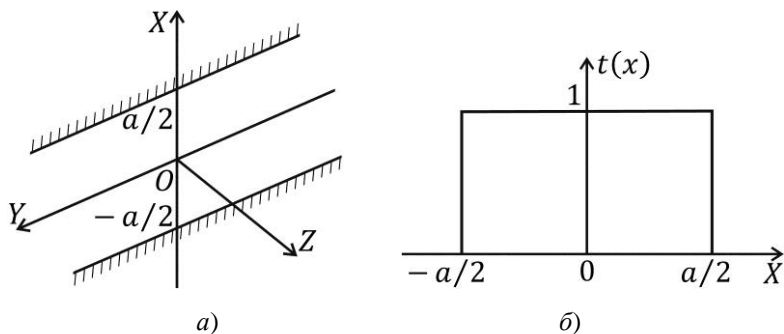


Рис. 5.8

Будем считать, что щель создана краями двух параллельных экранов; один край (нижний) расположен вдоль прямой $x = -a/2$, а другой (верхний) – вдоль прямой $x = a/2$. Тогда выражение (5.35) можно представить в виде

$$E'(x', 0, z) = CE_0 \int_{-\infty}^{a/2} \exp\left[ik \frac{(x' - x)^2}{2z}\right] dx +$$

$$+ CE_0 \int_{-a/2}^{\infty} \exp\left[ik \frac{(x' - x)^2}{2z}\right] dx - CE_0 \int_{-\infty}^{\infty} \exp\left[ik \frac{(x' - x)^2}{2z}\right] dx.$$

Здесь первое слагаемое описывает дифракцию на верхнем крае, второе – на нижнем крае. Интеграл в третьем слагаемом может быть вычислен с помощью формулы (5.25), если положить в ней $\alpha = -ik/2z$, $\beta = ikx'/z$. В результате получим комплексное число $\sqrt{i\lambda z} \exp\left(ik \frac{x'^2}{2z}\right)$. С учетом этого третье слагаемое примет вид

$$E_0 \frac{\exp(ikz)}{\sqrt{i\lambda z}} \sqrt{i\lambda z} \exp\left(ik \frac{x'^2}{2z}\right) = E_0 \exp\left(ik \frac{x'^2}{2z}\right) \exp(ikz).$$

Следовательно, третье слагаемое без учета квадратичного фазового множителя представляет собой комплексную амплитуду падающей плоской волны.

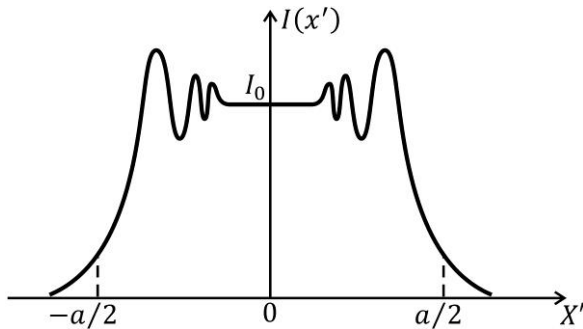


Рис. 5.9

Характер результирующей дифракционной картины определяется дифракцией на двух противоположных краях щели и падающей на щель плоской волной. Вид зависимости интенсивности результирующего поля от координаты x' существенно зависит от ширины щели, т.е. от соотношения между шириной щели a и расстоянием до нее z , точнее от безразмерного параметра $p = \sqrt{\lambda z} / a$. При большой ширине щели ($a \gg \sqrt{\lambda z}$) распределение интенсивности $I(x')$ световой волны в области напротив середины щели практически постоянно. Дифракционные полосы наблюдаются на экране напротив краев щели (рис. 5.9). Зависимость $I(x')$ при малой ширине щели, когда $a = \sqrt{2\lambda z}$, показана на рис. 5.10. Видим, что при такой ширине щели в центре дифракционной картины находится светлая полоса без каких-либо заметных колебаний интенсивности. Быстро

затухающие колебания интенсивности происходят лишь в области геометрической тени при $|x'| > a/2$.

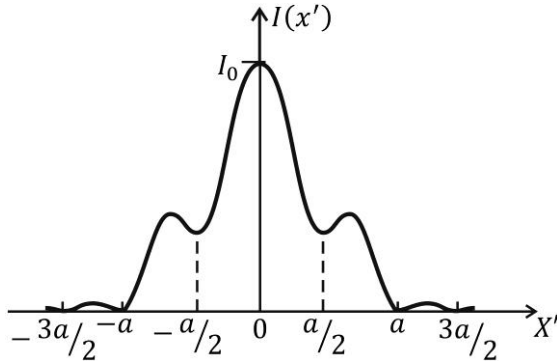


Рис. 5.10

Рассмотрим предельный случай бесконечно узкой щели. Апертурная функция такой щели описывается дельта-функцией $\delta(x)$. Тогда для поля в точке $P(x', 0, z)$ с учетом фильтрующего свойства δ -функции будем иметь

$$\begin{aligned} E'(x', 0, z) &= CE_0 \int_{-\infty}^{\infty} \delta(x) \exp \left[ik \frac{(x' - x)^2}{2z} \right] dx = \\ &= CE_0 \exp \left(ik \frac{x'^2}{2z} \right). \end{aligned}$$

Полученное выражение описывает поле цилиндрической волны, радиус волнового фронта которой равен расстоянию до источника z . При малых значениях x' (вблизи оси Z) экспоненциальный множитель можно заменить единицей. Тогда получим

$$E'(x', 0, z) = CE_0 = E_0 \frac{\exp(ikz)}{\sqrt{i\lambda z}}.$$

Это поле не зависит от x' и, следовательно, одинаково во всех точках плоскости $z = \text{const}$. Такое поле создается плоской волной, падающей на плоскость $z = \text{const}$.

Глава 6

ДИФРАКЦИЯ ФРАУНГОФЕРА

6.1. Приближение Фраунгофера

Перейдем теперь к рассмотрению второго предельного случая интегрального соотношения Кирхгофа – Френеля – Зоммерфельда, соответствующего дифракции Фраунгофера. Дифракция Фраунгофера имеет большое практическое значение для общей теории оптических приборов и теории дифракционных решеток.

Раскрыв скобки в аргументе экспоненты в (5.20) и вынеся из-под знака интеграла независящий от переменных интегрирования

x и y множитель $\exp\left(ik \frac{x'^2 + y'^2}{2z}\right)$, интеграл (5.20) можно

записать в виде

$$E'(x', y') = \frac{e^{ikz}}{i\lambda z} e^{ik \frac{x'^2 + y'^2}{2z}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp\left(ik \frac{x^2 + y^2}{2z}\right) \times \\ \times \exp\left[-i\left(\frac{kx'}{z}x + \frac{ky'}{z}y\right)\right] dx dy. \quad (6.1)$$

При достаточно больших расстояниях z , соответствующих

дальней зоне, когда $k \frac{(x^2 + y^2)_{\max}}{2z} \ll \pi$, т.е. когда

$$z \gg (x^2 + y^2)_{\max} / \lambda,$$

или учтя, что величина $(x^2 + y^2)_{\max} \approx R^2$, где R – максимальный размер отверстия, $z \gg R^2 / \lambda$, квадратичный фазовый множитель $\exp\left(ik \frac{x^2 + y^2}{2z}\right)$ в подынтегральном выражении (6.1) можно

положить равным единице. В результате интеграл суперпозиции приобретает вид

$$E'(x', y') = \frac{e^{ikz}}{i\lambda z} e^{ik \frac{x'^2 + y'^2}{2z}} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp\left[-i\left(\frac{kx'}{z}x + \frac{ky'}{z}y\right)\right] dx dy. \quad (6.2)$$

Если в плоскости $X'Y'$ рассматривать область, расположенную вблизи оптической оси Z , такую, что $k \frac{(x'^2 + y'^2)}{2z} \ll \pi$ или $R' = \sqrt{x'^2 + y'^2} \ll \sqrt{\lambda z}$, то квадратичный фазовый множитель перед интегралом (6.2) можно положить равным единице. Тогда получим

$$E'(x', y') = \frac{e^{ikz}}{i\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp\left[-i2\pi\left(\frac{x'}{\lambda z}x + \frac{y'}{\lambda z}y\right)\right] dx dy. \quad (6.3)$$

Полученный интеграл называют интегралом суперпозиции в приближении Фраунгофера. Он является основой описания дифракционных явлений Фраунгофера. Поскольку этот интеграл получается из интеграла (5.20), а последний, как было показано в п. 5.5, можно получить также и из представления Рэлея, то и интеграл (6.3) может быть получен из этого представления. Тем самым мы показали, что спектральный метод Рэлея приводит к тем же результатам, что и координатный метод Гюйгенса –

Френеля, выраженный в виде интеграла Кирхгофа – Френеля – Зоммерфельда.

Соотношение (6.3) можно переписать в виде

$$E'(x', y') = \frac{e^{ikz}}{i\lambda z} E\left(\frac{x'}{\lambda z}, \frac{y'}{\lambda z}\right),$$

где

$$E\left(\frac{x'}{\lambda z}, \frac{y'}{\lambda z}\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp\left[-i2\pi\left(\frac{x'}{\lambda z}x + \frac{y'}{\lambda z}y\right)\right] dx dy -$$

фурье-образ (спектральная плотность) поля $E(x, y)$.

Условия $z \gg R^2 / \lambda$, $x'^2 + y'^2 \ll \lambda z$ называют приближениями Фраунгофера, или условиями дифракции Фраунгофера. Неравенство $z \gg R^2 / \lambda$ (или $p = \sqrt{\lambda z} / R \gg 1$) является весьма серьезным ограничением на расстояние z и в дифракционных опытах выполняется крайне редко. Действительно, при $\lambda = 0,5$ мкм и максимальном размере отверстия $R = 1$ см имеем $z \gg 200$ м. Однако дифракцию Фраунгофера можно наблюдать и на расстояниях, не удовлетворяющих этому неравенству, если, например, используется сферическая волна, сходящаяся к наблюдателю. Дифракцию Фраунгофера можно получить и тогда, когда между плоскостями XU и $X'U'$ соответствующим образом расположить собирающую линзу. Так как отверстие освещается плоской волной (параллельным пучком света, получающимся если точечный источник S поместить в фокусе собирающей линзы, расположенной перед объектным экраном), то дифракционная картина наблюдается в фокальной плоскости линзы, поставленной на пути дифрагировавшего света. Дело обстоит так, как если бы расстояние z было равно бесконечности. В этом случае второе условие дифракции Фраунгофера выполняется автоматически,

поэтому квадратичный фазовый множитель $\exp\left(ik \frac{x'^2 + y'^2}{2z}\right)$

обращается в единицу. В теоретических рассуждениях от всех вспомогательных приспособлений, применяющихся при практическом осуществлении дифракции Фраунгофера, можно отвлечься и считать, что плоскость $X'Y'$ находится на расстоянии z от экрана, удовлетворяющем условиям дифракции Фраунгофера.

Распределение интенсивности света в плоскости $X'Y'$ в приближении Фраунгофера

$$I(x', y') = |E(x', y')|^2 = (\lambda z)^{-2} \left| E\left(\frac{x'}{\lambda z}, \frac{y'}{\lambda z}\right) \right|^2.$$

Практическая важность рассмотренных двух приближений (Френеля и Фраунгофера), таким образом, заключается в том, что первое из них описывает дифракцию Френеля, а второе – дифракцию Фраунгофера.

Пусть имеем распределение амплитуды поля в плоскости XU . При определении поля по другую сторону от экрана с отверстием или от транспаранта, создающих данное распределение поля в плоскости XU , пространство справа от этих оптических элементов можно условно разбить на две области: вблизи от экрана ($z \approx R^2 / \lambda$) – область Френеля, а дальше ($z \gg R^2 / \lambda$) – область Фраунгофера. Для определения поля в области дифракции Френеля следует использовать приближение Френеля, а в области дифракции Фраунгофера – приближение Фраунгофера. Случай $z \ll R^2 / \lambda$, как уже отмечалось, соответствует геометрической оптике (приближению тени), когда наблюдается резкое изображение препятствия, соответствующее фактически прямолинейному распространению света.

6.2. Дифракция Фраунгофера как пространственное преобразование Фурье

Соотношение (6.3) показывает, что в приближении Фраунгофера распределение комплексной амплитуды поля в плоскости $X'Y'$ пропорционально фурье-спектру распределения поля $E(x, y)$ в плоскости XY , расположенной непосредственно за отверстием, причем роль пространственных частот u и v играют соответственно величины $x' / \lambda z$ и $y' / \lambda z$:

$$\begin{aligned}
 E'(x', y') &= \frac{e^{ikz}}{i\lambda z} E\left(\frac{x'}{\lambda z}, \frac{y'}{\lambda z}\right) = \\
 &= \frac{e^{ikz}}{i\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp[-i2\pi(ux + vy)] dx dy = \\
 &= \frac{e^{ikz}}{i\lambda z} E(u, v). \tag{6.4}
 \end{aligned}$$

Тот факт, что поле $E'(x', y')$ определяется двумерным преобразованием Фурье функции $E(x, y)$, следует и из соотношения (6.2), когда условие $x'^2 + y'^2 \ll \lambda z$ не выполняется. Но в этом случае из-за наличия перед интегралом квадратичного фазового множителя $\exp\left(ik \frac{x'^2 + y'^2}{2z}\right)$, описывающего расходящуюся сферическую волну в параболическом приближении, фурье-образ функции $E(x, y)$ создается не в плоскости $X'Y'$, а на сфере радиуса z с центром в начале координат O системы XYZ . Эта сфера называется сферой Гаусса.

Если не учитывать несущественный множитель перед интегралом в соотношении (6.4), то к этому соотношению можно прийти и непосредственно, используя принцип Гюйгенса – Френеля и принцип суперпозиции. Возьмем в плоскости $X'Y'$,

расположенной непосредственно за экраном, некоторую точку M , являющуюся центром элемента поверхности $dxdy$ этой плоскости. Этот элемент испускает элементарную волну с комплексной амплитудой $E(x, y)dxdy$. В направлении, определяемом направляющими косинусами $\cos \alpha = u\lambda$, $\cos \beta = v\lambda$, $\cos \gamma = w\lambda$, волна, испущенная из точки M , опережает по фазе на $k_x x + k_y y = 2\pi(ux + vy)$ элементарную волну, испущенную из центра отверстия O . По принципу суперпозиции комплексная амплитуда волны в направлении, определяемом указанными направляющими косинусами, равняется сумме этих амплитуд:

$$\iint_{\sigma} E(x, y) \exp[-i2\pi(ux + vy)] dxdy ,$$

где интегрирование производится по площади отверстия σ .

Легко убедиться, что указанные выражения пространственных частот u и v совпадают с исходными их определениями как отношений косинусов направляющих углов к длине волны, если учесть, что $x' / z \approx \operatorname{tg}\varphi$, $y' / z \approx \operatorname{tg}\theta$ и что при малых углах (которые имеют место в фраунгоферовской дифракции) тангенсы углов равны их синусам. Введя затем направляющие углы $\alpha = 90 - \varphi$ и $\beta = 90 - \theta$, придем к указанным отношениям.

При практическом осуществлении дифракции Фраунгофера, когда плоскость наблюдения $X'Y'$ является фокальной плоскостью линзы F' , в выражениях пространственных частот u и v вместо z следует брать фокусное расстояние линзы f и тогда

$$u = x' / \lambda f, \quad v = y' / \lambda f.$$

Таким образом, оптическая система, состоящая из дифракционного элемента (например, отверстия в непрозрачном экране) и большой глубины свободного пространства, с точностью до фазовых и амплитудных множителей осуществляет

преобразование Фурье поля $E(x, y)$ в плоскости XU непосредственно за экраном по пространственным частотам

$$u = x' / \lambda z, \quad v = y' / \lambda z.$$

Плоские волны, соответствующие каждой фурье-компоненте, распространяются под различными углами к оси Z . В плоскости XU' каждая фурье-компонента дает световое пятно. Расстояние этого светового пятна от оптической оси (оси Z) пропорционально пространственной частоте спектральной составляющей сигнала: $x' = \lambda z u$, $y' = \lambda z v$. Следовательно, вблизи оптической оси распространяется световое поле, соответствующее низким пространственным частотам спектра сигнала. Более высоким пространственным частотам соответствуют поля, распространяющиеся под большими углами к оптической оси и дающие световые пятна при больших значениях x' и y' . Чем больше частоты u и v , тем под большими углами распространяются пространственные гармоники. Если сигнал $E(x, y)$ – действительная функция x и y , то спектр его пространственных частот симметричен относительно оси Z . Пространственное разделение волн, дифрагировавших в различных направлениях, позволяет в плоскости XU' (на практике – в задней фокальной плоскости линзы, расположенной за объектным экраном) наблюдать отдельные фурье-компоненты распределения поля $E(x, y)$.

Любая дифракционная оптическая система, как мы увидим в дальнейшем, с помощью когерентных волн ставит в соответствие освещаемому объекту, с одной стороны, его изображение в некоторой плоскости, определяемой законами геометрической оптики, а с другой – двумерный фурье-образ в плоскости, определяемой законами дифракции. Следовательно, образование изображения и преобразование Фурье – два проявления одного и того же явления, называемого дифракцией.

Как уже отмечалось, распределение поля $E(x, y)$ в плоскости XU может быть создано, например, путем освещения плоской

монохроматической волной отверстия в экране с апертурной функцией $t(x, y)$ или транспаранта с функцией пропускания по амплитуде $t(x, y)$ ³. Во всех случаях $E(x, y) = E_0 t(x, y)$, где E_0 – амплитуда падающей плоской волны, и тогда

$$E'(x', y') = \text{const} \cdot E_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y) \exp[-i2\pi(ux + vy)] dx dy, \quad (6.5)$$

где const включает в себя все несущественные множители, а пространственные частоты $u = x' / \lambda z$, $v = y' / \lambda z$. Интегрирование производится в пределах, определяемых областью существования функции $t(x, y)$. Следовательно, дифракционную картину Фраунгофера, создаваемую плоским объектом, можно рассматривать как преобразование Фурье от его функции пропускания. Если такой объект поместить в пучок параллельных лучей монохроматического света, то за объектом вследствие дифракции образуется совокупность пучков, идущих под различными углами к объекту. С помощью собирающей линзы, установленной за объектом, эти пучки можно свести в ее задней фокальной плоскости. При этом все волны, идущие по одному направлению, т.е. имеющие одинаковые пространственные частоты, собираются линзой в одной точке. В результате в задней фокальной плоскости линзы получим распределение интенсивности, которое будет соответствовать распределению пространственных частот объекта.

Таким образом, при дифракции плоской монохроматической волны на любом плоском объекте в фокальной плоскости линзы происходит разложение функции пропускания объекта на отдельные гармонические составляющие (пространственные

³ Отверстие можно рассматривать как транспарант, функция пропускания которого принимает только два значения: единица в отверстиях и нуль – вне пределов отверстия.

гармоники), каждая из которых может быть охарактеризована своими пространственными частотами (определяющими направление распространения волны) и фазами. Можно сказать, что дифракционная картина, создаваемая объектом, представляет собой проявление, почти физическое воплощение тех гармоник, которые составляют его функцию пропускания (или, как говорят, его оптическую структуру).

Отождествление фраунгоферовской дифракционной картины для апертурной функции с преобразованием Фурье от этой апертурной функции приводит к трактовке линзы как устройства, выполняющего преобразования Фурье (строгое доказательство этого утверждения будет дано ниже). По этой причине заднюю фокальную плоскость линзы называют дифракционной плоскостью, фурье-плоскостью (фурье-пространством) или, иначе, плоскостью (областью) частот (пространственных).

В одномерном случае, когда функция пропускания зависит только от одной переменной x поле в точке $(x', 0)$ можно получить исходя из соотношения (5.26). Раскрыв скобки в подынтегральном фазовом множителе и принимая, что $kx^2/2z \ll \pi$ и $kx'^2/2z \ll \pi$, придем к соотношению

$$E'(x') = \frac{e^{ikz}}{\sqrt{i\lambda z}} \int_{-\infty}^{\infty} E(x) \exp(-i2\pi ux) dx,$$

или

$$E'(x') = \frac{e^{ikz}}{\sqrt{i\lambda z}} E_0 \int_{-\infty}^{\infty} t(x) \exp(-i2\pi ux) dx. \quad (6.6)$$

Следовательно, если апертурная функция зависит только от одной координаты, то распределение поля в фраунгоферовской дифракционной картине будет описываться одномерным преобразованием Фурье этой функции.

Иногда бывает удобно комплексную амплитуду $E'(x', y') =$

$= E'(u\lambda z, v\lambda z)$ представить в виде

$$E'(x', y') = C + i S,$$

где C и S – действительные функции пространственных частот u и v . Тогда, если отверстие симметрично относительно осей X и Y ($t(x, y)$ – четная функция координат x и y), то $S = 0$, и соотношение (6.5), как легко убедиться (учтя, что $\operatorname{Re} e^{\pm i\varphi} = \cos\varphi$), принимает вид

$$\begin{aligned} E'(x', y') &= \\ &= \operatorname{const} \cdot E_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y) \cos(2\pi ux) \cos(2\pi vy) dx dy. \end{aligned} \quad (6.7)$$

В одномерном случае, когда амплитудное пропускание является функцией одной переменной (x),

$$E(x') = \operatorname{const} \cdot E_0 \int_{-\infty}^{\infty} t(x) \cos(2\pi ux) dx. \quad (6.8)$$

Интенсивность в этом случае определяется обычным способом: $I(x', y') = C^2$, а при несимметричной функции $t(x, y)$ – по формуле

$$I(x', y') = C^2 + S^2.$$

Ниже рассматривается применение соотношений (6.6) – (6.8) к изучению дифракции Фраунгофера на различных апертурах.

6.3. Дифракция Фраунгофера на двух точечных отверстиях

Изучение дифракции Фраунгофера начнем с простейшего случая, показанного на рис. 6.1. Плоская монохроматическая волна падает нормально на объектный экран с двумя очень

узкими (точечными) отверстиями (отверстиями, радиусы которых

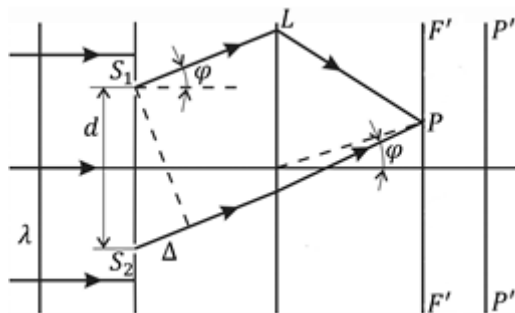


Рис. 6.1

меньше длины световой волны) S_1 и S_2 . Вследствие дифракции света на этих отверстиях в пространстве за экраном образуются два когерентных расходящихся перекрывающихся световых пучка –

две расходящиеся сферические волны.

Схема опыта такая же, как и схема интерференционного опыта Юнга. Однако для обеспечения выполнения условий наблюдения дифракции Фраунгофера в этой схеме дополнительно имеются две собирающие линзы, одна из которых располагается перед объектным экраном (на рис. 6.1 не показана) и служит для получения параллельного пучка света, освещающего объектный экран, а другая L располагается за этим экраном. Эта линза создает изображение точечных отверстий S_1 и S_2 в плоскости $P'P'$ (называемой плоскостью изображения).

В плане рассмотрения дифракции Фраунгофера непосредственный интерес для нас представляет задняя фокальная плоскость $F'F'$ линзы L . Возьмем на этой плоскости произвольную точку P , расположенную в направлении, определяемом углом φ к оптической оси линзы, совпадающей с координатной осью Z . В точке P сходятся и интерferируют друг с другом только волны, распространяющиеся от S_1 и S_2 в направлении угла φ (в опыте Юнга в точке P интерференция происходит между световыми волнами, распространяющимися от отверстий S_1 и S_2 в разных направлениях). Для определения результата сложения двух волн в точке P необходимо учесть

разность фаз δ между ними, обусловленную разностью хода Δ . Из рис. 6.1 имеем $\Delta = d \sin \varphi$, где d – расстояние между отверстиями, и тогда

$$\delta = k\Delta = kd \sin \varphi = k_x d = 2\pi u d,$$

Так как $k_x = 2\pi \sin \varphi / \lambda = 2\pi u$, $u = \sin \varphi / \lambda$ – пространственная частота волны, дифрагировавшей в направлении, определяемом углом φ . Интенсивность $I(P)$ в точке P определится формулой (3.1), если подставить в нее $\delta = (2\pi / \lambda) d \sin \varphi$:

$$I(P) = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos \left(\frac{2\pi}{\lambda} d \sin \varphi \right).$$

Интенсивности I_1 и I_2 равны, поскольку обе волны приходят в точку P под одним и тем же углом. Обозначив интенсивность, создаваемую одним отверстием через I_0 , с учетом сделанных замечаний получим

$$I(u) = 2I_0(1 + \cos 2\pi u d) = 4I_0 \cos^2 \pi u d. \quad (6.9)$$

К распределению (6.9) можно прийти и используя преобразование Фурье. Функцию пропускания объектного экрана можно представить в виде двух дельта-функций, определяющих положение отверстий на координатной оси X , проходящей через отверстия и начало отсчета которой расположено посередине расстояния между отверстиями:

$$t(x) = \delta(x + d/2) + \delta(x - d/2).$$

Поскольку функция пропускания зависит только от одной переменной (x), для определения поля в точке P следует использовать одномерное соотношение (6.6):

$$E(u) = E_0 \int_{-\infty}^{\infty} t(x) \exp(-i2\pi u x) dx,$$

где E_0 – амплитуда плоской волны, падающей на экран с отверстиями. Множитель $\exp(ikz)/\sqrt{i\lambda z}$ перед интегралом в (6.6) опущен, так как на распределение поля вдоль оси X он не влияет. Подставляя в этот интеграл вместо $t(x)$ выражение для функции пропускания и учитывая фильтрующее свойство дельта-функции, получим

$$E(u) = E_0 \int_{-\infty}^{\infty} \delta(x + d/2) \exp(-i2\pi ux) dx + \\ + E_0 \int_{-\infty}^{\infty} \delta(x - d/2) \exp(-i2\pi ux) dx = E_0 \exp(i\pi ud) + \\ + E_0 \exp(-i\pi ud) = 2E_0 \cos \pi ud.$$

Интенсивность светового поля в точке P

$$I(u) = 4I_0 \cos^2 \pi ud,$$

где $I_0 = |E_0|^2$ – интенсивность света в направлении распространения падающей на экран волны (при $\varphi = 0$). Таким образом, мы снова пришли к формуле (6.9).

Из выражения (6.9) видно, что максимумы интенсивности, равные $4I_0$, возникают при

$$ud = d \sin \varphi / \lambda = m,$$

т.е. в направлениях φ , определяемых соотношением

$$d \sin \varphi = m\lambda, \quad (6.10)$$

где $m = 0, \pm 1, \pm 2, \dots$. Число m называют порядком дифракции (или порядком дифракционного максимума). Следует отметить, что, хотя направления максимумов определяются величинами d и λ , их значения зависят от амплитуд волн, излучаемых в этих направлениях отдельными отверстиями.

Между максимумами лежат минимумы интенсивности, равные нулю. Таким образом, в задней фокальной плоскости линзы, расположенной за объектным экраном, возникает дифракционная картина в виде чередующихся светлых и темных полос. Так как интенсивность в дифракционной картине изменяется по закону $\cos^2(k\Delta/2)$, то эти полосы часто называют \cos^2 -полосами. Ширину h полос легко вычислить с помощью формулы (6.10). Если максимум первого порядка ($m = \pm 1$) расположен при $\varphi = \pm \varphi_1$, то вследствие малости φ_1 имеем $d \sin \varphi \approx d \cdot \varphi_1 = \lambda$, и тогда

$$h \approx f\varphi_1 = \lambda f / d, \quad (6.11)$$

где f – фокусное расстояние линзы.

Образование фраунгоферовской дифракционной картины в задней фокальной плоскости отображающей линзы, согласно теории Аббе – Портера, является первым (промежуточным) шагом формирования оптического изображения, выполняемом линзой. Второй (окончательный) шаг по этой теории заключается в следующем. Каждый дифракционный максимум в плоскости $F'F'$ представляет собой вторичный источник когерентных волн. Световые волны, распространяющиеся от этих вторичных источников, интерферируют в плоскости изображения PP' , образуя распределение освещенности (интенсивности), которое и является оптическим изображением объекта (в рассматриваемом случае – двух отверстий). Подробно вопросы формирования оптического изображения и, в частности, теория Аббе – Портера будут рассматриваться нами в главах 11 и 12.

6.4. Дифракция Фраунгофера на щели

В качестве другого простейшего примера применения приближения Фраунгофера рассмотрим снова дифракцию плоских волн на щели. Это приближение позволяет не только найти положение максимумов и минимумов в дифракционной

картине, но и получить функции распределения амплитуды и интенсивности поля в этой картине. Схема опыта показана на рис. 6.2. Как и при решении этой задачи в приближении Френеля, совместим коорди-

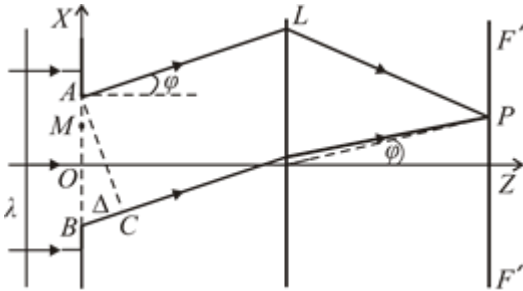


Рис. 6.2

натную плоскость XU с плоскостью, расположенной непосредственно за щелью, направив при этом ось X перпендикулярно краям щели и поместив начало координат в ее середине. Так как

длина щели b принимается очень большой ($b \gg a$), то дифракцию вдоль оси Y можно не учитывать и считать, что дифракция будет происходить только в направлениях, перпендикулярных оси Y ; волновой вектор луча, дифрагировавшего в направлении точки P , будет лежать в плоскости, параллельной плоскости XZ , образуя с направлением падающей на щель волны угол φ . Поэтому, как и в случае дифракции Френеля, можно решать одномерную задачу, рассматривая характер дифракционной картины вдоль оси X . Апертурная функция щели зависит только от одной переменной x : $t(x) = \text{rect}(x/a)$. В результате задача определения амплитуды поля в точке $P(x', 0)$, направление на которую определяется углом φ (пространственной частотой $u = \sin \varphi / \lambda$), математически сводится к одномерному преобразованию Фурье от апертурной функции $t(x)$, т.е. к вычислению интеграла (6.6). Следовательно,

$$E_1'(\varphi) = \frac{E_0 e^{ikz}}{\sqrt{i\lambda z}} \int_{-\infty}^{\infty} t(x) e^{-i2\pi ux} dx = \frac{E_0 e^{ikz}}{\sqrt{i\lambda z}} \int_{-a/2}^{a/2} e^{-i2\pi ux} dx =$$

$$= \frac{aE_0}{\sqrt{i\lambda z}} e^{ikz} \frac{\sin(\pi ua)}{\pi ua}.$$

Перепишем это выражение в виде

$$E_1'(\varphi) = E_1(0) \operatorname{sinc}\left(\frac{a \sin \varphi}{\lambda}\right) = E_1(0) \operatorname{sinc}(ua), \quad (6.12)$$

где $E_1(0) = \frac{aE_0}{\sqrt{i\lambda z}} e^{ikz}$ – комплексная амплитуда в направлении распространения падающей на экран волны (при $\varphi = 0$).

Соответствующее этой амплитуде распределение интенсивности будет

$$I_1(u) = I_1(0) \operatorname{sinc}^2(ua), \quad (6.13)$$

где $I_1(0) = |E_1(0)|^2$ – интенсивность волны в направлении $\varphi = 0$.

Учитывая, что $\sin \varphi \approx x'/f$, где x' – координата в плоскости наблюдения, f – фокусное расстояние линзы, $u = \sin \varphi / \lambda = x' / (\lambda f)$, запишем формулы (6.12) и (6.13) в координатной форме:

$$E_1'(x') = E_1(0) \operatorname{sinc}\left(\frac{ax'}{\lambda f}\right), \quad (6.14)$$

$$I_1(x') = I_1(0) \operatorname{sinc}^2\left(\frac{ax'}{\lambda f}\right). \quad (6.15)$$

Такой же результат для $E_1(P)$ получается и при использовании выражения (6.8). Применение этой формулы законно, так как при выбранном начале координат щель симметрична относительно оси X . Поэтому имеем

$$E_1'(\varphi) = \text{const} \cdot E_0 \int_{-\infty}^{\infty} t(x) \cos(2\pi ux) dx =$$

$$= \text{const} \cdot E_0 \int_{-a/2}^{a/2} \cos(2\pi ux) dx = E(0) \text{sinc}(ua) = E(0) \text{sinc}\left(\frac{a \sin \varphi}{\lambda}\right),$$

где $E_1(0) = \text{const} \cdot E_0$.

График зависимости интенсивности $I_1(u)$, определяемой формулой (6.13), приведен на рис. 6.3. Из графика видно, что интен-

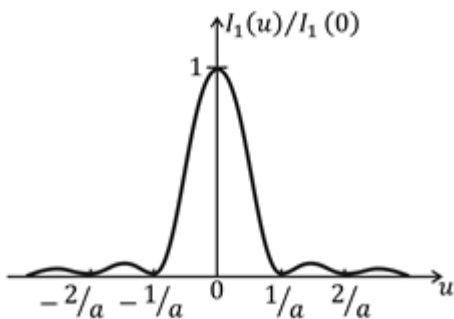


Рис. 6.3

сивность имеет ряд быстро убывающих максимумов, разделенных минимумами, в которых она обращается в нуль. Наибольшего значения, равного $I_1(0)$, интенсивность достигает в центре дифракционной картины, т.е. при $u = \sin \varphi / \lambda = 0$,

а значит, при $\varphi = 0$ (напомним, что $\max \text{sinc}(x) = 1$ при $x = 0$). В этом центральном максимуме сосредоточена основная доля светового потока, выходящего из щели. Минимальные значения, равные нулю, функция $I_1(u)$ имеет, когда ее аргумент $u = \sin \varphi / \lambda$ принимает значения, равные $\pm m/a$, где m – целое число, включая $m = 0$, т.е. в направлениях, определяемых условием

$$a \sin \varphi = m\lambda. \quad (6.16)$$

Из формулы (6.15) видно, что минимум интенсивности достигается на прямых $x' = m\lambda f/a$, параллельных краям

щели. Число $m = \pm 1, \pm 2, \pm 3, \dots$ называется порядком дифракционного минимума⁴. Минимумы первого порядка имеют координаты

$$x_1' = \pm \lambda \pi f / a. \quad (6.17)$$

Таким образом, дифракционная картина при дифракции Фраунгофера на щели представляет собой яркую светлую полосу, являющуюся изображением щели, по обеим сторонам которой симметрично расположены чередующиеся темные и светлые полосы. При этом освещенность боковых полос значительно меньше освещенности центральной полосы и быстро убывает с ростом номера полосы. Эти боковые полосы являются следствием интерференции испытывших дифракцию волн, распространяющихся из разных точек щели в различных направлениях и сведенных линзой в ее задней фокальной плоскости. Полосы с распределением интенсивности в дифракционной картине, описываемой функцией вида (6.13), называют sinc^2 -полосами.

Центральная светлая дифракционная полоса заключена между минимумами первых порядков ($m = \pm 1$), т.е. в пределах углов $\varphi = \pm \varphi_1$, где

$$\sin \varphi_1 = \lambda / a.$$

(Угол $\varphi_1 \approx \lambda / a$ называют дифракционной расходимостью пучка света за отверстием размера a .) Как видно из этого соотношения, ширина центрального максимума тем больше, чем уже щель. При $a \rightarrow \lambda$ центральный максимум расплывается на всю плоскость ($\varphi_1 \rightarrow \pi / 2$). Линейная ширина центрального дифракционного максимума

⁴ Здесь и ниже знаки \pm связаны с тем, что минимумы (или максимумы) всех порядков m , кроме $m = 0$, расположены симметрично относительно прямой $x' = 0$.

$$\Delta x_1' = 2\pi\lambda f / a. \quad (6.18)$$

Эта величина определяет и ширину изображения щели. Заметим, что ширина изображения тем больше, чем уже щель.

Если щель бесконечно узкая, ее апертурная функция аппроксимируется дельта-функцией Дирака $\delta(x)$ и тогда поле в точке P найдется как преобразование Фурье от дельта-функции:

$$E_1'(\varphi) = \text{const } E_0 \int_{-\infty}^{\infty} \delta(x) e^{-i2\pi x \varphi} dx = \text{const } E_0.$$

Мы пришли к тому же результату, что и в случае дифракции Френеля: амплитуда поля, создаваемого бесконечно узкой щелью, имеет одно и то же значение во всех точках плоскости наблюдения (в фокальной плоскости $F'F'$).

В заключение этого параграфа рассмотрим изменение характера дифракции при увеличении расстояния z от щели до плоскости наблюдения. В области приближения тени график интенсивности точно повторяет апертурную функцию щели и имеет вид импульса прямоугольной формы ширины a . Интенсивность света в любой точке оси Z совпадает с интенсивностью падающей волны. В области Френеля интенсивность поля получается как преобразование Френеля, а в области Фраунгофера – как преобразование Фурье от апертурной функции щели $t(x) = \text{rect}(x/a)$.

6.5. Дифракция Фраунгофера на двух параллельных щелях

Пусть теперь объектный экран содержит две одинаковые параллельные друг другу щели, расстояние между серединами которых равно d . Функция пропускания такого экрана показана на рис. 6.4. На экран нормально падает плоская монохроматическая волна. Свяжем с плоскостью, расположенной непосредственно за

экраном, координатную плоскость XU , ось X которой направим перпендикулярно, а ось U – параллельно краям щелей. Ось Z системы координат направим вдоль падающей волны; начало координат поместим посередине расстояния между щелями (схема опыта такая же, как и в случае точечных отверстий; рис. 6.1). В такой системе координат x -координаты середин щелей определяются как $x_{1,2} = \pm (1/2) d$.

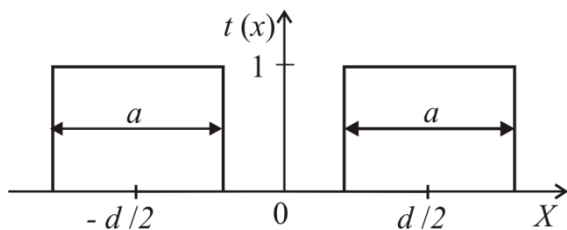


Рис. 6.4

Вследствие дифракции плоской волны на обеих щелях за экраном образуются две цилиндрические волны, которые, складываясь друг с другом в задней фокальной плоскости линзы, расположенной за экраном со щелями, создают систему интерференционных полос. В одном и том же направлении обе щели излучают свет совершенно одинаково. Комплексные амплитуды полей, создаваемых волнами от этих щелей в точке P , направление на которую определяется углом φ , будут одинаковы по модулю, но отличаться друг от друга фазовым множителем $\exp(-i\delta)$, учитывающим опережение колебаний от второй щели по сравнению с первой, так что $E_2 = E_1 \exp(-i\delta)$, где δ – разность фаз колебаний, создаваемых обеими щелями. Величина δ определяется так же, как и в случае двух узких отверстий, т.е.

$$\delta = (2\pi / \lambda) d \sin \varphi = 2\pi u d.$$

Результирующее поле в точке P , в соответствии с принципом суперпозиции, определится как

$$\begin{aligned}
 E'(u) &= E_1(u) + E_2(u) = E_1(u) + E_1(u) \exp(-i2\pi u d) = \\
 &= E_1(u) \exp(-i\pi u d) [\exp(i\pi u d) + \exp(-i\pi u d)] = \\
 &= 2E_1(u) \exp(-i\pi u d) \cos \pi u d.
 \end{aligned}$$

Распределение интенсивности в фокальной плоскости линзы будет описываться выражением

$$I(u) = |E(u)|^2 = 4|E_1(u)|^2 \cos^2 \pi u d, \quad (6.19)$$

или

$$I(u) = 2|E_1(u)|^2 (1 + \cos 2\pi u d).$$

Здесь $E_1(u)$ – огибающая, которая модулирует амплитуду полос, создаваемых одиночной щелью и определяемых выражением (6.12), $E_1(u) = E_1(0)\text{sinc}(ua)$. Сомножитель $|E_1(u)|^2$ в (6.19) описывает явление дифракции на одной щели, сомножитель $\cos^2 \pi u d$ характеризует двулучевую интерференцию, создаваемую лучами от каждой щели. График распределения интенсивности $I(u)$ при дифракции на двух щелях показан на рис. 6.5.

Эти выражения можно получить и с помощью преобразования Фурье. Функция пропускания $t(x)$ рассматриваемого объектного экрана представляет собой два прямоугольника высотой 1 и шириной a ; середины прямоугольников находятся на расстоянии d друг от друга. В выбранной системе координат

$$t(x) = 1 \text{ при } -(d/2 + a/2) \leq x \leq -(d/2 - a/2),$$

$$(d/2 - a/2) \leq x \leq (d/2 + a/2);$$

$$t(x) = 0 \text{ вне указанных интервалов.}$$

Тогда для поля в точке P будем иметь

$$E(u) = \frac{E_0}{\sqrt{i\lambda z}} \left(\int_{-d/2-a/2}^{-d/2+a/2} \exp(-i2\pi ux) dx + \int_{-d/2-a/2}^{d/2+a/2} \exp(-i2\pi ux) dx \right).$$

Множитель $\exp(ikz)$ опущен как не влияющий на распределение поля в плоскости XY' . Произведя интегрирование, после очевидных преобразований придем к выражению

$$E'(u) = \frac{aE_0}{\sqrt{i\lambda z}} \frac{\sin \pi ua}{\pi ua} 2 \cos \pi ud =$$

$$= 2E_1(0) \cdot \text{sinc}(ua) \cos \pi ud = 2E_1(u) \cos \pi ud.$$

Откуда получаем выражение для интенсивности, точно совпадающее с выражением (6.19).

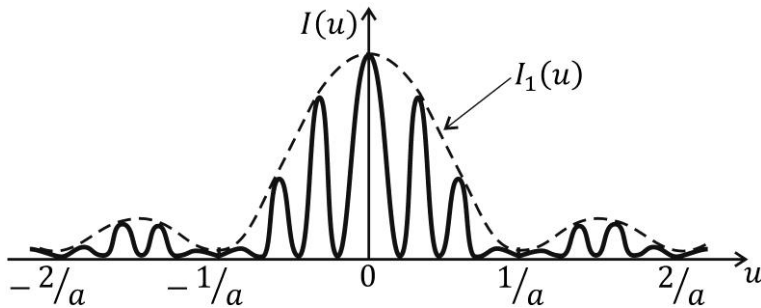


Рис. 6.5

Физические свойства системы полос можно очень просто описать используя свойство смещения преобразования Фурье, согласно которому фурье-преобразование $E(u, x_0)$ функции $E(x \pm x_0)$ равно фурье-преобразованию $E(u)$ функции $E(x)$, умноженному на фазовый множитель $\exp(\pm i2\pi ux_0)$. В этом можно убедиться, если в фурье-преобразовании

$$E(u, x_0) = \int_{-\infty}^{\infty} E(\xi \pm x_0) e^{-i2\pi u \xi} d\xi$$

функции $E(x \pm x_0)$ произвести замену переменной $x = \xi \pm x_0$:

$$\begin{aligned} E(u, x_0) &= \int_{-\infty}^{\infty} E(\xi \pm x_0) e^{-i2\pi u \xi} d\xi = \\ &= \left(\int_{-\infty}^{\infty} E(x) e^{-i2\pi u x} dx \right) e^{\pm i2\pi u x_0} = E(u) e^{\pm i2\pi u x_0}. \end{aligned}$$

Пусть теперь фурье-преобразованием поля $E_1(x)$, созданного одной щелью, будет $E_1(u)$. Тогда фурье-преобразованием поля $E_1(x - d/2)$, созданного нижней щелью, на основании свойства смещения фурье-преобразования будет

$$E_1(u) \exp[-i2\pi u(d/2)] = E_1(u) \exp(-i\pi u d),$$

а поля $E_1(x + d/2)$, созданного верхней щелью, – соответственно

$$E_1(u) \exp[i2\pi u(d/2)] = E_1(u) \exp(i\pi u d).$$

Фурье-преобразование суммарного поля

$$E_1(x + d/2) + E_1(x - d/2),$$

созданного обеими щелями, запишется как

$$E_1(u) [\exp(i\pi u d) + \exp(-i\pi u d)] = 2E_1(u) \cos \pi u d.$$

Отсюда видно, что наблюдаемое распределение интенсивности в фокальной плоскости линзы будет описываться выражением (6.19), полученным другим способом.

Согласно (6.19), интенсивность $I(u)$ достигает максимального значения, равного $4|E_1(u)|^2$, при значениях u , удовлетворяющих условию $2\pi u d = 2m\pi$, т.е. в направлениях, определяемых углами φ , которые удовлетворяют условию

$$d \sin \varphi = m\lambda, \quad (6.20)$$

где $m = 0, \pm 1, \pm 2, \dots$. Эти максимумы называют главными, так как в направлениях, определяемых условием (6.20), действие одной щели усиливается действием другой щели. Минимумы интенсивности, равные нулю, лежат в направлениях, удовлетворяющих условию

$$d \sin \varphi = (m + 1/2)\lambda.$$

Эти минимумы называют добавочными. Главные минимумы наблюдаются в направлениях, в которых лежат минимумы, даваемые одной щелью. Наибольшего значения интенсивность $I(u)$ достигает, когда максимальную интенсивность дает одиночная щель, т.е. в направлении угла $\varphi = 0$ (при $m = 0$). Поэтому имеют значение только те максимумы и минимумы, которые лежат в пределах центрального максимума от одиночной щели. Полосы, расположенные за пределами этого максимума, имеют столь малую освещенность, что оказываются практически незаметными.

Ширина интерференционных полос определяется выражением (6.11): $h = \lambda f / d$. Разделив ширину центрального максимума от одной щели $\Delta x'$, определяемую формулой (6.18), на ширину интерференционных полос h , получим число полос, укладывающихся на ширине центрального максимума:

$$n = \Delta x' / h = (2 \lambda f / a) / (f \lambda / d) = 2 d / a.$$

Отсюда следует, что при переходе от одной щели к двум щелям дифракционные максимумы становятся более узкими. С удалением от центра дифракционной картины интенсивность полос уменьшается, обращаясь в нуль при выполнении условия

$$a \sin \varphi_1 = \pm \lambda.$$

Дифракция на двух щелях применяется в астрономии для измерения угловых диаметров звезд и угловых расстояний между двойными звездами.

6.6. Дифракция Фраунгофера на прямолинейном крае протяженного экрана

Решим теперь задачу о дифракции Фраунгофера на прямолинейном крае протяженного экрана (оптическом ноже). Как и при рассмотрении дифракции Френеля, функцию пропускания экрана $t(x)$ будем представлять в виде единичной функции Хевисайда $1(x)$, т.е. будем считать, что $t(x) = 1(x)$. Ось Y при этом направлена вдоль края экрана, а ось X – перпендикулярно ему. При такой функции пропускания данная задача математически сводится к одномерному преобразованию Фурье функции $1(x)$ и последующему умножению результата на $\text{const} \cdot E_0$, где E_0 – амплитуда плоской волны, падающей на экран.

Поскольку для указанной функции $t(x)$

$$\int_{-\infty}^{\infty} |t(x)| dx \rightarrow \infty,$$

преобразование Фурье в обычной форме

$$F(u) = \int_{-\infty}^{\infty} f(x) \exp(-i2\pi ux) dx$$

не может быть применено непосредственно. Это затруднение можно обойти, выделив в подынтегральном выражении действительную и мнимую части. Тогда получим

$$t(u) = F\{t(x)\} = \int_{-\infty}^{\infty} t(x) \exp(-i2\pi ux) dx =$$

$$= \int_0^{\infty} \exp(-i2\pi ux) dx = \int_0^{\infty} \cos(2\pi ux) dx - i \int_0^{\infty} \sin(2\pi ux) dx.$$

Первый интеграл в правой части

$$\begin{aligned} \int_0^{\infty} \cos(2\pi ux) dx &= \lim_{\varepsilon \rightarrow \infty} \int_0^{\varepsilon/2} \cos(2\pi ux) dx = \\ &= \frac{1}{2} \lim_{\varepsilon \rightarrow \infty} \frac{\sin \varepsilon \pi u}{\pi u} = \frac{1}{2} \delta(u). \end{aligned}$$

Смысл этого результата заключается в том, что в точке $x' = 0$, соответствующей значению $u = x' / (\lambda f) = 0$, вклады всех гармонических составляющих суммируются, создавая бесконечно большой пик, а в точках $x' \neq 0$ вклады этих же составляющих взаимно компенсируются. Второй интеграл легко вычислить, если учесть, что

$$\int_0^{\infty} \sin x dx = 1.$$

Тогда для фурье-преобразования функции пропускания $t(x)$, представляющей единичную функцию Хевисайда, получим

$$t(u) = F\{1(x)\} = \frac{1}{2} \delta(u) - i \frac{1}{2\pi u} = \frac{1}{2} \delta(u) + \frac{1}{i2\pi u}. \quad (6.21)$$

Наличие дельта-функции в спектре единичной функции соответствует постоянной составляющей, а второе слагаемое обусловлено тем, что при $x < 0$ функция $t(x)$ обращается в нуль.

С учетом сказанного для комплексной амплитуды дифрагировавшего поля в области $u = x' / (\lambda f) < 0$, т.е. $x' < 0$, где $\delta(u) = 0$, будем иметь

$$E'(x', 0) = \text{const} \cdot E_0 \frac{\lambda f}{i2\pi x'} = \text{const} \cdot E_0 \frac{\lambda f}{2\pi x'} \exp\left(-i\frac{\pi}{2}\right).$$

Распределение интенсивности в указанной области

$$I(x', 0) = C \left(\frac{\lambda f}{2\pi x'} \right)^2,$$

где C – постоянная.

Полученное выражение показывает, что при дифракции Фраунгофера на крае протяженного экрана, как и при дифракции Френеля, резкой границы между светом и тенью нет. В области $x' < 0$, где по законам геометрической оптики должна наблюдаться тень, интенсивность убывает обратно пропорционально квадрату расстояния от края экрана. В области же $x' > 0$ затухающие осцилляции интенсивности, как это имеет место при дифракции Френеля, отсутствуют.

Указанное выше затруднение, связанное с расходимостью интеграла при непосредственном применении преобразования Фурье, можно обойти и если вместо $t(x)$ в виде единичной функции Хевисайда использовать функцию $t_1(x) = 1(x) \exp(-\alpha x)$, где α – положительное число, стремящееся к нулю, и в конечном выражении перейти к пределу $\alpha \rightarrow 0$. Так как

$$t_1(x) = \begin{cases} 0, & x < 0, \\ \exp(-\alpha x), & x > 0, \end{cases}$$

то

$$\begin{aligned} E'(x') &= \text{const} E_0 \lim_{\alpha \rightarrow 0} \int_0^{\infty} \exp[-(\alpha + i2\pi u)x] dx = \\ &= \text{const} E_0 \lim_{\alpha \rightarrow 0} \frac{1}{\alpha + i2\pi u} = \text{const} E_0 \frac{1}{i2\pi u}. \end{aligned}$$

От полученного ранее выражения оно отличается отсутствием члена с дельта-функцией. Однако это не имеет здесь принципиального значения.

Как и в случае дифракции Френеля, распределение поля в картине при дифракции Фраунгофера на щели можно рассматривать как дифракцию на двух краях протяженных экранов, сближенных на расстояние, равное ширине щели a . (Такая возможность существует в связи с принципом суперпозиции, согласно которому спектр суммы или разности функций равен сумме или разности соответствующих этим функциям спектров.) Функцию пропускания щели, как было показано в п. 5.9, можно представить в виде разности двух единичных функций Хевисайда, начало одной из которых находится в точке $x = -a/2$, а другой – в точке $x = a/2$. Воспользовавшись уже использованным выше свойством смещения преобразования Фурье, согласно которому фурье-спектр функций $f(x \pm d)$ равен фурье-спектру функции $f(x)$, умноженному на фазовые множители $\exp(\pm i2\pi ud)$, а также свойством дельта-функции, согласно которому

$$f(x)\delta(x) = f(0)\delta(x),$$

получим

$$\begin{aligned} 1(u, a/2) &= 1(u) \exp[i2\pi u(a/2)] = \\ &= \frac{\exp(i\pi ua)}{i2\pi u} + \frac{1}{2} \exp(i\pi ua) \delta(u) = \frac{\exp(i\pi ua)}{i2\pi u} + \frac{1}{2} \delta(u). \end{aligned}$$

Аналогично находим

$$1(u, -a/2) = \frac{\exp(-i\pi ua)}{i2\pi u} + \frac{1}{2} \delta(u).$$

С учетом принципа суперпозиции получаем фурье-спектр функции пропускания щели

$$F\{t(x)\} = 1(u, a/2) - 1(u, -a/2) = \\ = \frac{1}{i2\pi u} [\exp(i\pi u a) - \exp(-i\pi u a)] = a \cdot \frac{\sin \pi u a}{\pi u a}.$$

Умножив это выражение на $\text{const} \cdot E_0$, получим искомое распределение поля в дифракционной картине при дифракции Фраунгофера на щели.

6.7. Дифракция Фраунгофера на прямоугольном отверстии

Рассмотрим дифракцию Фраунгофера на двумерных структурах. Простейшей такой структурой является прямоугольное отверстие в экране, когда стороны a и b прямоугольника не существенно отличаются друг от друга (могут быть и равными). Схема и условия опыта такие же, как и при дифракции на щели, с той лишь разницей, что в этом опыте вместо щели стоит экран с прямоугольным отверстием. Ось X системы координат XU направляется вдоль размера a , Y – вдоль размера b . Начало координат помещено в центр отверстия. Теперь, когда размеры a и b соизмеримы, следует учесть дифракцию и на размере b . Волновой вектор дифрагировавшей волны в данном случае уже не будет перпендикулярен оси Y . Его направление будет определяться направляющими углами α и β (рис. 6.6). При этом, как видно из рисунка, направляющие косинусы

$$\cos \alpha = \sin \varphi = x' / r = \lambda u, \quad \cos \beta = \sin \theta = y' / r = \lambda v,$$

где φ и θ – углы, дополняющие углы α и β до $\pi/2$, r – длина отрезка OP , соединяющего центр отверстия O с точкой наблюдения P . Задача становится двумерной. Поле в любой точке $P(x', y')$, в которой собираются дифрагировавшие волны, направление распространения которых определяется

пространственными частотами u и v , можно найти с помощью двумерного преобразования Фурье (6.5).

При выбранном начале координат O края отверстия будут иметь координаты $x_{1,2} = \pm a / 2$, $y_{1,2} = \pm b / 2$. Функция пропускания прямоугольного отверстия в выбранной системе координат будет иметь вид

$$t(x, y) = \begin{cases} 1, & |x| \leq a/2, |y| \leq b/2; \\ 0, & |x| > a/2, |y| > b/2. \end{cases}$$

(Очевидно, что $t(x, y) = \text{rect}(x/a) \cdot \text{rect}(y/b)$). При такой апертурной функции переменные x и y в интеграле (6.5) разделяются. С учетом этого находим

$$\begin{aligned} E'(u, v) &= \text{const} \cdot E_0 \int_{-a/2}^{a/2} e^{-i2\pi ux} dx \int_{-b/2}^{b/2} e^{-i2\pi vy} dy = \\ &= E(0) \cdot \left(\frac{\sin \pi ua}{\pi ua} \right) \cdot \left(\frac{\sin \pi vb}{\pi vb} \right) = E(0) \cdot \text{sinc}(ua) \text{sinc}(vb), \quad (6.22) \end{aligned}$$

где $E(0) = \text{const} \cdot abE_0$ – амплитуда поля, соответствующая значениям $u = 0$, $v = 0$, т.е. в направлении падающей на отверстие волны.

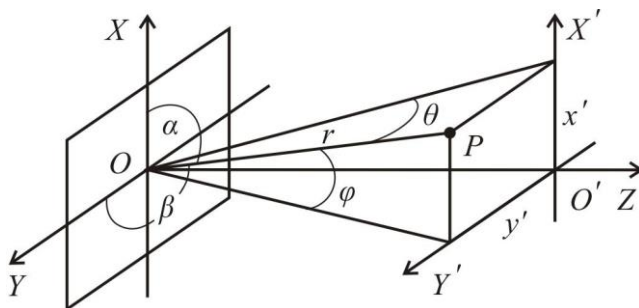


Рис. 6.6

К такому же результату мы придем и при использовании соотношения (6.7). Возможность применения этой формулы обусловлена тем, что при выбранном начале координат прямоугольное отверстие симметрично относительно осей X и Y . Так как переменные x и y разделяются, то

$$\begin{aligned}
 E'(u, v) &= \text{const} \cdot E_0 \int_{-a/2}^{a/2} \cos(2\pi ux) dx \int_{-b/2}^{b/2} \cos(2\pi vy) dy = \\
 &= E(0) \cdot \left(\frac{\sin \pi ua}{\pi ua} \right) \cdot \left(\frac{\sin \pi vb}{\pi vb} \right) = E(0) \text{sinc}(ua) \text{sinc}(vb).
 \end{aligned}$$

Подставляя в (6.21) выражения пространственных частот u и v , получим распределение амплитуды поля по экрану

$$E'(x', y') = E(0) \text{sinc} \left(\frac{ax'}{\lambda f} \right) \text{sinc} \left(\frac{by'}{\lambda f} \right) \quad (6.23)$$

и распределение интенсивности

$$I(x', y') = I(0) \text{sinc}^2 \left(\frac{ax'}{\lambda f} \right) \text{sinc}^2 \left(\frac{by'}{\lambda f} \right). \quad (6.24)$$

Величины $E(0)$ и $I(0)$ – представляют собой соответственно амплитуду и интенсивность поля в центре дифракционной картины. График распределения интенсивности $I(x', y')$ показан на рис. 6.7, а дифракционная картина – на рис. 6.8.

Выражение (6.24) показывает, что в дифракционной картине от прямоугольного отверстия распределение интенсивности пропорционально произведению распределений интенсивностей от взаимно перпендикулярных щелей. Такую картину, как видно из

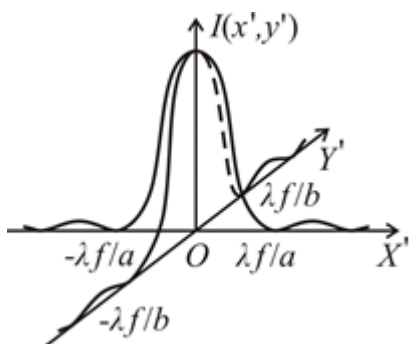


Рис. 6.7

перпендикулярных прямых $x' = m_1 \lambda f / a$ и $y' = m_2 \lambda f / b$, где $m_1, m_2 = \pm 1, \pm 2, \pm 3, \dots$. Таким образом, дифракционная картина содержит два ряда взаимно перпендикулярных линий, на которых интенсивность равна нулю. Темные линии находятся на постоянном расстоянии друг от друга, и только в центре дифракционной картины это расстояние вдвое больше. Области максимальной интенсивности находятся между этими линиями. Поэтому дифракционная картина представляет собой ряд светлых прямоугольников, пересеченных темными линиями. Она производит впечатление креста, который становится светлее к центру и переключины которого лежат параллельно сторонам дифракционного отверстия. Как и в случае щели, основная часть светового потока будет приходиться на область, занимаемую центральным максимумом. Эта область представляет собой прямоугольник и является изображением диафрагмы. Центр этого прямоугольника имеет координаты $x' = 0, y' = 0$, где обе sinc^2 -функции достигают своего наибольшего значения, равного единице. Центры других, менее освещенных прямоугольников, имеют координаты

рис. 6.8, можно получить, налагая друг на друга две взаимно перпендикулярные картины, одна из которых получается при дифракции на щели шириной a , а другая – на щели шириной b .

Интенсивность равна нулю всюду, где величина $ax' / (\lambda f)$ или $by' / (\lambda f)$ есть целое число, отличное от нуля, т.е. на взаимно

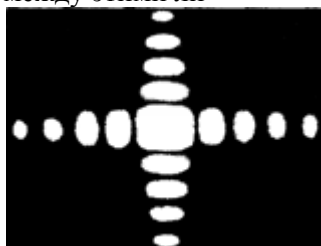


Рис. 6.8

$$x' = (m_1 + 1/2)\lambda f / a, \quad y' = (m_2 + 1/2)\lambda f / b,$$

где m_1, m_2 – целые положительные и отрицательные числа. Размеры центральной области можно определить как удвоенные расстояния между началом координат и первыми нулями функции $I(x', y')$. Имеем: $ax_0' / \lambda f = 1$, $by_0' / \lambda f = 1$. Откуда получаем

$$2x_0' = 2\lambda f / a, \quad 2y_0' = 2\lambda f / b.$$

Как видим из этих соотношений, размеры изображения обратно пропорциональны размерам дифракционного отверстия (диафрагмы). Следовательно, чем больше отверстие в экране, тем меньше его дифракционное изображение. При этом длинная сторона прямоугольника изображения соответствует короткой стороне диафрагмы, а короткая сторона – длинной стороне диафрагмы, так как дифракция тем заметнее, чем уже освещаемый участок. Изображение оказывается повернутым в плоскости диафрагмы на угол 90° .

Перепишем выражение (6.22) в виде

$$E'(u, v) = \text{const } E_0 a \text{ sinc}(ua) \frac{\sin \pi v b}{\pi v} \quad (6.25)$$

и перейдем к пределу при $b \rightarrow \infty$ в множителе $\frac{\sin \pi v b}{\pi v}$.

Получим δ -функцию $\delta(v)$, которая равна нулю при любом значении v , кроме значения $v = 0$. С учетом этого выражение (6.25) принимает вид

$$E'(u, 0) = \text{const } E_0 a \text{ sinc}(ua) = E_1(0) \text{ sinc}(ua).$$

Это выражение описывает распределение поля при дифракции на щели. Следовательно, при $b \rightarrow \infty$ формула (6.21), как и должно быть, переходит в формулу (6.12).

Для апертуры квадратной формы распределения комплексной амплитуды и интенсивности в дифракционной картине можно получить из выражений (6.23) и (6.24), положив в них $b = a$. Будем иметь

$$E'(x', y') = E(0) \operatorname{sinc}\left(\frac{ax'}{\lambda f}\right) \cdot \operatorname{sinc}\left(\frac{ay'}{\lambda f}\right)$$

– для распределения комплексной амплитуды и

$$I(x', y') = I(0) \operatorname{sinc}^2\left(\frac{ax'}{\lambda f}\right) \cdot \operatorname{sinc}^2\left(\frac{ay'}{\lambda f}\right)$$

– для распределения интенсивности. Изображение квадратного отверстия тоже будет иметь форму квадрата со стороной

$$2x_0' = 2y_0' = 2\lambda f / a.$$

6.8. Дифракция Фраунгофера на круглом отверстии

Другим примером дифракции на двумерной структуре является дифракция на круглом отверстии. Схема опыта и в этом случае такая же, как на рис. 6.2, только теперь вместо экрана со щелью стоит экран с круглым отверстием. Как и при рассмотрении френелевской дифракции на круглом отверстии, перейдем к полярным координатам ρ и θ в плоскости $X'Y'$, расположенной непосредственно за отверстием. Начало координат O расположим в центре отверстия. Дифракционная картина наблюдается в плоскости $X'Y'$. Апертурная функция отверстия в полярных координатах $t(\rho) = \operatorname{circ}(\rho/R)$, где $\rho = \sqrt{x'^2 + y'^2}$ – расстояние от центра отверстия до его произвольной точки, R – радиус

отверстия. Направление вектора \mathbf{k} дифрагировавшей волны на точку P будем определять углом φ , который образует этот вектор с осью Z , и азимутальным углом ψ – углом между осью X и проекцией k_{\perp} вектора \mathbf{k} на плоскость XY (рис. 6.9). При этом

$$k_{\perp} = \sqrt{k_x^2 + k_y^2} = 2\pi\sqrt{u^2 + v^2}.$$

Тогда

$$k_x = k_{\perp}\cos\psi, \quad k_y = k_{\perp}\sin\psi, \quad k_{\perp} = k \sin\varphi.$$

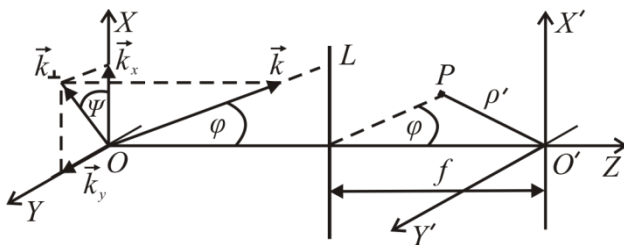


Рис. 6.9

Как и в предыдущем случае, задача решается с помощью двумерного преобразования Фурье. Однако здесь удобнее перейти от пространственных частот u и v к циклическим пространственным частотам k_x и k_y . В этом случае интеграл (6.5) принимает вид

$$\begin{aligned} E'(x', y') &= \\ &= \text{const} \cdot E_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y) \exp[-i(k_x x + k_y y)] dx dy. \end{aligned} \quad (6.26)$$

Выражая в (6.26) переменные x , y , k_x , k_y через введенные переменные r , θ , φ , ψ , получим

$$x = \rho \cos\theta, \quad y = \rho \sin\theta, \quad dx dy = \rho dr d\theta,$$

$$k_x x + k_y y = \rho k_{\perp} (\cos \theta \cos \psi + \sin \theta \sin \psi) = \rho k_{\perp} \cos(\theta - \psi).$$

Распределение амплитуды в плоскости наблюдения XY'

$$E'(\varphi) = \text{const} \cdot E_0 \int_0^R \rho d\rho \int_0^{2\pi} \exp[-i\rho k_{\perp} \cos(\theta - \psi)] d\theta. \quad (6.27)$$

Интегрирование в (6.27) проводится по площади круга радиуса R .

Используя интегральное представление функции Бесселя первого рода нулевого порядка

$$J_0(z) = \frac{1}{2\pi} \int_0^{2\pi} \exp[-i z \cos(\theta - \psi)] d\theta,$$

получим

$$E(\varphi) = \text{const} \cdot 2\pi E_0 \int_0^R J_0(\rho k_{\perp}) \rho d\rho. \quad (6.28)$$

Положив в (6.27) $\rho k_{\perp} = \xi$, $d\rho = d\xi / k_{\perp}$, будем иметь

$$E(\varphi) \sim \frac{2\pi}{k_{\perp}^2} \int_0^{Rk_{\perp}} \xi J_0(\xi) d\xi,$$

или

$$E(\varphi) \sim \frac{2\pi}{k_{\perp}^2} \cdot Rk_{\perp} \cdot J_1(Rk_{\perp}),$$

где учтено, что

$$\int_0^z \xi J_0(\xi) d\xi = z J_1(z).$$

Функция $J_1(z)$ называется функцией Бесселя первого рода первого порядка.

Функции Бесселя, или, как их еще называют, цилиндрические функции, принадлежат к классу так называемых специальных функций. Эти функции нельзя представить аналитически в виде конечного алгебраического выражения. Их можно представить в виде рядов. Функция $J_0(z)$ может быть представлена следующим быстро сходящимся знакопеременным рядом

$$\begin{aligned} J_0(z) &= 1 - \frac{z^2}{2^2 (1!)^2} + \frac{z^4}{2^4 (2!)^2} - \frac{z^6}{2^6 (3!)^2} + \dots = \\ &= 1 + \sum_{m=1}^{\infty} (-1)^m \frac{z^{2m}}{2^{2m} (m!)^2}, \end{aligned}$$

а функция $J_1(z)$ – рядом

$$\begin{aligned} J_1(z) &= \frac{z}{2} - \frac{z^3}{2^2 \cdot 4} + \frac{z^5}{2^2 \cdot 4^2 \cdot 6} - \frac{z^7}{2^2 \cdot 4^2 \cdot 6^2 \cdot 8} + \dots = \\ &= \frac{z}{2} + \sum_{n=1}^{\infty} (-1)^n \frac{z^{2n+1}}{\left(\prod_{m=1}^n (2m)^2 \right) [2(n+1)]}. \end{aligned}$$

При больших значениях z имеют место асимптотические формулы

$$J_0(z) = \frac{\cos(z - \pi/4)}{\sqrt{\pi z/2}}, \quad J_1(z) = \frac{\sin(z - \pi/4)}{\sqrt{\pi z/2}}.$$

Функции Бесселя часто встречаются в различных задачах физики, астрономии и других наук. Численные значения этих функций протабулированы.

Проведя указанное интегрирование, придем к выражению амплитуды поля в произвольной точке P :

$$E'(\varphi) = \text{const} \cdot \pi R^2 \frac{2J_1(\alpha)}{\alpha}, \quad (6.29)$$

где $\alpha = R k_{\perp} = kR \sin \varphi = (2\pi / \lambda) R \sin \varphi \approx 2\pi R \varphi / \lambda$, так как рассматриваются малые углы φ (поэтому $\sin \varphi \approx \varphi$), const – другая постоянная, включающая в себя множитель E_0 .

Распределение интенсивности в дифракционной картине будет определяться выражением

$$I(\alpha) = I(0) \left(\frac{2J_1(\alpha)}{\alpha} \right)^2. \quad (6.30)$$

Заметим, что поскольку в выражение для α входит произведение

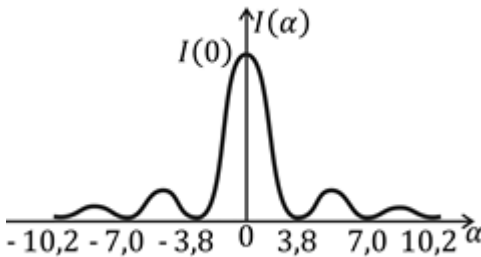


Рис. 6.10

$2\pi R$, то, в отличие от дифракции Френеля, где интенсивность зависит от площади отверстия (πR^2), при дифракции Фраунгофера интенсивность зависит от длины окружности

($2\pi R$), ограничивающей отверстие. График зависимости $I(\alpha)$ показан на рис. 6.10. Эта кривая очень похожа на кривую sinc^2 -функции. Однако, в отличие от последней, нули функции $I(\alpha)$ имеют место при несколько больших значениях аргумента. При значении аргумента, равном нулю, обе функции достигают наибольшего значения, равного единице (при этом $J_1(x)/x = 1/2$ при $x = 0$).

Как и в случае щели, интенсивность $I(\alpha)$ имеет максимальное значение, равное $I(0)$, при $\alpha = 0$, т.е. в направлении распространения падающей волны ($\varphi = 0$), и с ростом α , осциллируя, быстро затухает. Дифракционная картина

в плоскости $X'Y'$ представляет собой яркое пятно, в котором освещенность быстро убывает от центра к периферии, и ряд светлых концентрических колец, разделенных темными промежутками, в которых освещенность падает до нуля. С удалением от центра картины интенсивность светлых колец резко падает, так что практически можно заметить только одно или два кольца. Расчет показывает, что 83,8% всего светового потока приходится на центральное светлое пятно в дифракционной картине, на первое светлое кольцо – 7,2%, на второе светлое кольцо – 2,8%, на третье – 1,4% и 4,8% – на все остальные кольца (рис. 6.11). Это пятно называют кружком (или диском) Эри, который впервые исследовал эту задачу в общем виде. Центр диска Эри расположен в точке O' плоскости наблюдения.

Первый минимум (равенство нулю) интенсивности, как видно на рис. 6.8, имеет место при $\alpha = 3,83$, т.е. при $2\pi R\varphi / \lambda = 3,83$. Откуда находим направление на этот минимум:

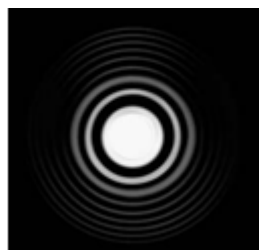


Рис. 6.11

$$\varphi_1 = 0,61\lambda / R = 1,22\lambda / D, \quad (6.31)$$

где D – диаметр отверстия. Этот минимум определяет границу центрального светового пятна (т.е. угловой размер диска Эри) в дифракционной картине. Заметим, что размер (угловой) центрального светлого пятна, как и при дифракции на прямоугольном отверстии, обратно пропорционален размеру (диаметру) отверстия.

Радиальное распределение интенсивности можно получить, если принять, что $\varphi \approx \rho' / f$, где $\rho' = \sqrt{x'^2 + y'^2}$ – расстояние от точки наблюдения $P(x', y')$ до точки O' , f – фокусное расстояние линзы. Тогда $\alpha = (2\pi R / \lambda)\varphi = [2\pi R / (\lambda f)]\rho'$, а распределение интенсивности

$$I(\rho') = I(0) \left(\frac{2J_1([2\pi R/(\lambda f)]\rho')}{[2\pi R/(\lambda f)]\rho'} \right)^2.$$

Используя соотношение (6.30), найдем линейный размер (радиус) кружка Эри:

$$\rho'_1 \approx f\varphi_1 = 0,61\lambda f/R. \quad (6.32)$$

Заметим, что радиус кружка Эри тем меньше, чем больше радиус диафрагмы (отверстия). В области пространственных частот радиусу кружка Эри соответствует полоса пространственных частот шириной

$$\eta = \sqrt{u^2 + v^2} = \frac{1}{\lambda f} \sqrt{x'^2 + y'^2} = \frac{\rho'_1}{\lambda f} = 0,61/R,$$

так как $u = x'/(\lambda f)$ $v = y'/(\lambda f)$.

Следует отметить, что рассмотрение дифракции Фраунгофера на прямоугольном и круглом отверстиях имеет большое практическое значение. В дифракционной теории изображающих оптических систем формулы (6.31) и (6.32) играют первостепенную роль. Дело в том, что в оптических приборах оправы линз и объективов, а также диафрагмы имеют круглую форму. В оптических приборах формирования изображения диафрагма служит для регулирования способности собирать падающий на оптическую систему свет, т.е. для пространственного ограничения падающего на них пучка света. Эта так называемая апертурная диафрагма обычно помещается между различными линзовыми элементами. Применяются и прямоугольные диафрагмы. При прохождении света через такие диафрагмы и через оправы линз из-за ограничения фронта падающей волны возникает более или менее заметная дифракция. В фокальной плоскости линзы или в плоскости, где возникает изображение, образуется дифракционная картина. Если на линзу падает плоская волна, возникает дифракция Фраунгофера. В

плоскости наблюдения получится картина, показанная на рис. 6.11. Формулы (6.31) и (6.32) будут применимы и к линзе, в этом случае D – диаметр оправы линзы. Световые лучи, исходящие из точечного источника света, находящегося на достаточно большом удалении от линзы, при попадании в линзу можно считать почти параллельными. Условия френгоферовской дифракции будут выполнены и в этом случае. В плоскости изображения возникает дифракционная картина в виде диска Эри, окруженного чередующимися темными и светлыми кольцами. Эта картина и будет представлять собой изображение бесконечно удаленного точечного источника света, размытое дифракцией на апертуре линзы.

6.9. Применение метода подобия к изучению дифракции Фраунгофера

Метод подобия позволяет без новых вычислений получить распределение амплитуды светового поля в френгоферовской дифракционной картине от отверстия по известной дифракционной картине от другого отверстия, связанного с первым преобразованием сжатия или растяжения.

Будем исходить из интегрального соотношения (6.4), которое запишем в виде

$$E(\mathbf{k}) = C \iint_{\sigma} E(x, y) e^{-i2\pi(ux+vy)} dx dy, \quad (6.33)$$

где интегрирование производится не по всей плоскости XU , а по области σ , занимаемой отверстием, C – постоянная. Вектор $\mathbf{k}(2\pi u, 2\pi v)$ определяет направление на точку наблюдения P . Перейдем к новым координатам

$$x' = \mu x, \quad y' = \lambda y, \quad (6.34)$$

где μ и λ – постоянные. Значения этих постоянных, большие единицы, соответствуют преобразованиям растяжения, а меньшие

единицы – преобразованиям сжатия. Преобразования (6.34) называются преобразованиями подобия. Область σ плоскости XU при этом преобразуется в область σ' плоскости XU' . Эта область получается из области σ равномерным растяжением или сжатием (в зависимости от того, больше или меньше единицы коэффициенты μ и λ) вдоль осей X и U . Интеграл (6.33) при этом преобразуется в интеграл

$$E(\mathbf{k}') = \frac{C}{\mu\lambda} \iint_{\sigma} E(x', y') e^{-i2\pi(u'x' + v'y')} dx' dy', \quad (6.35)$$

где пространственные частоты

$$u' = u / \mu, v' = v / \lambda. \quad (6.36)$$

Но интеграл (6.35) представляет собой амплитуду светового поля $E(u', v')$ в штрихованной системе координат в направлении вектора $\mathbf{k}'(2\pi u', 2\pi v')$ при френгоферовской дифракции на отверстии σ' . Следовательно, $E(\mathbf{k}') = \mu\lambda E(\mathbf{k})$. Это соотношение и устанавливает связь между распределением амплитуды светового поля в френгоферовских дифракционных картинах от отверстий σ и σ' , получаемых одно из другого равномерным сжатием или растяжением в направлении, определяемом соотношением параметров μ и λ , т.е. связанных между собой преобразованием подобия.

Масштаб дифракционной картины вдоль координатных осей X и U определяется составляющими $2\pi u'$ и $2\pi v'$ вектора \mathbf{k}' . Поэтому, как видно из соотношений (6.34) и (6.36), при равномерном растяжении или сжатии отверстия в каком-либо направлении дифракционная картина сжимается или растягивается в том же направлении. Например, при растяжении или сжатии круглого отверстия в каком-либо направлении оно переходит в эллиптическое. Дифракционные кольца при этом сжимаются или растягиваются, также принимая эллиптическую форму.

В заключение этого параграфа отметим важное свойство дополнительных экранов. Дополнительными экранами называют два экрана с такими апертурами, что при их наложении друг на друга получается полностью непрозрачный экран. Свойство таких экранов определяется теоремой Бабине, согласно которой дифракционные картины, создаваемые такими экранами, совпадают. В частности, например, дифракция на непрозрачном диске или шарике такая же, как и рассмотренная выше дифракция на круглом отверстии такого же радиуса в непрозрачном экране. В обоих случаях интенсивность света описывается выражением (6.30), в котором R – радиус отверстия или диска (или шарика).

6.10. Пространственная когерентность и дифракция Фраунгофера

В п. 4.5 было показано, что степень пространственной когерентности $|\gamma_{12}(0)|$ световых колебаний в точках S_1 и S_2 , расположенных в поперечном сечении пучка, и функция видности интерференционных полос V в случае протяженного источника в виде равномерно светящейся полосы определяются как

$$|\gamma_{12}(0)| = V(a) = \left| \operatorname{sinc} \left(\frac{ad}{\lambda l_0} \right) \right|, \quad (6.37)$$

где a – ширина полосы, d – расстояние между точками S_1 и S_2 , l_0 – расстояние от источника до точек S_1 и S_2 , λ – длина волны света, излучаемого источником. Найдем теперь выражение для модуля амплитуды светового поля в точке S_1 фраунгоферовской дифракционной картины, создаваемой щелью, имеющей ширину полосы a и расположенной на том же расстоянии l_0 от точек S_1 и S_2 (рис. 4.6), что и полоска, и сравним его с соотношением (6.37). Это выражение определяется выражением (6.14). Положим в нем $x' = d$ и заменим f на l_0 (так как точки S_1 и S_2 расположены не в

фокальной плоскости линзы, а на расстоянии l_0 от щели). Тогда для модуля амплитуды поля в точке S_1 при дифракции на щели будем иметь

$$|E_1'(S_1)| = \left| E_1(0) \cdot \text{sinc}\left(\frac{ad}{\lambda l_0}\right) \right|.$$

Сравнивая это соотношение с (4.21), видим, что

$$|\gamma_{12}(0)| = V(a) \sim |E_1'(S_1)|.$$

Аналогичные соотношения между $|\gamma_{12}(0)|$ и $|E'(S_1)|$ имеют место для протяженных источников любой формы. Это непосредственно вытекает из рассмотренной в главе 4 теоремы Ван-Циттера – Цернике.

Согласно Ван-Циттеру и Цернике, комплексная степень пространственной когерентности колебаний светового поля в точках S_1 и S_2 , освещенными протяженным некогерентным источником, с точностью до несущественного фазового множителя $\exp(i\psi_{12})$ является нормированным фурье-образом распределения интенсивности $I_S(\xi, \eta)$ источника:

$$\gamma_{12}(0) = e^{i\psi_{12}} \frac{\iint I_S(\xi, \eta) \exp[-i2\pi(u\xi + v\eta)] d\xi d\eta}{\iint_{\sigma} I_S(\xi, \eta) d\xi d\eta}.$$

Интегрирование производится по площади σ источника S . Если интенсивность распределена по источнику равномерно, т.е. для равномерно светящегося источника, когда $I_S(\xi, \eta) = \text{const}$, интеграл в знаменателе будет равен $\text{const} \cdot \sigma$, и тогда для степени пространственной когерентности в точках S_1 и S_2 получим

$$|\gamma_{12}(0)| = \frac{1}{\sigma} \left| \iint_{\sigma} \exp[-i2\pi(u\xi + v\eta)] d\xi d\eta \right|. \quad (6.38)$$

Если отверстие в непрозрачном экране, имеющее форму источника, и апертурную функцию, равную единице в пределах отверстия и нулю за ее пределами, осветить плоской волной единичной амплитуды, то интеграл в (6.38) с точностью до несущественного постоянного множителя будет, очевидно, определять комплексную амплитуду поля в точке $S_1(x', y')$ фраунгоферовской дифракционной картины, центр которой расположен в точке $S_2(0, 0)$. На основании этого Ван-Циттер и Цернике сформулировали следующее утверждение: степень пространственной когерентности световых колебаний, возбуждаемых протяженным источником в точках S_1 и S_2 , расположенных в поперечном сечении пучка, пропорциональна модулю амплитуды напряженности поля в точке S_1 фраунгоферовской дифракционной картины с центром в точке S_2 , создаваемой отверстием, име-

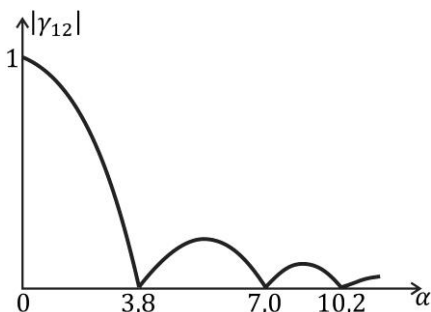


Рис. 6.12

ющим форму и размеры источника. Как видим, теорема Ван-Циттера – Цернике сводит вычисление степени пространственной когерентности $|\gamma_{12}(0)|$ к соответствующей задаче фраунгоферовской дифракции.

Применим теорему Ван-Циттера – Цернике к определению степени пространственной когерентности $|\gamma_{12}(0)|$ и видности интерференционных полос V в случае протяженных источников в виде равномерно светящихся круга и прямоугольника. Используя результаты п. 6.8, находим для светящегося круга

$$|\gamma_{12}(0)| = V(0) \sim \left| \frac{2J_1(\alpha)}{\alpha} \right|,$$

где при малых φ переменная $\alpha = (2\pi/\lambda)R\sin\varphi \approx \pi Dd/(\lambda l_0)$, D – диаметр круга. График функции $|\gamma_{12}(0)|$ от переменной α показан на рис. 6.12. Точно такой же вид имеет и график зависимости $V(\alpha)$.

Из графика на рис. 6.12 видно, что при $\alpha = 3,83$ степень пространственной когерентности и функция видности обращаются в нуль. Отсюда следует, что если рассматриваемые точки S_1 и S_2 находятся на расстоянии $d = 3,83\lambda l_0 / (\pi D) = 1,22\lambda/\theta$ друг от друга, то колебания светового поля в этих точках будут полностью не когерентны. Здесь $\theta = D/l_0$ – угол, под которым диаметр источника виден из точки S_1 или S_2 . Величину $d = 1,22\lambda/\theta$ принимают в качестве радиуса когерентности для светящегося круга, угловой размер которого равен θ

$$\rho_{\text{ког}} = 1,22\lambda/\theta.$$

Таким образом, внутри круга диаметра, равного

$$\rho_{\text{ког}} = 1,22\lambda/\theta = 1,22\lambda l_0/D,$$

обеспечивается пространственная когерентность. Задаваясь допустимым уменьшением функции $|\gamma_{12}(0)|$, можно выбрать диаметр источника света, который должен быть меньше

$$D = 1,22\lambda l_0/\rho_{\text{ког}}.$$

В случае источника в виде светящегося прямоугольника с учетом изложенного в п. 6.7 получим

$$|\gamma_{12}(0)| = V(0) \sim |\text{sinc}(ua) \cdot \text{sinc}(vb)|. \quad (6.39)$$

При $ua = 1$ и $vb = 1$ величины $|\gamma_{12}(0)|$ и V обращаются в нуль. Учитывая, что $u = \cos\alpha/\lambda$ и $v = \cos\beta/\lambda$, где $\cos\alpha$ и $\cos\beta$ – направляющие косинусы вектора \mathbf{k} волны, дифрагировавшей в направлении точки S_1 , эти равенства приводим к виду $\cos\alpha =$

$= \lambda / a$, $\cos\beta = \lambda / b$. Поскольку сумма квадратов направляющих косинусов равна единице, то ввиду малости угла γ (γ – угол между вектором \mathbf{k} и осью Z), будем иметь

$$d^2 / l_0^2 = \operatorname{tg}^2 \gamma \approx \sin^2 \gamma = 1 - \cos^2 \gamma = \cos^2 \alpha + \cos^2 \beta.$$

Подставив сюда значения $\cos\alpha$ и $\cos\beta$, получим

$$d^2 / l_0^2 = \lambda^2 / a^2 + \lambda^2 / b^2.$$

Отсюда находим расстояние d между точками S_1 и S_2 , при котором колебания светового поля в этих точках будут полностью не когерентны: $d = \lambda l_0 \sqrt{a^{-2} + b^{-2}}$. Это значение d можно считать радиусом когерентности источника в виде светящегося прямоугольника, длины сторон которого равны a и b :

$$\rho_{\text{кор}} = \lambda l_0 \sqrt{a^{-2} + b^{-2}}. \quad (6.40)$$

Пространственная когерентность обеспечивается внутри прямоугольника со сторонами $2\lambda l_0 / a$, $2\lambda l_0 / b$.

Если источник имеет форму квадрата со стороной a , то его степень и радиус пространственной когерентности можно найти, положив в формулах (6.39) и (6.40) $a = b$. Тогда получим соответственно:

$$\begin{aligned} |\gamma_{12}(0)| &= V(0) \sim |\operatorname{sinc}(ua) \cdot \operatorname{sinc}(va)|, \\ \rho_{\text{кор}} &= \sqrt{2} \lambda l_0 / a. \end{aligned} \quad (6.41)$$

Формулу (6.41) можно переписать в виде

$$\rho_{\text{кор}} = \sqrt{2} \lambda / \theta,$$

где θ – угловой размер квадратного источника при наблюдении его из точки S_1 . При квадратном источнике света

пространственная когерентность обеспечивается внутри квадрата со стороной, равной λ_0 / a .

Щелевой источник света применяют для обеспечения когерентности по одной оси X . В этом случае ширина щели, при которой обеспечивается размер когерентности d , будет равна $a = \lambda_0 / d$. Следовательно, радиус когерентности щелевого источника равен

$$\rho_{\text{ког}} = \lambda_0 / a.$$

Заметим, что это соотношение можно получить из соотношения (6.40), если перейти к пределу при $b \rightarrow \infty$.

Глава 7

ДИФРАКЦИЯ СВЕТА НА ПЕРИОДИЧЕСКИХ СТРУКТУРАХ

7.1. Амплитудная гармоническая решетка

Дифракция Фраунгофера на периодических структурах имеет большое практическое применение. Она широко используется во многих экспериментальных методах спектрального анализа. Дифракция рентгеновских лучей на кристаллах позволяет получить важные сведения о структуре кристаллов, определить межатомное расстояние. Любую пространственно-периодическую структуру, способную изменить (модулировать) амплитуду или фазу падающей на нее световой волны, в оптике называют дифракционной решеткой. Периодически меняющимся параметром может быть, например, прозрачность, геометрическая толщина или коэффициент преломления. Первая решетка называется амплитудной, а последние две – фазовыми дифракционными решетками. Решетки, свойства структуры которых периодически меняются только в одном направлении,

называются одномерными или линейными. Если периодичность решетки имеет место в двух или трех направлениях, то решетку называют соответственно двух- или трехмерной.

Рассмотрим сначала простой случай одномерной амплитудной гармонической решетки.

Представим себе плоский прозрачный объект (транспарант), функция пропускания которого по амплитуде поля изменяется в одном направлении, например в направлении оси X , по гармоническому закону (рис. 7.1)

$$t(x, y) = t(x) = t_0 + t_1 \cos \frac{2\pi}{d} x. \quad (7.1)$$

Такой объект называют амплитудной гармонической дифракционной решеткой. В выражении (7.1) t_0 и t_1 – постоянные действительные коэффициенты, удовлетворяющие тем условиям, чтобы функция $t(x)$ была неотрицательной и не превышала значения, равного единице, т.е. чтобы $t_0 + t_1 \leq 1$, d – период изменения функции $t(x)$, называемый периодом решетки; величину $2\pi / d$ называют круговой (или циклической) пространственной частотой решетки, а число периодов решетки, приходящихся на единицу длины, т.е. величину $1 / d$ – (линейной) пространственной частотой решетки. Часто имеют дело с решетками, у которых $t_0 = t_1 = 1 / 2$, а также с решетками, у которых $t_0 = 1 / 2$, а $t_1 = (1/2)m$, где m – параметр, равный разности между максимальным и минимальным значениями функции пропускания (коэффициент модуляции). Функцию пропускания одномерной гармонической решетки иногда записывают в виде

$$t(x) = B \left(1 + m \cos \frac{2\pi}{d} x \right),$$

где $B = t_0$ – среднее за период функции пропускания $t(x)$ решетки, $m = t_1 / t_0$ – коэффициент модуляции решетки.

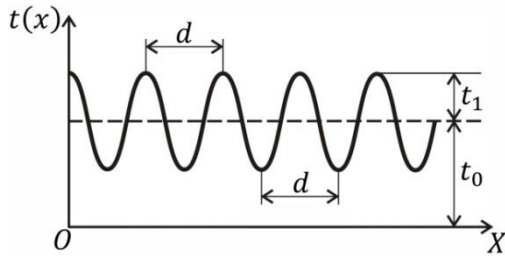


Рис. 7.1

Плоскость XU располагается в плоскости решетки, начало системы координат помещено в середину решетки.

Амплитудную гармоническую решетку можно получить, если косинусоидальное распределение интенсивности (3.3), возникающее при интерференции плоских волн, распространяющихся под углом друг к другу, зарегистрировать на фотопластинке.

Пусть на гармоническую решетку нормально падает плоская монохроматическая световая волна, радиус когерентности которой превышает размеры решетки; амплитуда плоской волны в плоскости $z = 0$ равна E_0 . Схема опыта такая же, как показано на рис. 6.2, в рассматриваемом случае A и B – края решетки, $a = AB$ – ширина решетки. Действие решетки на падающую волну состоит в том, что она пространственно модулирует амплитуду этой волны и не вносит фазового сдвига, так как функция $t(x)$ – вещественная. Световое поле за решеткой в плоскости $z > 0$, согласно приближению тени, получим, умножив амплитуду поля $E(x, y, 0) = E_0 t(x)$ в плоскости $z = 0$ на фазовый множитель $\exp(ikz)$. Представим функцию пропускания решетки в комплексной форме. Поскольку $t(x)$ – вещественная функция, то

$$t(x) = t_0 + \frac{1}{2} t_1 \exp\left(i \frac{2\pi}{d} x\right) + \frac{1}{2} t_1 \exp\left(-i \frac{2\pi}{d} x\right).$$

С учетом этого для амплитуды поля за решеткой будем иметь

$$E(x, z) = t_0 E_0 \exp(ikz) + \frac{1}{2} t_1 E_0 \exp \left[i \left(kz + \frac{2\pi}{d} x \right) \right] + \\ + \frac{1}{2} t_1 E_0 \exp \left[i \left(kz - \frac{2\pi}{d} x \right) \right].$$

Первый член полученного выражения описывает плоскую волну, распространяющуюся вдоль оси Z . Амплитуда волны равна $t_0 E_0$, волновой вектор направлен вдоль оси Z (циклическая пространственная частота $k_x = 0$). Второй член также описывает плоскую волну с амплитудой $(1/2)t_0 E_0$. Для определения направления ее распространения сравним ее фазу с фазой плоской волны той же амплитуды, распространяющейся в плоскости XZ под углом φ к оси Z , которая, согласно формуле (1.16), равна

$$\frac{2\pi}{\lambda} \sin \varphi \cdot x + \frac{2\pi}{\lambda} \cos \varphi \cdot z. \quad (7.2)$$

Из сравнения видно, что $\cos \varphi = 1$, $\sin \varphi / \lambda = 2\pi / d$. Это означает, что плоская волна, представленная вторым членом, распространяется под углом φ к оси Z , причем этот угол мал, так как $\cos \varphi \approx 1$. Аналогично, сравнивая фазовый множитель третьего члена с (7.2), видим, что $\sin \varphi / \lambda = -2\pi / d$ или $\sin \varphi_1 / \lambda = 2\pi / d$, где $\varphi_1 = -\varphi$. Следовательно, третий член выражения для $E(x, z)$ описывает плоскую волну той же амплитуды, что и плоская волна, описываемая вторым членом, но распространяющуюся под углом $-\varphi$ к оси Z .

Таким образом, плоская волна, проходя через гармоническую решетку, разбивается на три волны (на три световых пучка). Первая волна (с круговой пространственной частотой $k_x = 0$) распространяется в том же направлении, что и падающая на

решетку волна; ее называют дифрагировавшей волной нулевого порядка ($m = 0$). Две другие волны (с круговыми пространственными частотами $k_x = 2\pi/d$ и $k_x = -2\pi/d$) распространяются симметрично оси Z под углами φ и $-\varphi$ к ней; их называют дифрагировавшими волнами плюс первого ($m = +1$) и минус первого ($m = -1$) порядков соответственно. Углы $\pm\varphi$ при этом определяются соотношениями

$$d \sin \varphi = \pm m \lambda. \quad (7.3)$$

Распределение амплитуды поля в дальней зоне, т.е. в области дифракции Фраунгофера, с точностью до несущественного постоянного множителя получим, найдя фурье-преобразование функции распределения амплитуды поля непосредственно за решеткой и учтя, что функция пропускания отлична от нуля только в интервале $(-a/2, a/2)$, где a – ширина решетки:

$$\begin{aligned} E(\varphi) &= \int_{-\infty}^{\infty} E(x) e^{-i2\pi u x} dx = t_0 E_0 \int_{-a/2}^{a/2} \exp(-i2\pi u x) dx + \\ &+ \frac{1}{2} t_1 E_0 \int_{-a/2}^{a/2} \exp[-i2\pi(u - 1/d)x] dx + \\ &+ \frac{1}{2} t_1 E_0 \int_{-a/2}^{a/2} \exp[-i2\pi(u + 1/d)x] dx. \end{aligned}$$

Вычислив интегралы, получим

$$\begin{aligned} E(\varphi) &= a t_0 E_0 \cdot \text{sinc}(ua) + \frac{a}{2} t_1 E_0 \cdot \text{sinc}[(u - 1/d) a] + \\ &+ \frac{a}{2} t_1 E_0 \cdot \text{sinc}[(u + 1/d) a]. \end{aligned}$$

Возведя это выражение в квадрат, получим распределение интенсивности света в дальней зоне. Если период решетки d много меньше ее полуширины $a/2$ ($d \ll a/2$), то перекрытием трех sinc-функций можно пренебречь. Тогда получим

$$I(\varphi) = (at_0E_0)^2 \operatorname{sinc}^2(ua) + \left(\frac{a}{2}t_1E_0\right)^2 \operatorname{sinc}^2\left[\left(u - \frac{1}{d}\right)a\right] + \left(\frac{a}{2}t_1E_0\right)^2 \operatorname{sinc}^2\left[\left(u + \frac{1}{d}\right)a\right].$$

Из полученного результата следует, что помимо основного максимума, определяемого направлением $\varphi = 0$, которое получается из условия максимума первого слагаемого $ua = 0$, т.е. $k_x = 2\pi u = 0$, возникают два симметричных относительно основного побочных максимума, направления на которые определяются условиями максимума второго $((u - 1/d)a = 0)$ и третьего $((u + 1/d)a = 0)$ слагаемых, т.е. условиями $k_x = \pm 2\pi/d$ или (7.3). Пространственный спектр гармонической решетки (его называют также Винеровским спектром), таким образом, состоит из трех частот: $u = 0$ и $u = \pm 1/d$.

График зависимости интенсивности I от пространственной частоты $u = \sin \varphi / \lambda$ представлен на рис. 7.2. На рисунке видно, что побочные максимумы очень близко примыкают к центральному максимуму, а их интенсивности меньше интенсивности основного максимума, так как амплитуды отклоненных волн меньше, чем амплитуда неотклоненной волны.

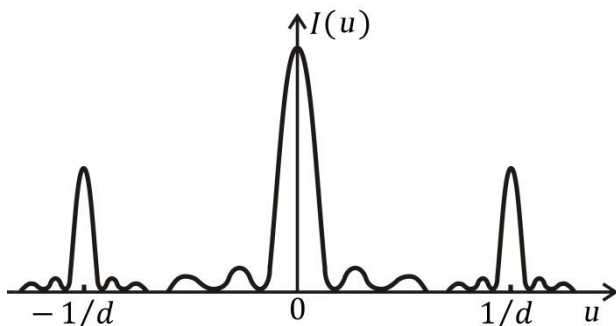


Рис. 7.2

Таким образом, гармоническое изменение амплитудной функции пропускания решетки приводит к перераспределению энергии от центрального максимума к двум боковым максимумам (максимумам первого порядка). Центральный максимум получается из неотклоненной волны, а максимумы ± 1 -го порядка – из боковых (отклоненных) волн. Других максимумов (максимумов более высокого порядка) амплитудная гармоническая решетка не создает. Координаты максимумов можно найти, переписав условие (7.3) в виде $u d = 0, \pm 1$ и подставив значение $u = x' / \lambda f$. Тогда получим

$$x' = 0, \pm \lambda f / d. \quad (7.4)$$

Отсюда видно, что чем больше период решетки d , тем ближе боковые максимумы располагаются к центральному максимуму.

Полуширину побочных максимумов $\Delta x_1'$ можно определить из условия $\text{sinc}\{[(x_1' + \Delta x_1') / \lambda f - 1 / d] a\} = 0$, или

$$[(x_1' + \Delta x_1') / \lambda f - 1 / d] a = 1, \quad (7.5)$$

где учтено, что $u = x' / \lambda f$ и $\text{sinc}(1) = 0$; x_1' – координата максимума первого порядка. Подставляя в (7.5) значение $x_1' = \lambda f / d$, получим

$$\Delta x_1' = \lambda f / a.$$

Полуширина центрального максимума $\Delta x_0'$ определяется из условия

$$[(0 + \Delta x_0') / \lambda f] a = 1$$

($x_0' = 0$). Откуда

$$\Delta x_0' = \lambda f / a.$$

Как видим, полуширина центрального максимума такая же, как и у побочных максимумов. Угловая полуширина всех максимумов также одинакова и равна

$$\Delta \theta = \Delta x' / f = \lambda / a.$$

Эта величина определяется только шириной решетки (при заданном значении длины волны падающего на решетку монохроматического света).

На экране, расположенном в плоскости $X'Y'$, пространственный спектр гармонической решетки наблюдается в виде трех световых пятен, соответствующих нулевому и ± 1 -му порядкам.

Если ширина решетки велика ($a \gg d$), т.е. при $a \rightarrow \infty$, спектр в области фраунгоферовской дифракции, как легко убедиться, будет описываться тремя дельта-функциями:

$$E(\varphi) = t_0 E_0 \cdot \delta(ua) + \frac{1}{2} t_1 E_0 \cdot \delta(u - 1/d) + \frac{1}{2} t_1 E_0 \cdot \delta(u + 1/d),$$

и будет изображаться тремя резкими линиями, исходящими из точек $x' = 0$, $x' = \lambda f / d$ и $x' = -\lambda f / d$. Им соответствуют три строго параллельных световых пучка. При не очень большой ширине решетки эти пучки не будут строго параллельными, что и приводит к конечной ширине дифракционных максимумов (когда они покрывают некоторый интервал углов).

В случае, когда амплитудное пропускание транспаранта является периодической с периодом d , но не гармонической функцией, эту функцию можно разложить в ряд Фурье:

$$t(x) = \sum_{m=-\infty}^{\infty} t_m \exp\left(i \frac{2m\pi}{d} x\right),$$

$$t_m = \frac{1}{d} \int_{-d/2}^{d/2} t(x) \exp\left(-i \frac{2m\pi}{d} x\right) dx.$$

При падении на такую решетку плоской волны с комплексной амплитудой E_0 поле на выходе транспаранта будет описываться функцией

$$E(x) = E_0 t(x) = E_0 \sum_{m=-\infty}^{\infty} t_m \exp\left(i \frac{2m\pi}{d} x\right),$$

а на расстоянии z от транспаранта – функцией

$$E(x, z) = E_0 \sum_{m=-\infty}^{\infty} t_m \exp\left[i\left(kz + \frac{2m\pi}{d} x\right)\right] =$$

$$= E_0 t_0 \exp(ikz) + E_0 t_{-1} \exp\left[i\left(kz - \frac{2\pi}{d} x\right)\right] +$$

$$+ E_0 t_1 \exp\left[i\left(kz + \frac{2\pi}{d} x\right)\right] + E_0 t_{-2} \exp\left[i\left(kz - \frac{4\pi}{d} x\right)\right] +$$

$$+ E_0 t_2 \exp\left[i\left(kz + \frac{4\pi}{d} x\right)\right] + \dots$$

Это соотношение показывает, что за транспарантом с периодической функцией пропускания будет распространяться система плоских волн. При $m = 0$ возникает одна волна нулевого порядка, при $m = \pm 1$ – две волны первого порядка, при $m = \pm 2$ – две волны второго порядка и т.д. Направление распространения этих волн определяется углами φ , удовлетворяющими равенству

$$d \sin \varphi = m\lambda,$$

где $m = 0, \pm 1, \pm 2, \dots$, а интенсивность – коэффициентами разложения t_m .

7.2. Фазовая гармоническая решетка

Мы рассмотрели дифракционную решетку, изменяющую только амплитуду падающей на нее волны (модулирующую волну по амплитуде). Существуют одномерные решетки, которые изменяют только фазу волны (модулирующие волну по фазе). Такие решетки называются фазовыми одномерными дифракционными решетками. Их функция пропускания является комплексной величиной

$$t(x) = A \exp[i\varphi(x)].$$

Модуль A этой функции не зависит от координаты x , а фаза $\varphi(x)$ изменяется по периодическому (в частности, по гармоническому) закону. Рассмотрим здесь простой случай, когда фаза $\varphi(x)$ изменяется по синусоидальному закону с периодом d и коэффициентом модуляции m ; среднее значение фазы равно φ_0 . Тогда

$$\varphi(x) = m \sin \frac{2\pi}{d} x + \varphi_0,$$

а функция пропускания запишется в виде

$$t(x) = B \exp\left(i \sin \frac{2\pi}{d} x\right),$$

где $B = A \exp(i\varphi_0)$. Используя тождество

$$\exp(m \sin \alpha) = \sum_{n=-\infty}^{\infty} J_n(m) \exp(in\alpha),$$

где $J_n(m)$ – функция Бесселя первого рода n -го порядка, аргумента m , функцию пропускания решетки представим в виде

$$t(x) = B \sum_{n=-\infty}^{\infty} J_n(m) \exp\left(in \frac{2\pi}{d} x\right).$$

Если на такую решетку падает плоская волна с амплитудой E_0 , то световое поле в плоскости непосредственно за решеткой будет иметь вид

$$E(x) = E_0 t(x) = C \sum_{n=-\infty}^{\infty} J_n(m) \exp\left(in \frac{2\pi}{d} x\right),$$

где $C = BE_0$, а в плоскости $z > 0$ – вид

$$E(x, z) = C \sum_{n=-\infty}^{\infty} J_n(m) \exp\left[i\left(kz + n \frac{2\pi}{d} x\right)\right].$$

Распределение поля в дальней зоне, как и в случае амплитудной решетки, получим, найдя преобразование Фурье от распределения поля непосредственно за решеткой:

$$\begin{aligned} E(\varphi) &= \int_{-a/2}^{a/2} \left[C \sum_{n=-\infty}^{\infty} J_n(m) \exp\left(in \frac{2\pi}{d} x\right) \right] e^{-i2\pi u x} dx = \\ &= C \sum_{n=-\infty}^{\infty} J_n(m) \int_{-a/2}^{a/2} \exp\left[-i2\pi\left(u - \frac{n}{d}\right)x\right] dx = \\ &= Ca \sum_{n=-\infty}^{\infty} J_n(m) \operatorname{sinc}\left[\left(u - \frac{n}{d}\right)a\right], \end{aligned}$$

ИЛИ

$$E(x') = Ca \sum_{n=-\infty}^{\infty} J_n(m) \operatorname{sinc}\left[\frac{a}{\lambda f} \left(x' - n \frac{\lambda f}{d}\right)\right].$$

Распределение интенсивности вдоль оси X' , как и в предыдущем случае, можно найти, возведя это выражение в квадрат. Считая, что период решетки $d \ll a / 2$, перекрытием различных дифракционных порядков можно пренебречь. Тогда получим

$$I(x') = (Ca)^2 \sum_{n=-\infty}^{\infty} J_n^2(m) \operatorname{sinc}^2 \left[\frac{a}{\lambda f} \left(x' - n \frac{\lambda f}{d} \right) \right].$$

Видим, что фазовая гармоническая решетка, в отличие от амплитудной решетки, кроме центрального максимума интенсивности (если m не является корнем уравнения $I_0^2(m) = 0$), дает не два, а множество побочных максимумов. Эти максимумы имеют координаты $x_n' = n \frac{\lambda f}{d}$, $n = 0, \pm 1, \pm 2, \pm 3, \dots$. Значение интенсивности в n -ом максимуме пропорционально квадрату функции Бесселя n -го порядка. Из соотношения

$$[(x_n' + \Delta x_n') / \lambda f - n / d] a = 1$$

определим полуширину максимума n -го порядка. Подставляя сюда значение x_n' , получим $\Delta x_n' = \lambda f / a$. Эта величина одинакова для всех максимумов (не зависит от n) и определяется тем же выражением, что и для амплитудной решетки. Она не зависит от периода решетки, а определяется ее шириной (при заданных значениях λ и f).

В общем случае дифракционные решетки могут вызывать изменение и амплитуды, и фазы волны. Функция пропускания таких решеток также является комплексной величиной. Но в этом случае зависят от координат и ее модуль, и фаза.

Пользуясь преобразованием Фурье, любую реальную решетку можно представить в виде суммы некоторого числа гармонических амплитудных и фазовых решеток. В общем случае дифракционного экрана (пользуясь опять-таки преобразованием Фурье) его функцию пропускания можно заменить суммой бесконечного числа гармонических решеток с всевозможными

пространственными частотами и различными коэффициентами модуляции. В результате дифракции света на этих решетках мы получаем тот сложный спектр пространственных частот, который наблюдается при дифракции на всевозможных экранах.

Полученные в этих двух параграфах результаты находят широкое применение в оптике, в частности, при описании механизма образования голографических изображений.

7.3. Плоская одномерная щелевая решетка

7.3.1. Дифракция света на щелевой решетке

Одномерная щелевая решетка представляет собой систему одинаковых и расположенных в одной плоскости параллельных друг другу щелей одинаковой ширины, разделенных непрозрачными промежутками; середины щелей (или любые соответственные точки) находятся на одинаковом расстоянии d друг от друга. В простом случае – это прозрачная стеклянная плоскопараллельная пластинка с нанесенными на ней непрозрачными штрихами. Величину d называют периодом (или постоянной) щелевой решетки, а обратную ей величину $1/d$ – пространственной частотой решетки; $d = a + b$, где a – ширина щели, b – ширина непрозрачного промежутка. Лучшие решетки имеют пространственную частоту до 1200 периодов на 1 мм. Эта решетка тоже является амплитудной решеткой.

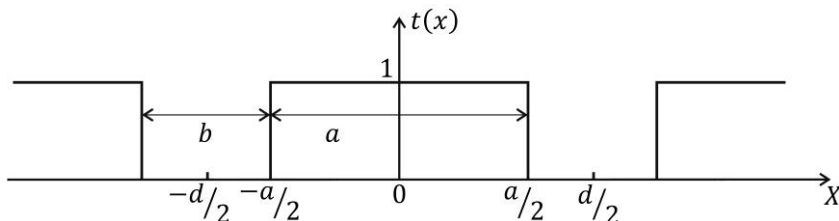


Рис. 7.3

Плоскую одномерную щелевую дифракционную решетку можно рассматривать как транспарант, функция пропускания (или, как ее еще называют, оптическая структура) которого зависит только от одной координаты x : $t(x, y) = t(x)$. Если начало системы координат, плоскость XU которой расположена в плоскости решетки, поместить в середину первой щели, то график функции $t(x)$ щелевой решетки можно изобразить в виде системы прямоугольных импульсов (шириной a и расстоянием между их серединами d), как показано на рис. 7.3. Аналитически ее можно записать в виде

$$t(x) = \sum_{n=1}^N t_1(x - x_n),$$

где N – число щелей решетки, $x_m = (m - 1)d$, $m = 1, 2, 3, \dots, N$, x_n – координаты центров щелей

$$t_1(x) = \begin{cases} 1, & |x| \leq a/2, \\ 0, & |x| > a/2, \end{cases}$$

$t_1(x)$ – функция пропускания одной щели.

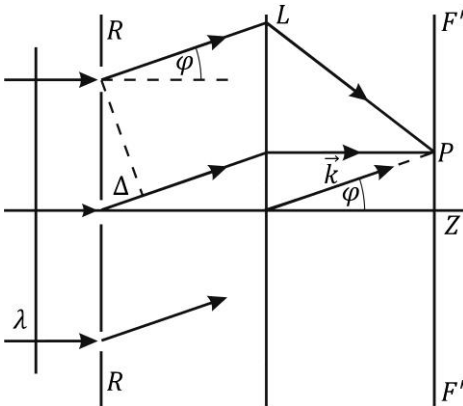


Рис. 7.4

Будем считать, что
плоская

монохроматическая волна с амплитудой E_0 падает на плоскую
одномерную щелевую решетку нормально
(рис. 7.4). Найдем результирующую амплитуду $E'(\varphi)$ колебаний
светового поля в точке P плоскости FF' , в которой собираются
параллельные лучи от всех щелей решетки, падающие на линзу L
под углом φ к оптической оси (оси Z), т.е. световые волны,
имеющие пространственную частоту $u = \sin\varphi / \lambda$.
Распределение амплитуды светового поля в плоскости FF'
определится как преобразование Фурье от апертурной функции
 $t(x)$ решетки:

$$E'(\varphi) = CE_0 \int_{-\infty}^{\infty} t(x) \exp(-i2\pi ux) dx, \quad (7.6)$$

где постоянная $C = \frac{\exp(ikz)}{\sqrt{i\lambda z}}$. Подставив в формулу (7.6)

функцию пропускания решетки, будем иметь

$$E'(\varphi) = CE_0 \sum_{n=1}^N \int_{-\infty}^{\infty} t_1[x - (n-1)d] \exp(-i2\pi ux) dx.$$

Введем новую переменную $x - (n-1)d = \zeta$. Тогда
 $x = \zeta + (n-1)d$, $dx = d\zeta$. Получим

$$\begin{aligned} E'(\varphi) &= CE_0 \sum_{n=1}^N \int_{-\infty}^{\infty} t_1(\zeta) \exp[-i2\pi u[\zeta + (n-1)d]] d\zeta = \\ &= CE_0 \left(\int_{-\infty}^{\infty} t_1(\zeta) \exp(-i2\pi u\zeta) d\zeta \right) \sum_{n=1}^N \exp[-i2\pi u(n-1)d]. \end{aligned}$$

Интеграл, заключенный в скобки, определяет дифракцию на одиночной щели. Поскольку все они одинаковы, в качестве таковой можно взять первую щель (соответствующую $n = 1$). В этом случае $\xi = x$, а пределы интегрирования будут $x = \pm a/2$, и мы получим

$$\begin{aligned} E'(\phi) &= CE_0 \int_{-a/2}^{a/2} \exp(-i2\pi ux) dx \cdot \sum_{n=1}^N \exp[-i2\pi u(n-1)d] = \\ &= CE_0 a \frac{\sin(\pi ua)}{\pi ua} f_N(u) = E_1(0) \operatorname{sinc}(ua) f_N(u), \end{aligned}$$

где $E_1(0)\operatorname{sinc}(ua)$ – амплитуда поля волны, испытавшей дифракцию на отдельной щели, $E_1(0)$ – амплитуда поля, создаваемого одной щелью в направлении, определяемом углом $\varphi = 0$,

$$f_N(u) = \sum_{n=1}^N \exp[-i2\pi u(n-1)d]$$

– фактор многолучевой интерференции, описывающий совокупное действие всех щелей. Его можно найти как сумму первых $n = N$ членов геометрической прогрессии, первый член которой $b_1 = 1$, а знаменатель $q = \exp(-i2\pi ud)$:

$$f_N(x) = \frac{\exp(-i2\pi Nud) - 1}{\exp(-i2\pi ud) - 1}.$$

Преобразовав это выражение, приведем его к виду

$$f_N(x) = \frac{\exp(i\pi Nud)}{\exp(i\pi ud)} \cdot \frac{\sin \pi Nud}{\sin \pi ud} = \frac{\sin N\pi ud}{\sin \pi ud} \cdot \exp[i(N-1)\pi ud].$$

С учетом этого находим

$$E'(\varphi) = (E_1(0) \cdot \text{sinc}(ua)) \cdot \left(\frac{\sin N\pi ud}{\sin \pi ud} \cdot \exp[-i(N-1)\pi ud] \right). \quad (7.7)$$

Мы видим, что результирующая амплитуда поля в произвольной точке P представляется в виде произведения двух сомножителей, заключенных соответственно в квадратные и круглые скобки. Сомножитель в квадратных скобках характеризует распределение амплитуды светового поля от дифракции на одной щели; он представляет собой результат интерференции волн, дифрагировавших на каждой щели в отдельности. Сомножитель в круглых скобках обусловлен наличием решетки, т.е. правильной периодической структуры из N щелей.

Легко убедиться, что этот сомножитель учитывает интерференцию пучков, исходящих из всех щелей решетки. Действительно, в одном и том же направлении все щели решетки излучают свет совершенно одинаково, поэтому комплексные амплитуды колебаний, создаваемых в точке P волнами, посылаемыми всеми N щелями решетки, одинаковы по модулю, но отличаются друг от друга фазовым множителем $\exp(-i\delta)$, учитывающим опережение по фазе на $\delta = (2\pi/\lambda)\Delta$ колебаний от каждой следующей щели по сравнению с предыдущей, обусловленное, как видно из рис. 7.4, разностью хода волн $\Delta = d \sin \varphi$ от соседних щелей до точки наблюдения. Следовательно,

$$E_2 = E_1 \exp(-i\delta), E_3 = E_2 \exp(-i\delta) = E_1 \exp(-i2\delta), \dots,$$

$$E_n = E_{n-1} \exp(-i\delta) = E_1 \exp[-i(n-1)\delta], \dots$$

Световые пучки от всех щелей когерентны между собой, так как они являются составными частями единого исходного пучка света, падающего на решетку, и в них представлены различные

части одних и тех же элементарных пучков. Поэтому между этими когерентными пучками будет происходить интерференция. Тогда, в соответствии с правилом сложения комплексных амплитуд при интерференции волн, для результирующего поля в произвольной точке P можно записать

$$E'(\varphi) = \sum_{n=1}^N E_n = E_1(\varphi) \sum_{n=1}^N e^{-i(n-1)\delta} = E_1(\varphi) \sum_{n=1}^N e^{-i2\pi(n-1)ud},$$

где учтено, что $\delta = (2\pi/\lambda)\Delta = (2\pi/\lambda)d \sin\varphi = 2\pi ud$. Мы снова

получили множитель $f_N(u) = \sum_{n=1}^N e^{-i2\pi(n-1)ud}$ и тем самым

показали, что этот множитель учитывает интерференцию многочисленных когерентных пучков, посылаемых отдельными щелями, или, другими словами, определяет суммарное поле в направлении, задаваемом углом φ , создаваемое системой N щелей. Этот дополнительный множитель называют решеточным множителем.

Интенсивность света, распространяющегося под углом φ к направлению падающей на решетку волны,

$$I(u) = I_1(0) \operatorname{sinc}^2(ua) \left(\frac{\sin(N\pi ud)}{\sin(\pi ud)} \right)^2 = I_1(\varphi) \left(\frac{\sin(N\pi ud)}{\sin(\pi ud)} \right)^2, \quad (7.8)$$

где $I_1(0)$ – интенсивность света, создаваемая одной щелью, в направлении падающей на решетку волны ($\varphi = 0$), $I_1(\varphi)$ – интенсивность от одной щели в направлении угла φ . Таким образом, при дифракции плоской волны на дифракционной решетке происходит многолучевая интерференция, при которой интенсивности дифрагировавших пучков одинаковы, а фаза каждого следующего пучка отличается от фазы предыдущего на величину $\delta = 2\pi ud$.

Переходя к анализу полученного результата, выясним сначала роль интерференции N пучков. Из (7.8) видно, что при $\pi u d = m\pi$, т.е. при $u = \sin\varphi / \lambda = m / d$, где $m = 0, \pm 1, \pm 2, \dots$, решеточный множитель

$$I_N = \left(\frac{\sin(N\pi u d)}{\sin(\pi u d)} \right)^2,$$

учитывающий указанную многолучевую интерференцию, достигает своего максимального значения, равного N^2 . В этом случае интенсивность света, создаваемая решеткой в направлении, определяемом углом φ , в N^2 раз больше интенсивности света, посылаемого в данном направлении одной щелью:

$$I(u) = N^2 I_1(u).$$

Возникающие при этом максимумы называют главными дифракционными максимумами; в направлении этих максимумов распространяется основная часть падающего на решетку света. Целое число m называют порядком главного дифракционного максимума.

Условие $\sin\varphi / \lambda = m / d$ можно записать в виде

$$d \sin\varphi = m\lambda. \quad (7.9)$$

Это соотношение определяет направление на главный дифракционный максимум m -го порядка. Заметим, что формула (7.3) является частным случаем этой формулы при $m = 0, \pm 1$.

Из соотношения (7.9) вытекает также условие образования дифракционной картины на решетке: $d \geq \lambda$. Действительно, только при таком соотношении между периодом решетки d и длиной волны λ света (при $m \neq 0$) равенство (7.9) может выполняться (так как $\sin\varphi \leq 1$).

Таким образом, если разность хода волн, испускаемых в направлении угла φ двумя эквивалентными точками соседних щелей, будет равна целому числу длин волн, то в этом направлении волны в результате их интерференции усилят друг друга, создавая результирующую интенсивность в N^2 раз большую интенсивности, даваемой одной щелью. Это есть прямое следствие правильности структуры. Если бы N щелей располагались в решетке беспорядочно, то при наложении волн от каждой щели взаимные сдвиги фаз δ изменялись бы хаотически; посылаемые щелями волны были бы некогерентными. В результате имело бы место простое сложение интенсивностей волн. Суммарная интенсивность была бы только в N раз больше интенсивности, создаваемой одной щелью.

В отличие от гармонической решетки, дающей только три максимума, щелевая решетка дает большое число максимумов. Это число можно получить с помощью формулы (7.9) с учетом условия $\sin \varphi \leq 1$:

$$m_{\max} = \text{целое число} \leq d / \lambda. \quad (7.10)$$

Однако заметную интенсивность, как и в случае двух щелей, имеют лишь те дифракционные максимумы, которые лежат в пределах центрального дифракционного максимума от одной щели.

Таким образом, можно констатировать, что если на щелевую решетку падает плоская световая волна, то ее волновой фронт после решетки представляется множеством плоских волн, распространяющихся в определенных направлениях, удовлетворяющих условию (7.9).

Решеточный множитель I_N обращается в нуль всякий раз, когда числитель $\sin(N\pi ud) = 0$, а знаменатель $\sin(\pi ud) \neq 0$, т.е. при $N\pi ud = (mN + n)\pi$, или

$$d \sin \varphi = (m + n / N)\lambda,$$

где $n = 1, 2, 3, \dots, N - 1$. В соответствующих таким значениям пространственной частоты u и направлениях φ лежат минимумы интенсивности. Эти минимумы называются добавочными минимумами. Главными минимумами называются минимумы, возникающие от одной щели решетки.

График зависимости решеточного множителя I_N от пространственной частоты u показан на рис. 7.5. Из этого рисунка видно, что в промежутках между главными максимумами, расположенными при $u = m / d$, имеются так называемые побочные максимумы, интенсивность которых исчезающе мала по сравнению с интенсивностью света в главных максимумах. Видно также, что главные максимумы дифракционных порядков $\pm m$ расположены парами симметрично относительно центрального (нулевого порядка) максимума и чем меньше период решетки d , тем дальше главные максимумы отстоят друг от друга.

Ближайший к максимуму нулевого порядка дополнительный минимум определяется условием $N\pi u d = \pi$. Откуда, учитывая, что $Nd \gg \lambda$, получаем $\varphi_1 \approx \sin \varphi_1 = \lambda / (Nd)$. Это соотношение определяет угловую полуширину главного дифракционного максимума:

$$\delta\varphi = \varphi_1 \approx \lambda / (Nd).$$

Линейная полуширина

$$\Delta x = f \delta\varphi = \lambda f / (Nd),$$

где f – фокусное расстояние линзы. Как видим, максимум тем острее, т.е. тем резче, чем большее число щелей N содержит решетка.

Выясним теперь влияние, которое оказывает на распределение интенсивности одна щель. Сомножитель $I_1(u) = I_1(0)\text{sinc}^2(ua)$ в (7.9), определяющий это распределение, дает огибающую колебаний решеточного сомножителя. Для углов,

удовлетворяющих условию $a \sin \varphi = m\lambda$, $m = \pm 1, \pm 2, \pm 3, \dots$, $I_1(u) = 0$ и $I(u) = 0$. Как видим, эти углы соответствуют минимумам при дифракции на одной щели и определяют направления главных минимумов интенсивности при дифракции на решетке.

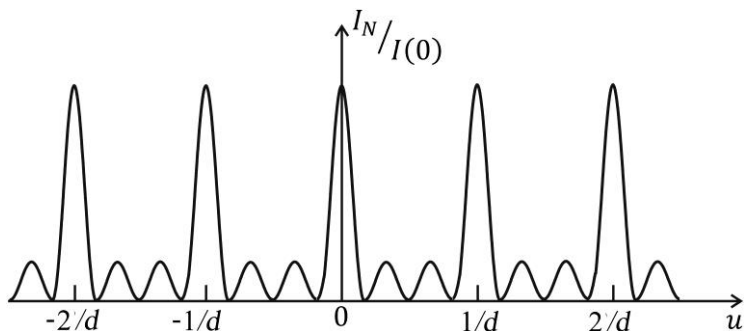


Рис. 7.5

Найдем теперь значение $I_1(u)$ при выполнении условия $u = m / d$. Подставив это значение u в выражение для $I_1(u)$, получим

$$I_{1m}(u) = I_1(0) \operatorname{sinc}^2\left(\frac{\pi m a}{d}\right). \quad (7.11)$$

Тогда интенсивность m -го главного максимума

$$I_m(u) = N^2 I_{1m}(u) = I_1(0) \operatorname{sinc}^2\left(\frac{\pi m a}{d}\right). \quad (7.12)$$

или

$$I_m(u) = N^2 I_1(0) d^2 \frac{\sin^2(\pi m a / d)}{\pi^2 m^2 a^2}. \quad (7.13)$$

Из выражения (7.13) видно, что $I_m(u) \sim 1/m^2$, т.е. с увеличением порядка дифракции m интенсивность соответствующего главного максимума уменьшается пропорционально $1/m^2$. Видно также, что эта интенсивность существенно зависит и от отношения ширины щели a к периоду решетки d . Если $m(a/d) = n$ – целое число, то интенсивность света в этом m -ом главном максимуме равна нулю. Но при этом направления на главные максимумы и главные минимумы совпадают:

$$(\sin \varphi)_{\max} = m\lambda/d = n\lambda/a = (\sin \varphi)_{\min}.$$

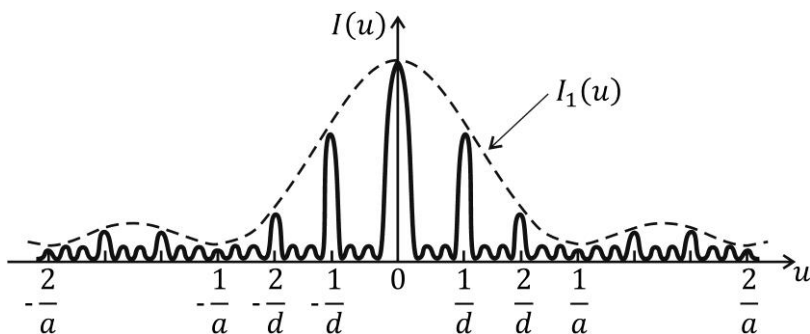


Рис. 7.6

Следовательно, в этих направлениях главные максимумы не возникают. Например, если у дифракционной решетки $a = b$ (где b – ширина непрозрачного промежутка), то $d = 2a$, и при четном m ($m = 2n$) величина $I_m(u) = 0$, поэтому в дифракционной картине от такой решетки исчезают все четные главные максимумы, при $a/d = 1/3$ выпадает каждый третий главный максимум ($m = 3n$), при $a/d = 1/4$ – каждый четвертый ($m = 4n$) и т.д.

График суммарного распределения интенсивности показан на рис. 7.6. Сплошная кривая на этом рисунке изображает главные и

добавочные максимумы и минимумы. Пунктирная кривая, проходящая через вершины главных максимумов (огibaющая), показывает распределение интенсивности при дифракции от одной щели, умноженной на N^2 .

7.3.2. Разложение апертурной функции щелевой решетки в ряд Фурье

Для дальнейшего изучения свойств щелевой дифракционной решетки проведем разложение ее апертурную функцию $t(x)$ в ряд Фурье. Число щелей N решетки велико. Можно считать, что это число $N \rightarrow \infty$, поэтому апертурную функцию $t(x)$ решетки можно считать периодической с периодом, равным d : $t(x) = t(x + nd)$, где n – целое число. Учитывая также, что эта функция имеет разрывы первого рода, а значит, условия Дирихле выполняются, ее можно разложить в ряд Фурье

$$t(x) = \frac{A_0}{2} + \sum_{m=1}^{\infty} \left(A_m \cos \frac{\pi mx}{l} + B_m \sin \frac{\pi mx}{l} \right),$$

где

$$A_0 = \frac{1}{l} \int_{-l}^l t(x) dx,$$

$$A_m = \frac{1}{l} \int_{-l}^l t(x) \cos \frac{\pi mx}{l} dx,$$

$$B_m = \frac{1}{l} \int_{-l}^l t(x) \sin \frac{\pi mx}{l} dx$$

– коэффициенты ряда Фурье, $\pm l$ – границы интервала разложения (l – половина периода функции $t(x)$). Ряд Фурье для такой функции будет зависеть от того, где расположено начало координат: в середине щели или на ее границе. В первом случае разложение будет проводиться по косинусам (когда все $B_m = 0$), а во втором – по синусам (когда все $A_m = 0$).

Поместим начало координат в середину одной из щелей. Тогда в интервале $(-d/2, d/2)$ функция $t(x)$ будет иметь вид

$$t(x) = \begin{cases} 1, & -a/2 \leq x \leq a/2; \\ 0, & -d/2 \leq x < -a/2, a/2 < x \leq d/2. \end{cases}$$

График этой функции показан на рис. 7.3.

При выбранном начале координат функция $t(x)$ является четной, поэтому в интервале $(-d/2, d/2)$ она может быть разложена в ряд Фурье по косинусам:

$$t(x) = \frac{A_0}{2} + \sum_{m=1}^{\infty} A_m \cos \frac{\pi m x}{l}.$$

В нашем случае $l = d/2$. Следовательно,

$$\begin{aligned} A_0 &= \frac{2}{d} \int_{-d/2}^{d/2} t(x) dx = \frac{2}{d} \int_{-a/2}^{a/2} dx = \frac{2a}{d}; \\ A_m &= \frac{2}{d} \int_{-d/2}^{d/2} t(x) \cos \frac{2\pi m x}{d} dx = \frac{2}{d} \int_{-a/2}^{a/2} \cos \frac{2\pi m x}{d} dx = \\ &= \frac{2a}{d} \cdot \operatorname{sinc} \left(\frac{ma}{d} \right). \end{aligned} \quad (7.14)$$

С учетом этого получаем

$$t(x) = \frac{a}{d} + \sum_{m=1}^{\infty} A_m \cos \frac{2\pi m x}{d}. \quad (7.15)$$

Поскольку косинус – четная функция, этот ряд можно представить в виде

$$t(x) = \sum_{m=-\infty}^{\infty} \frac{A_m}{2} \cos \frac{2\pi mx}{d}. \quad (7.16)$$

Коэффициент с $m = 0$ в сумме (7.16) тот же, что и в сумме (7.15): $A_0/2 = a/d$. При других значениях m коэффициенты $A_m/2$ образуют пары $A_m/2$ и $A_{-m}/2$, равные половине первоначального значения A_m . Члены ряда

$$t_m(x) = \frac{A_m}{2} \cos \frac{2\pi mx}{d}$$

называются пространственными гармониками номера m , числа m/d – пространственными частотами, а обратные им величины d/m – пространственными периодами. Причем, поскольку учитываются положительные и отрицательные значения m , каждому значению m соответствует не одна, а пара гармоник с амплитудами $A_{\pm m}/2$ и пространственными частотами $\pm m/d$.

Таким образом, функцию пропускания щелевой решетки можно представить в виде бесконечной совокупности пар ($\pm m$) пространственных гармоник (фурье-компонент) с амплитудами $A_{\pm m}/2$ и пространственными частотами $\pm m/d$. Другими словами, щелевую решетку можно рассматривать как совокупность бесконечного числа гармонических решеток с периодами d/m , $m = 1, 2, 3, \dots$

Амплитуды $A_m/2$ пар ($\pm m$) гармоник $t_m(x)$ апертурной функции $t(x)$ пропорциональны амплитудам поля главных дифракционных максимумов m -го порядка, создаваемых щелевой решеткой. Действительно, подставив в выражение амплитуды поля решетки (7.7) значение $u = m/d$ и перейдя к пределу при $u \rightarrow m/d$ в сомножителе в круглых скобках, получим

$$E_m'(\varphi) = N E_1(0) \operatorname{sinc}\left(\frac{ma}{d}\right) e^{-i(N-1)m\pi}.$$

Сравнивая это соотношение с соотношением (7.14), убеждаемся в справедливости указанного утверждения, т.е. в том, что $A_m / 2 \sim E_m'(\varphi)$. Следовательно, амплитуды пар гармоник m -го порядка определяют амплитуды пар m -го порядка главных дифракционных максимумов, создаваемых решеткой с апертурной функцией $t(x)$. Направления φ_m этих дифракционных максимумов также определяются гармониками пространственных частот m/d , так как $m/d = \sin \varphi_m / \lambda$. Отсюда вытекает, что пары главных дифракционных максимумов $\pm m$ -го порядка обусловлены соответствующими парами пространственных гармоник $t_m(x)$ в разложении апертурной функции решетки $t(x)$ в ряд Фурье. Поэтому если в этом разложении отсутствует какой-либо член, то будут отсутствовать и главные дифракционные максимумы соответствующего порядка. Это мы видели на примере гармонической дифракционной решетки. В разложении ее апертурной функции отличны от нуля только коэффициенты $A_0 / 2 (= t_0)$ и $A_1 / 2 (= t_1)$, поэтому она создает дифракционную картину с максимумами только нулевого ($m = 0$) и двух первых ($m = \pm 1$) порядков.

Проведенное рассмотрение показывает, что дифракционная картина, создаваемая щелевой решеткой, представляет собой проявление, почти физическое воплощение, тех гармоник, которые составляют функцию пропускания решетки. При дифракции плоской волны на решетке происходит как бы физическое разложение функции пропускания $t(x)$ решетки в ряд Фурье, причем в фокальной плоскости линзы наблюдается дифракционная картина, максимумы в которой соответствуют отдельным фурье-компонентам функции $t(x)$.

При исследовании оптических систем часто используют периодический тест-объект в виде миры, которую называют

прямоугольной или П-образной мрой. Она представляет собой щелевую решетку с шириной непрозрачного промежутка b , равной ширине щели a : $a = b = d/2$. Полагая в формулах коэффициентов разложения A_0 и A_m значение $a = d/2$, получим

$$A_0 = 1, A_m = \frac{2}{m\pi} \sin \frac{m\pi}{2}.$$

Разложение функции пропускания мрой в ряд Фурье по косинусам будет иметь вид

$$t(x) = \frac{1}{2} + \frac{2}{\pi} \cos \frac{2\pi x}{d} - \frac{2}{3\pi} \cos \frac{3 \cdot 2\pi x}{d} + \frac{2}{5\pi} \cos \frac{5 \cdot 2\pi x}{d} - \dots$$

Если начало координат совместить с границей щели, получим разложение функции пропускания мрой по синусам. В этом случае

$$A_m = \frac{2}{d} \int_0^{d/2} t(x) \sin \frac{2\pi mx}{d} dx = -\frac{1}{m\pi} [\cos m\pi - 1].$$

Следовательно, указанное разложение будет иметь вид

$$t(x) = \frac{1}{2} + \frac{2}{\pi} \sin \frac{2\pi x}{d} + \frac{2}{3\pi} \sin \frac{3 \cdot 2\pi x}{d} + \frac{2}{5\pi} \sin \frac{5 \cdot 2\pi x}{d} + \dots$$

Заметим, что в отличие от знакопеременного ряда по косинусам, ряд по синусам состоит только из положительных членов.

7.3.3. Распределение амплитуды поля в дифракционной картине щелевой решетки как произведение двух фурье-преобразований

Распределение амплитуды светового поля в дифракционной картине, создаваемой щелевой решеткой, согласно (7.7),

определяется произведением двух сомножителей, первый из которых $\text{sinc}(ua)$ является фурье-преобразованием апертурной функции щели. Покажем, что и второй (решеточный) сомножитель

$$f_N(x) = \frac{\sin N\pi ud}{\sin \pi ud} \exp[-i(N-1)\pi ud]$$

также является фурье-преобразованием, но от так называемой решеточной структуры, на которой распределены отдельные щели, т.е. от регулярной структуры, представляющей собой последовательность идентичных апертур (щелей), образующих решетку. Для этого решеточную структуру математически представим в виде последовательности дельта-функций $\delta(x - x_n)$, расположенных в точках $x_n = (n - 1)d$ ($n = 1, 2, 3, \dots, N - 1$), являющихся координатами середин щелей, т.е. для описания распределения щелей в решетке используем функцию вида

$$f(x) = \sum_{n=1}^N \delta(x - x_n) = \sum_{n=1}^N \delta[x - (n - 1)d]. \quad (7.17)$$

Эта функция является частным случаем часто используемой в оптике так называемой гребенчатой функции

$$\text{comb}(x) = \sum_{m=-\infty}^{\infty} \delta(x - md),$$

где m – число периодов d вдоль оси X .

Учтем, что фурье-преобразование дельта-функции $\delta(x - x_m)$

$$\begin{aligned} F\{\delta(x - x_m)\} &= \int_{-\infty}^{\infty} \delta(x - x_m) e^{-i2\pi ux} dx = \\ &= e^{-i2\pi ux_m} = e^{-i2\pi u(n-1)d}. \end{aligned} \quad (7.18)$$

Тогда с учетом свойства фурье-преобразования, согласно которому фурье-преобразование суммы функций равно сумме фурье-преобразований этих функций, будем иметь

$$\begin{aligned} F\{f(x)\} &= F\left\{\sum_{n=1}^N \delta(x-x_n)\right\} = \sum_{n=1}^N F\{\delta[x-(n-1)d]\} = \\ &= \sum_{n=1}^N e^{-i2\pi u(n-1)d} = \frac{\sin \pi Nud}{\sin \pi ud} \cdot e^{-i(N-1)\pi ud}. \end{aligned}$$

Полученное выражение совпадает с решеточным множителем соотношения (7.7). Тем самым мы показали, что преобразование Фурье от линейной последовательности N регулярно расположенных δ -функций является решеточным множителем N регулярно расположенных щелей.

Таким образом, множитель, соответствующий одиночной щели, и решеточный множитель в выражении (7.7) являются преобразованием Фурье соответственно одноапертурной функции и последовательности δ -функций, определяющих структуру решетки. Дифракционная картина, создаваемая решеткой, является произведением указанных двух фурье-преобразований.

Из соотношения (7.18) следует, что фурье-преобразование функции $\delta(x)$ равно единице ($F\{\delta(x)\} = 1$). Учитывая это и используя обратное фурье-преобразование, можно записать

$$\delta(x) = \int_{-\infty}^{\infty} F\{\delta(x)\} e^{i2\pi ux} du = \int_{-\infty}^{\infty} e^{i2\pi ux} du.$$

В общем случае

$$\delta(x) = \int_{-\infty}^{\infty} e^{\pm i2\pi ux} du = \lim_{\varepsilon \rightarrow \infty} \int_{-\varepsilon/2}^{\varepsilon/2} e^{\pm i2\pi ux} du = \lim_{\varepsilon \rightarrow \infty} \frac{\sin \varepsilon \pi x}{\pi x},$$

так как δ -функция четная. Эту формулу можно считать условным определением δ -функции. Для двумерной δ -функции

$$\begin{aligned} \delta(x, y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\pm i2\pi(ux+vy)} dudv = \\ &= \lim_{\varepsilon \rightarrow \infty} \int_{-\varepsilon/2}^{\varepsilon/2} \int_{-\varepsilon/2}^{\varepsilon/2} e^{\pm i2\pi(ux+vy)} dudv = \lim_{\varepsilon \rightarrow \infty} \frac{\sin \varepsilon\pi x}{\pi x} \frac{\sin \varepsilon\pi y}{\pi y}. \end{aligned}$$

Эта формула будет использована нами при рассмотрении вопросов формирования оптического изображения в главах 11 и 12.

7.3.4. Разрешающая способность щелевой решетки

Из соотношения $d \sin \varphi = m\lambda$ видно, что положения главных максимумов, кроме центрального ($m = 0$), в дифракционной картине от щелевой решетки зависят от длины волны используемого света λ . Поэтому если решетка освещается белым или другим некогерентным светом, то для разных значений λ все дифракционные максимумы, кроме центрального, окажутся пространственно разделенными. В результате в дифракционной картине решетки, освещаемой белым светом, центральный максимум будет иметь вид белой полосы, а все остальные – радужных полос, называемых дифракционными спектрами первого ($m = \pm 1$), второго ($m = \pm 2$) и т.д. порядков. В спектрах каждого порядка наиболее отклоненными будут красные лучи (с большим значением λ , так как $\sin \varphi \sim 1 / \lambda$), а наименее – фиолетовые (с меньшим значением λ). Спектры получаются тем более четкими (в смысле разделения цветов), чем больше щелей N содержит решетка. Это следует из того, что линейная полуширина максимума обратно пропорциональна числу щелей N). Максимальное число наблюдаемых дифракционных спектров

определяется соотношением (7.11). Таким образом, дифракционная решетка производит разложение сложного излучения на отдельные монохроматические составляющие, т.е. проводит гармонический анализ падающего на него излучения.

Свойство дифракционной решетки разлагать сложное излучение на гармонические составляющие используется в спектральных аппаратах – приборах, служащих для исследования спектрального состава излучения, т.е. для получения спектра излучения и определения длин волн и интенсивностей всех его монохроматических компонент.

Как спектральный прибор дифракционная решетка характеризуется разрешающей способностью, под которой понимают способность решетки давать раздельно две близкие спектральные линии. Эта величина определяется отношением

$$R = \lambda / \delta\lambda,$$

где $\delta\lambda$ – минимальная разность длин волн двух спектральных линий, при которой эти линии воспринимаются как раздельные спектральные линии. Величину $\delta\lambda$ называют разрешаемым спектральным расстоянием. Вследствие дифракции на действующем отверстии объектива⁵, расположенного за дифракционной решеткой, каждая спектральная линия изображается спектральным аппаратом не в виде линии, а в виде дифракционной картины, распределение интенсивности в которой имеет вид sinc^2 -функции. Так как спектральные линии с различными длинами волн не когерентны, то результирующая дифракционная картина, создаваемая такими линиями, будет представлять собой простое наложение дифракционных картин от

⁵ Этот объектив создает монохроматическое изображение щели, через которую проходит свет, падающий на дифракционную решетку, в излучении разных длин волн в фокальной плоскости объектива. Эти изображения (спектральные линии) в своей совокупности и составляют спектр исследуемого излучения.

каждой щели в отдельности; результирующая интенсивность будет равна сумме интенсивностей

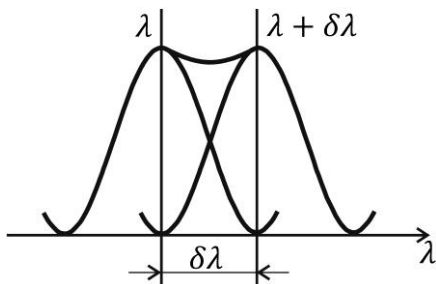


Рис. 7.7

обеих линий. Согласно критерию Рэля, спектральные линии с близкими длинами волн λ и $\lambda + \delta\lambda$ считаются разрешенными, если они находятся на таком расстоянии $\delta\lambda$, при котором главный дифракционный максимум

одной линии совпадает по своему положению с первым дифракционным минимумом другой линии. В этом случае на кривой суммарного распределения интенсивности (рис. 7.7) образуется провал (глубиной, равной $0,26I_0$, где I_0 – максимальная интенсивность, одинаковая для обеих спектральных линий), что и позволяет глазу воспринимать такую картину как двойную спектральную линию. В противном случае две близко расположенные спектральные линии воспринимаются как одна уширенная линия.

Положение m -го главного дифракционного максимума, соответствующего длине волны λ , определяется координатой

$$x_{m1}' = f \operatorname{tg} \varphi \approx f \sin \varphi = m \lambda f / d.$$

Аналогично находим и положение m -го максимума, соответствующего длине волны $\lambda + \delta\lambda$:

$$x_{m2}' = m(\lambda + \delta\lambda) f / d.$$

При выполнении критерия Рэля расстояние между этими максимумами составит

$$\Delta x' = x_{m2}' - x_{m1}' = m(\delta\lambda) f / d.$$

Оно равно полуширине $\delta x = \lambda f / d$ этих максимумов (здесь, как и выше, полуширину мы определяем по первому нулю интенсивности). Отсюда находим

$$\delta \lambda = \lambda / (mN),$$

и, следовательно, разрешающая способность дифракционной решетки как спектрального прибора

$$R = mN.$$

Таким образом, разрешающая способность дифракционной решетки пропорциональна числу щелей N и порядку спектра m . Положив $m = m_{\max} \approx d / \lambda$, получим максимальную разрешающую способность:

$$R_{\max} = (\lambda / \delta \lambda)_{\max} = m_{\max} N \approx L / \lambda,$$

где $L = Nd$ – ширина рабочей части решетки. Как видим, максимальная разрешающая способность щелевой решетки определяется только шириной рабочей части решетки и средней длиной волны исследуемого излучения. Зная R_{\max} , найдем минимально разрешимый интервал длин волн:

$$(\delta \lambda)_{\min} \approx \lambda^2 / L.$$

Что касается разрешающей способности гармонической решетки, то для ее определения заметим, что чем уже побочные максимумы в ее дифракционной картине и чем больше расстояние между ними, тем легче наблюдать отдельно эти линии. Поэтому разрешающую способность гармонической дифракционной решетки определяют как отношение расстояния между соседними максимумами (которое равно координате максимума первого порядка) x_1' к полуширине максимума $\Delta x'$: $R = x_1' / \Delta x'$. Подставив сюда значения $x_1' = \lambda f / d$ и $\Delta x_1' = \lambda f / a$, получим

$$R = a / d.$$

Следовательно, разрешающая способность гармонической решетки равна числу ее периодов.

7.4. Решетки периодические относительно квадратичных переменных

7.4.1. Решетка периодическая относительно одной квадратичной переменной

Рассмотрим одномерную дифракционную решетку, функция пропускания которой

$$t(x) = t_0 + t_1 \cos\left(k \frac{x^2}{2a}\right),$$

где t_0 и t_1 – безразмерные постоянные, такие что $t_0 + t_1 \leq 1$, $t_1 < t_0$, a – постоянная, имеющая размерность длины, $k = 2\pi / \lambda$ – волновое число, λ – длина волны, на которую рассчитана решетка. Эта решетка является периодической по квадрату координаты x . Решетку с такой функцией пропускания можно получить, зарегистрировав на фотопластинке распределение интенсивности (3.19), которое возникает при интерференции плоской и цилиндрической волн. Представим функцию пропускания решетки $t(x)$ в виде трех слагаемых:

$$t(x) = t_0 + \frac{1}{2} t_1 \exp\left(ik \frac{x^2}{2a}\right) + \frac{1}{2} t_1 \exp\left(-ik \frac{x^2}{2a}\right).$$

Если на такую решетку падает плоская монохроматическая волна с длиной волны λ и с амплитудой E_0 , то в плоскости непосредственно за решеткой (плоскости $z = 0$) распределение комплексной амплитуды светового поля будет

$$E(x) = E_0 t(x) = t_0 E_0 + \frac{1}{2} t_1 E_0 \exp\left(ik \frac{x^2}{2a}\right) + \frac{1}{2} t_1 E_0 \exp\left(-ik \frac{x^2}{2a}\right). \quad (7.19)$$

В соответствии с приближением геометрической оптики такое же распределение поля будет и в плоскости $z > 0$. Выражение (7.19) показывает, что за рассматриваемой дифракционной решеткой распространяются три волны. Первая – плоская волна с амплитудой $t_0 E_0$ и с направлением распространения, совпадающим с направлением падающей на решетку плоской волны; вторая и третья – цилиндрические волны, причем одна из них (определяемая вторым членом в (7.19)) – расходящаяся, а другая (третий член в (7.19)) – сходящаяся. Радиусы цилиндрических волновых фронтов равны $\pm a$. Полученные три волны соответствуют дифракционным порядкам $m = 0, \pm 1$. Максимумы других порядков отсутствуют.

7.4.2. Зонная решетка Габора

Рассмотрим теперь дифракционную решетку с круговой симметрией, функция пропускания которой является периодической по квадрату расстояния $\rho = \sqrt{x^2 + y^2}$ от центра до произвольной точки решетки:

$$t(\rho) = t_0 + t_1 \cos\left(k \frac{\rho^2}{2a}\right), \quad (7.20)$$

где t_0 , t_1 и a – постоянные, причем t_0 и t_1 удовлетворяют тем же условиям, что и для рассмотренной выше решетки с цилиндрической симметрией. Координатная плоскость XU расположена в плоскости решетки, начало координат

расположено в центре решетки. График функции $t(\rho)$ показан на рис. 7.8. Круговая решетка с функцией пропускания вида (7.20) называется зонной решеткой Габора. Она представляет собой трафарет из колец радиуса $\rho_m = \sqrt{m\lambda a}$, где m – целое число (четное для светлых колец и нечетное для темных), a – расстояние от волнового фронта до точки наблюдения, λ – рабочая длина волны, в которой прозрачные кольца чередуются с непрозрачными. Такую решетку можно

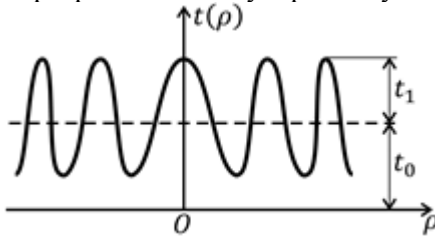


Рис. 7.8

получить, если зарегистрировать на фотопластинке распределение интенсивности (3.17), возникающее при интерференции плоской и сферической волн. Теория дифракции на такой

решетке имеет большое значение в голографии.

Представим выражение для $t(\rho)$ в виде суммы трех слагаемых:

$$t(x, y) = t_0 + \frac{1}{2} t_1 \exp\left(ik \frac{x^2 + y^2}{2a}\right) + \frac{1}{2} t_1 \exp\left(-ik \frac{x^2 + y^2}{2a}\right).$$

Предположим, что такая решетка освещается плоской монохроматической волной с длиной волны λ и с амплитудой E_0 . Тогда в плоскости непосредственно за решеткой ($z = 0$) распределение комплексной амплитуды светового поля будет иметь вид

$$E(x, y) = E_0 t(x, y) = t_0 E_0 + \frac{1}{2} t_1 E_0 \exp\left(ik \frac{x^2 + y^2}{2a}\right) +$$

$$+ \frac{1}{2} t_1 E_0 \exp \left(-ik \frac{x^2 + y^2}{2a} \right).$$

Такое же распределение поля будет иметь вид и в плоскости $z > 0$. Первый член полученной суммы соответствует плоской волне, распространяющейся в направлении падающей на решетку волны, а второй и третий – соответственно расходящейся и сходящейся сферическим волнам с радиусами волновых фронтов $R = \pm a$. Это вытекает из выражения сферической волны в параксиальном приближении.

Таким образом, в результате дифракции плоской монохроматической волны на зонной решетке с периодической функцией пропускания возникают три волны (рис. 7.9): плоская волна, распространяющаяся в направлении падающей волны, и две сферические волны, одна из которых является расходящейся из точки S_1 , расположенной на расстоянии a перед решеткой, а другая – сходящаяся в точке S_2 , расположенной на расстоянии a за решеткой. Плоская волна имеет амплитуду $t_0 E_0$, сферические волны – амплитуду $(1/2)t_1 E_0$. Эти волны соответствуют дифракционным порядкам $m = 0, \pm 1$. Других порядков зонная решетка Габора не дает.

Свойством преобразовывать плоскую волну в волну сферическую обладает линза. Следовательно, действие зонной решетки

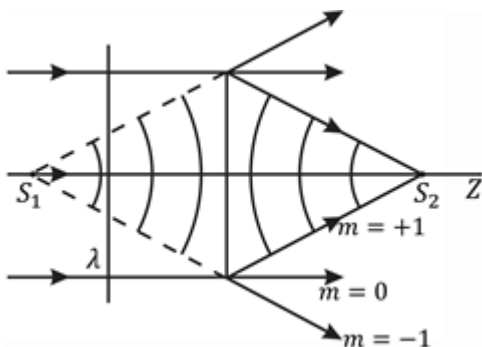


Рис. 7.9

Габора на плоскую волну равносильно действию двух линз: одной собирающей, другой – рассеивающей.

Положительная и отрицательная линзы, соответствующие зонной решетке, имеют фокусные расстояния $f = \pm a$. Действия

зонной пластинки и линзы на падающие на них лучи отличаются друг от друга лишь тем, что в зонной пластинке лучи отклоняются за счет дифракции, а в линзе – за счет рефракции.

В общем случае транспаранта с периодической по $q = \rho^2$, но не гармонической функцией пропускания $t(\rho)$ эту функцию можно представить рядом Фурье:

$$t(\rho) = \sum_{m=-\infty}^{\infty} t_m \exp\left(im \frac{k}{2a} q\right) = \sum_{m=-\infty}^{\infty} t_m \exp\left(im \frac{k}{2a} \rho^2\right).$$

Коэффициенты разложения

$$\begin{aligned} t_m &= \frac{1}{2\pi} \cdot \frac{k}{2a} \int_0^{4\pi a/k} t(q) \exp\left(-im \frac{k}{2a} q\right) dq = \\ &= \frac{k}{2\pi a} \int_0^{\sqrt{4\pi a/k}} t(\rho) \exp\left(-im \frac{k}{2a} \rho^2\right) \rho d\rho. \end{aligned}$$

Если на транспарант с такой функцией пропускания падает плоская волна с комплексной амплитудой E_0 , то на его выходе будем иметь распределение амплитуды поля

$$\begin{aligned}
E(\rho) &= E_0 t(\rho) = E_0 \sum_{m=-\infty}^{\infty} t_m \exp\left(i m \frac{k}{2a} \rho^2\right) = \\
&= E_0 t_0 + E_0 t_{-1} \exp\left(-i \frac{k}{2a} \rho^2\right) + E_0 t_1 \exp\left(i \frac{k}{2a} \rho^2\right) + \\
&+ E_0 t_{-2} \exp\left(-i 2 \frac{k}{2a} \rho^2\right) + E_0 t_2 \exp\left(i 2 \frac{k}{2a} \rho^2\right) + \dots
\end{aligned}$$

Первый член в этом разложении соответствует волне нулевого порядка – плоской волне. Следующие два члена соответствуют волнам первого порядка – сходящимся и расходящимся сферическим волнам с радиусами волновых фронтов $R_1 = \pm a$. Последующие два члена соответствуют волнам второго порядка; они тоже описывают сходящуюся и расходящуюся сферические волны, но с радиусами фронтов $R_2 = \pm a / 2 = R_1 / 2$ и т.д. Следовательно, за транспарантом с периодической по ρ^2 функцией пропускания распространяется совокупность пар сходящихся и расходящихся сферических волн с радиусами волновых фронтов

$$R_m = \pm a / m = R_1 / m.$$

Интенсивность этих волн определяется коэффициентами разложения t_m .

7.4.3. Зонная пластинка Френеля

Рассмотрим свободно распространяющуюся плоскую волну. Пусть плоский фронт волны в момент времени t занимает положение $\Phi(t)$, при котором он находится на расстоянии z от точки наблюдения P (рис. 7.10, a). Для определения светового поля в точке P фронт волны $\Phi(t)$ мысленно разобьем на кольцевые участки (зоны), построенные так, чтобы вторичные

волны Гюйгенса, приходящие в точку P от двух соседних участков (зон), оказались в противофазе, т.е. чтобы расстояния от границ (или любых других соответственных точек) двух соседних зон до точки P отличались на $\lambda / 2$:

$$M_1P - M_0P = M_2P - M_1P = \dots = M_mP - M_{m-1}P = \lambda / 2.$$

Построенные таким способом кольцевые участки волнового фронта (зоны) называются зонами Френеля (рис. 7.10, б, цифры 1, 2, 3 обозначают номера зон).

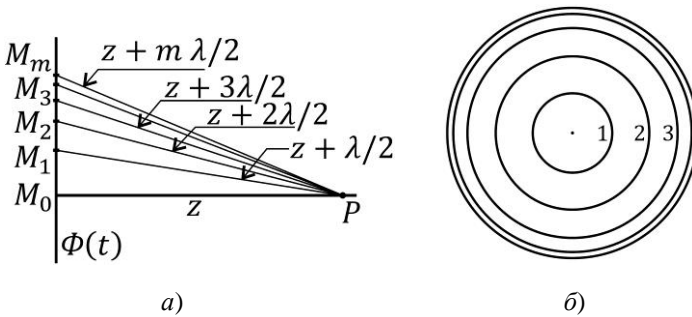


Рис. 7.10

Найдем радиус ρ_m внешней границы m -ой зоны Френеля (рис. 7.11). Так как расстояние M_mP от внешнего края m -ой зоны до точки P равно $z + m \frac{\lambda}{2}$, то, как видно на рис. 7.11,

$$\rho_m = \sqrt{\left(z + m \frac{\lambda}{2}\right)^2 - z^2} = \sqrt{m\lambda z + m^2 \frac{\lambda^2}{4}}.$$

Если ограничиться рассмотрением не слишком больших m , $m \ll \ll 4z / \lambda$, то вторым членом в подкоренном выражении можно пренебречь, и тогда для радиуса внешней границы m -ой зоны получим

$$\rho_m = \sqrt{m\lambda z}. \quad (7.21)$$

Из формулы (7.21) имеем $\rho_m^2 = m\lambda z$, откуда $2\rho_m \Delta\rho_m = \lambda z \Delta m$. Положив в этом соотношении $\Delta m = 1$, найдем расстояние между соседними кольцевыми зонами Френеля (ширину колец):

$$\Delta\rho_m = \lambda z / 2\rho_m.$$

Следовательно, ширина кольцевых зон Френеля $\Delta\rho_m$ убывает с увеличением их радиуса ρ_m . При этом площади всех зон одинаковы (не зависят от номера зоны m):

$$\Delta\sigma = \pi\rho_m \Delta\rho_m = (1/2)\pi\lambda z = \text{const.}$$

Результирующая амплитуда светового поля в точке P будет равна сумме амплитуд E_m ($m = 1, 2, 3, \dots$), создаваемых в точке P каж-

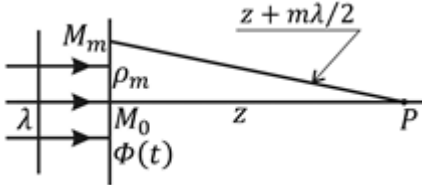


Рис. 7.11

дой из зон в отдельности. Колебания, возбуждаемые в точке P двумя соседними зонами, находятся в противофазе, поэтому при наложении волн эти колебания будут ослаблять друг друга. Это означает,

что амплитуда поля, создаваемая каждой последующей зоной, будет отличаться от амплитуды поля, создаваемой предыдущей зоной, не только модулем, но и фазовым множителем $\exp(i\pi) = -1$, т.е. знаком. Следовательно, результирующая амплитуда в точке P будет представляться знакопеременным рядом

$$E(P) = \sum_m (-1)^{m+1} E_m = E_1 - E_2 + E_3 - \dots \pm E_m \mu \dots \quad (7.22)$$

Вследствие увеличения расстояния от центра каждой последующей зоны до точки наблюдения P величины амплитуд монотонно убывают с ростом номера зоны.

В качестве допустимого приближения можно принять, что амплитуда поля E_m от каждой m -ой зоны Френеля равна среднему арифметическому от амплитуд примыкающих к ней зон:

$$E_m = \frac{1}{2}(E_{m-1} + E_{m+1}).$$

Тогда, представив каждый нечетный член ряда (7.22) в виде двух его половин, с учетом этого приближения находим

$$E(P) = (1/2) E_1.$$

Следовательно, амплитуда светового поля, создаваемая плоской волной в точке P при полностью открытом волновом фронте, равна половине амплитуды, создаваемой первой зоной Френеля. Таким образом, вследствие интерференции вторичных волн Гюйгенса действие всего открытого фронта сводится к действию половины первой (центральной) зоны Френеля.

Интенсивность света $I(P) = |E(P)|^2$ в точке P можно многократно увеличить, если на пути света расположить экран, который перекрывает все четные или все нечетные зоны Френеля. В этом случае вторичные волны от всех открытых зон Френеля будут приходить в точку P в одинаковой фазе и в результате интерференции усилят друг друга. Такой экран называют амплитудной зонной пластинкой Френеля.

Эта пластинка представляет собой систему концентрических чередующихся светлых и темных колец, построенных по принципу расположения зон Френеля. Она имеет прямоугольное радиальное распределение пропускания. График этого распределения показан на рис. 7.12, *a*. Из этого рисунка видно, что как ширина прямоугольных импульсов, так и расстояние между ними с удалением от центра пластинки уменьшается; высота прямоугольников остается одинаковой и равной 1.

Простейшая амплитудная зонная пластинка – это стеклянная пластинка, на которой (для определенных условий наблюдения, т.е. для определенных значений z и λ) нанесены зоны Френеля.

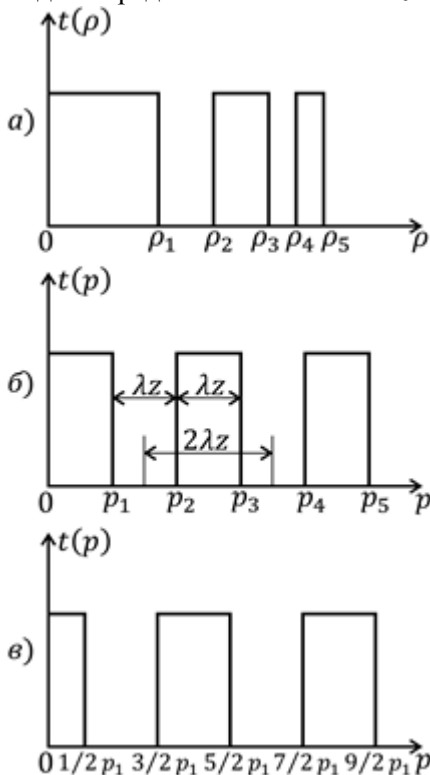


Рис. 7.12

Каждая открытая зона создает в точке P амплитуду, примерно равную E_1 . Поэтому суммарная амплитуда в точке P определится как

$$E = (N / 2)E_1 = E_1 R^2 / (2\lambda z).$$

Но $E_1 \approx 2E_0$, где E_0 – амплитуда плоской волны, падающей на пластинку. С учетом этого находим интенсивность в точке P :

Действие амплитудной зонной пластинки описывается рядом (7.22), в котором оставлены члены только одного знака. Усиление интенсивности света зонной пластинкой аналогично фокусирующему действию собирающей линзы, поэтому с ее помощью можно получать удовлетворительные оптические изображения.

Найдем интенсивность света в точке фокусировки P . Если радиус пластинки равен R , то число N зон Френеля, укладываемых на пластинке, определится из условия $R = \rho_N = \sqrt{N\lambda z}$, откуда $N = R^2 / (z\lambda)$. Число открытых зон равно $(1 / 2)N = R^2 /$

$$I = |E|^2 = |E_0|^2 R^4 / (\lambda b)^2 = I_0 R^4 / (\lambda z)^2.$$

Выигрыш в интенсивности составляет

$$I / I_0 = R^4 / (\lambda z)^2.$$

Например, при $R = 1$ см, $z = 10$ см и $\lambda = 0,5$ мкм получаем $I / I_0 \approx 4 \cdot 10^6$, т.е. несколько миллионов.

На рис. 7.13 изображена амплитудная зонная пластинка Френеля со светлым центром, у которой нечетные зоны оставлены прозрачными, а четные – зачернены (функция пропускания, изображенная на рис. 7.12, а, соответствует именно такой пластинке). Такая пластинка удаляет из волнового фронта все четные зоны Френеля. У нее внутренние радиусы колец пропорциональны квадратным корням из последовательных нечетных чисел, а внешние – из четных. При этом у светлых колец внутренние радиусы пропорциональны квадратным корням из последовательных четных чисел, а внешние – из нечетных. Можно изготовить и аналогичную пластинку с темным центром. Эта пластинка удаля-



Рис. 7.13

ет все нечетные зоны. Как видим, зонная пластинка Френеля имеет такую же структуру, что и зонная решетка Габора.

Если положить $\rho_m^2 = p_m$, то $p_m = m\lambda z$. В координатах t, p функция пропускания t такой пластинки будет представлять собой систему прямоугольных импульсов высотой, равной 1, шириной и расстоянием между ними, равными λz (рис. 7.12, б). Период такой решетки $d = 2\lambda z$. Поместив начало координат в центр пластинки, получим четную функцию $t = t(\rho)$ (рис. 7.12, в). Такую функцию можно разложить в ряд Фурье по косинусам

$$t(x) = \frac{A_0}{2} + \sum_{m=1}^{\infty} A_m \cos \frac{\pi mx}{l},$$

где

$$A_0 = \frac{2}{l} \int_0^l t(x) dx, \quad A_m = \frac{2}{l} \int_0^l t(x) \cos \frac{\pi mx}{l} dx$$

– коэффициенты ряда Фурье, l – верхняя граница интервала разложения (половина периода функции $t(p)$). В интервале разложения $(0, \lambda z)$ функция $t = t(p)$ отлична от нуля в интервале $(0, \lambda z/2)$, поэтому в качестве верхнего предела в интегралах следует взять величину $\lambda z/2$. Тогда получим

$$\begin{aligned} A_0 &= \frac{2}{\lambda z} \int_0^{\lambda z/2} dp = 1, \quad A_m = \frac{2}{\lambda z} \int_0^{\lambda z/2} \cos \frac{\pi mp}{\lambda z} dp = \frac{2}{m\pi} \sin \frac{m\pi}{2} = \\ &= \frac{2(-1)^{m+1}}{\pi 2m-1}. \end{aligned}$$

Возвращаясь к старой переменной ρ , получим разложение функции $t(\rho)$ в ряд Фурье:

$$t(\rho) = \frac{1}{2} + \frac{2}{\pi} \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{2m-1} \cos \left(mk \frac{\rho^2}{2z} \right)$$

или

$$t(\rho) = \frac{1}{2} + \frac{1}{\pi} \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{2m-1} \left[\exp \left(imk \frac{\rho^2}{2z} \right) + \exp \left(-imk \frac{\rho^2}{2z} \right) \right].$$

Если осветить такую пластинку плоской волной, распространяющейся вдоль оси Z перпендикулярно пластинке, то на выходе получим одну плоскую волну с уменьшенной в два раза амплитудой и бесконечно большое число сходящихся и

расходящихся сферических волн с амплитудами $\frac{E_0}{\pi} \frac{(-1)^{m+1}}{2m-1}$, где

E_0 – комплексная амплитуда падающей на пластинку плоской волны, и с радиусами, равными $\pm z / m$.

7.5. Двумерные решетки

7.5.1. Решетка, полученная наложением двух скрещенных щелевых решеток

Двумерной периодической структурой (двумерной решеткой) называют плоскую структуру, свойства которой периодически изменяются в двух различных направлениях. Для ее описания требуется две координаты.

Простейшей двумерной структурой является решетка, полученная путем наложения двух скрещенных щелевых дифракционных решеток, когда щели одной решетки направлены перпендикулярно щелям другой решетки. Эта структура будет представлять собой периодически расположенную систему прямоугольных отверстий. Как и в одномерном случае, координаты x и y свяжем с плоскостью решетки, а координаты x' и y' – с плоскостью наблюдения. Задача о распределении поля в дифракционной картине решается с помощью двумерного преобразования Фурье (6.5). Направление на точку наблюдения P будем определять угла-

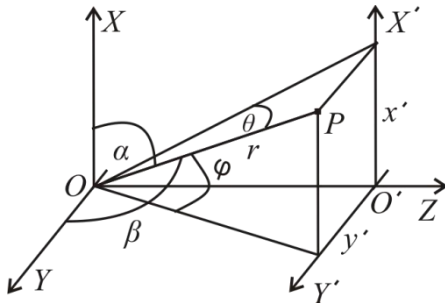


Рис. 7.14

ми φ и θ , дополняющими направляющие углы α и β до $\pi/2$, как показано на рис. 7.14. При этом

$$\sin \phi = x'/r, \quad \sin \theta = y'/r,$$

где r – длина отрезка OP , соединяющего начало координат O координатной плоскости

$X'Y'$ с точкой наблюдения P . При таком выборе угловых координат пространственные частоты $u = \sin \varphi / \lambda$ и $v = \sin \theta / \lambda$. Амплитудный коэффициент пропускания двумерной структуры будет зависеть от двух переменных x и y , но в случае ортогональных скрещенных решеток переменные разделяются:

$$t(x, y) = t_1(x)t_2(y),$$

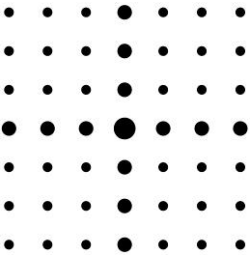
где $t_1(x)$ и $t_2(y)$ – коэффициенты пропускания одномерных решеток. При этом двойной интеграл в (6.5) распадается на произведение двух однократных интегралов, и дальнейший расчет фактически будет повторять сделанный в п. 7.3.1 расчет для одномерной решетки. Интенсивность света, распространяющегося под углами α и β к направлению падающей на решетку волны,

$$I(\varphi, \theta) = I_1(\varphi, \theta) \left(\frac{\sin(N_1 \pi u d_1)}{\sin(\pi u d_1)} \cdot \frac{\sin(N_2 \pi v d_2)}{\sin(\pi v d_2)} \right)^2, \quad (7.23)$$

где

$$I_1(\varphi, \theta) = I(0, 0) \text{sinc}^2(ua_1) \text{sinc}^2(va_2),$$

a_1 и a_2 – размеры прямоугольных отверстий, d_1 и d_2 – периоды



решетки вдоль осей X и Y соответственно, N_1 и N_2 – число щелей в направлении координатных осей X и Y . При $\pi u d_1 = m_1 \pi$ и $\pi v d_2 = m_2 \pi$, где m_1, m_2 – целые числа, решеточный множитель (множитель, стоящий за $I_1(\varphi, \theta)$ в соотношении (7.23)) достигает своего максимального значения.

Рис.7.15

Следовательно, направления на главные максимумы дифракционной картины будут определяться равенствами

$$d_1 \sin \varphi = m_1 \lambda, \quad d_2 \sin \theta = m_2 \lambda.$$

Дифракционная картина показана на рис. 7.15. Такой вид дифракционной картины объясняется тем, что решетка, штрихи которой направлены вертикально, дает систему лучей, лежащих в горизонтальной плоскости, а решетка, штрихи которой направлены горизонтально, дает систему лучей, лежащих в вертикальной плоскости.

Если положение точки P определять с помощью углов α и β (углов между прямой OP и осями координат X и Y), то, как видно из рис. 7.14, направляющие косинусы

$$\cos \alpha = \sin \varphi = x' / r, \quad \cos \beta = \sin \theta = y' / r.$$

Направления на главные максимумы при таком выборе угловых координат будут определяться формулами

$$d_1 \cos \alpha = m_1 \lambda, \quad d_2 \cos \beta = m_2 \lambda.$$

7.5.2. Решетка, состоящая из точечных отверстий

Рассмотрим теперь случай, когда двумерная решетка представляет собой систему большого числа периодически

расположенных точечных отверстий в непрозрачном экране. Решетка является прямоугольной. На рис. 7.16, *a* такая решетка изображена в виде точек, расположенных в узлах прямоугольной сетки. Свяжем с решеткой прямоугольную систему координат XU , начало O которой поместим в центр какого-либо отверстия, а оси X и U направим вдоль линий, соединяющих ближайшие соседние отверстия. Периоды решетки вдоль осей X и U обозначим через a и b . Тогда координаты отверстий будут $x_{mn} = ma$, $y_{mn} = nb$, где m и n – целые числа. Принимая направляющие векторы \mathbf{a} и \mathbf{b} за базисные, радиус-вектор, определяющий положение отверстия в решетке, представим в виде

$$\mathbf{p}_{mn} = m\mathbf{a} + n\mathbf{b}.$$

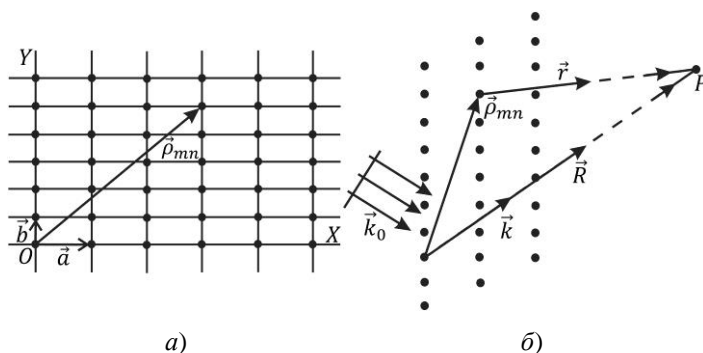


Рис. 7.16

Пусть на решетку падает плоская монохроматическая волна, волновой вектор которой \mathbf{k}_0 . В соответствии с принципом Гюйгенса каждое отверстие создает вторичные сферические волны, распространяющиеся за экраном. Нас интересует амплитуда светового поля в произвольной точке P , находящейся на большом удалении от решетки (рис. 7.16, *б*). Положение этой точки будем определять радиус-вектором \mathbf{R} , проведенным из начала координат O . Для определения указанной амплитуды следует учесть интерференцию полей, создаваемых в точке P

всеми отверстиями, т.е. просуммировать амплитуды полей от каждого отверстия. Для отверстия с центром, расположенным в точке $\mathbf{\rho}_{mn}$ решетки, эта амплитуда, очевидно, будет

$$E_{mn}(P) = CE(\mathbf{\rho}_{mn}) \frac{\exp(ikr)}{r}, \quad (7.24)$$

где r – расстояние от рассматриваемого отверстия до точки P , C – коэффициент пропорциональности, величина которого зависит от особенностей рассеивающего центра (отверстия), k – модуль волнового вектора \mathbf{k} волны, рассеянной в направлении точки P , $E(\mathbf{\rho}_{mn}) = E_0 \exp(i\mathbf{k}_0 \cdot \mathbf{\rho}_{mn})$ – амплитуда поля в точке $\mathbf{\rho}_{mn}$, создаваемая падающей на решетку плоской волной, E_0 – амплитуда падающей волны в точке O ; фазовый множитель падающей волны в точке O полагается равным единице. С помощью рис. 7.17, б получаем $\mathbf{r} = \mathbf{R} - \mathbf{\rho}_{mn}$. Так как $\rho_{mn} \ll R$, то пренебрегая членами порядка $(\rho_{mn}/R)^2$ и выше, для расстояния от отверстия $\mathbf{\rho}_{mn}$ до точки наблюдения P будем иметь

$$\begin{aligned} r &= \sqrt{(\mathbf{R} - \mathbf{\rho}_{mn})^2} = R\sqrt{1 - 2\mathbf{R} \cdot \mathbf{\rho}_{mn} / R^2 + (\rho_{mn} / R)^2} \approx \\ &\approx R - \mathbf{R} \cdot \mathbf{\rho}_{mn} / R. \end{aligned}$$

С учетом этого фазовый множитель $\exp(ikr)$ в (7.24) примет вид $\exp(ikR)\exp(-i\mathbf{k} \cdot \mathbf{\rho}_{mn})$, где учтено, что угол между векторами \mathbf{R} и $\mathbf{\rho}_{mn}$, как видно на рис. 7.16, б равен углу между векторами \mathbf{k} и $\mathbf{\rho}_{mn}$. Подставляя в (7.24) выражение для $E(\mathbf{\rho}_{mn})$ и положив в знаменателе $r = R$, получим

$$E_{mn}(P) = C \frac{E_0 \exp(ikR)}{R} \exp(-i\Delta\mathbf{k} \cdot \mathbf{\rho}_{mn}), \quad (7.25)$$

где $\Delta \mathbf{k} = \mathbf{k} - \mathbf{k}_0$ – так называемый вектор рассеяния. Амплитуду поля в точке P от всей решетки получим, суммируя выражения (7.25) по всем отверстиям решетки. Эта величина будет пропорциональна сумме фазовых множителей

$$E(P) \sim \sum_{m,n} \exp(-i\Delta \mathbf{k} \cdot \boldsymbol{\rho}_{mn}). \quad (7.26)$$

Амплитуда в точке P (в направлении вектора \mathbf{k} рассеянной волны) имеет максимальное значение, если каждое слагаемое суммы (7.26) равно единице, что возможно, если

$$\Delta \mathbf{k} \cdot \boldsymbol{\rho}_{mn} = \Delta \mathbf{k} \cdot (m\mathbf{a} + n\mathbf{b}) = 2\pi \cdot (\text{целое число}). \quad (7.27)$$

Условие (7.27) будет выполнено, если одновременно будут выполняться соотношения

$$\Delta \mathbf{k} \cdot \mathbf{a} = 2\pi h, \quad \Delta \mathbf{k} \cdot \mathbf{b} = 2\pi p \quad (7.28)$$

при любых целых h и p . Соотношения (7.28) называют уравнениями дифракции Лауэ.

Представим вектор рассеяния $\Delta \mathbf{k}$ в виде

$$\Delta \mathbf{k} = h\mathbf{a}^* + p\mathbf{b}^*,$$

где \mathbf{a}^* и \mathbf{b}^* – подлежащие определению векторы. Подставив это выражение в уравнения Лауэ, видим, что $\Delta \mathbf{k}$ является решением этих уравнений, если выполняются следующие условия:

$$\mathbf{a}^* \cdot \mathbf{a} = 2\pi, \quad \mathbf{a}^* \cdot \mathbf{b} = 0, \quad \mathbf{b}^* \cdot \mathbf{a} = 0, \quad \mathbf{b}^* \cdot \mathbf{b} = 2\pi. \quad (7.29)$$

Отсюда следует, что вектор \mathbf{a}^* перпендикулярен вектору \mathbf{b} , а вектор \mathbf{b}^* перпендикулярен вектору \mathbf{a} . Но поскольку в рассматриваемой прямоугольной решетке векторы \mathbf{a} и \mathbf{b} взаимно перпендикулярны, то векторы \mathbf{a}^* и \mathbf{a} , \mathbf{b}^* и \mathbf{b} будут сонаправлены. Поэтому для прямоугольной сетки

$$\mathbf{a}^* = 2\pi/a, \quad \mathbf{b}^* = 2\pi/b. \quad (7.30)$$

Вектор $\Delta \mathbf{k}$ определяет положение точки в так называемом обратном пространстве, или пространстве Фурье. Название «обратное пространство» связано с тем, что вектор $\Delta \mathbf{k}$ имеет размерность обратной длины, что видно из соотношений (7.30). Переби-

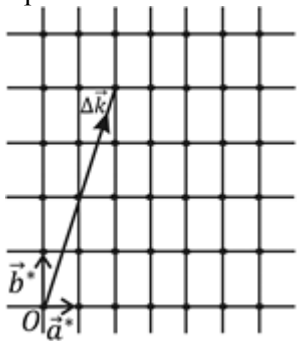


Рис. 7.17

рая различные целые значения h и p , получим двумерную периодическую структуру, называемую обратной решеткой (рис. 7.17). Величины a^* и b^* , определяемые соотношениями (7.30), являются периодами обратной решетки. Обратная решетка представляет собой дифракционную картину, создаваемую двумерной дифракционной решеткой. На опыте дифракционная картина наблюдается в виде системы световых пятен,

распологающихся в узлах решетки (рис. 7.16, б), периоды которой равны a^* и b^* . Эти пятна представляют собой максимумы дифракционной картины и называются пятнами Лауэ.

Интересно отметить, что дифракционная решетка из точечных отверстий, расположенных, как пятна на рис. 7.16, б дает дифракционную картину, подобную изображенной на рис. 7.16, а (концепция взаимности).

7.6. Трехмерные решетки

Трехмерные пространственные решетки обладают периодичностью в трех различных направлениях в пространстве. Естественными пространственными решетками являются кристаллы. Период кристаллических решеток обычно составляет от 1 нм до 10 нм. Световые волны видимого диапазона на такой решетке дифрагировать не могут, так как для таких волн условие дифракции на решетке, $d \geq \lambda$, не выполняется. Это условие

выполняется для рентгеновских лучей, которые имеют примерно такой же порядок длины волны, что и период кристаллической решетки. Кроме того, электронные оболочки атомов рассеивают рентгеновские лучи, и в этом смысле атомы ведут себя подобно точечным отверстиям в непрозрачном экране. Следовательно, кристаллы могут служить трехмерными дифракционными решетками для рентгеновских лучей.

Все рассуждения относительно дифракции на двумерных решетках остаются в силе и для трехмерных решеток. Однако поскольку в трехмерной решетке положение узла (рассеивающего элемента) определяется вектором

$$\mathbf{r}_{mnl} = m\mathbf{a} + n\mathbf{b} + l\mathbf{c},$$

где \mathbf{c} – направляющий вектор вдоль оси Z (c – период решетки в направлении оси Z), l – целое число, то в выражениях (7.24) – (7.27) вектор \mathbf{r}_{mn} следует заменить на вектор \mathbf{r}_{mnl} , а суммирование по m и n в (7.26) – суммированием по m , n и l , и тогда вместо двух условий дифракции получим три условия (три уравнения Лауэ):

$$\Delta\mathbf{k} \cdot \mathbf{a} = 2\pi h, \quad \Delta\mathbf{k} \cdot \mathbf{b} = 2\pi p, \quad \Delta\mathbf{k} \cdot \mathbf{c} = 2\pi q, \quad (7.31)$$

где q – целое число.

Вектор рассеяния $\Delta\mathbf{k}$ в трехмерном случае представляется в виде

$$\Delta\mathbf{k} = h\mathbf{a}^* + p\mathbf{b}^* + q\mathbf{c}^*,$$

где векторы \mathbf{a}^* , \mathbf{b}^* , \mathbf{c}^* находятся из уравнений (7.31). Этот вектор определяет положение узла в трехмерной обратной решетке, периоды которой равны a^* , b^* , c^* . Дифракционная картина будет представлять собой систему пятен Лауэ, образующих трехмерную обратную решетку с периодами a^* , b^* , c^* . Эта обратная трехмерная решетка взаимосвязана со структурой кристалла, как и в случае двумерных решеток. На практике обычно

дифракционную картину получают на фотопластинке, располагающейся на пути дифрагировавшего (на атомах кристалла) пучка. На фотопластинке после ее проявления будет наблюдаться сложная, но симметричная картина световых пятен Лауэ. Эта картина представляет собой двумерное изображение (проекцию на плоскость) трехмерной дифракционной картины.

Другой подход к рассмотрению дифракции рентгеновских лучей на кристаллах был предложен Ю.В. Вульфом и Л. Брэггом.

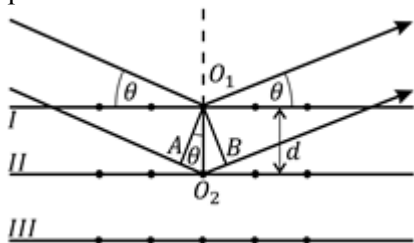


Рис. 7.18

Следуя Вульфу и Брэггу, предположим, что волновой фронт рентгеновских лучей падает на систему параллельных равноотстоящих друг от друга атомных плоскостей (плоскостей, в которых атомы упакованы наиболее плотно (рис. 7.18)). Каждый

атом становится источником рассеянного рентгеновского излучения, формируя отраженную от плоскости волну. Если поглощение излучения кристаллом незначительно, падающая волна будет отражаться от многих атомных плоскостей. Для лучей, отраженных от соседних плоскостей, разность хода, как это видно на рис. 7.18, равна $\Delta = AB = 2d \sin \theta$, где θ – угол, отсчитываемый от атомной плоскости (он называется углом скольжения), d – межплоскостное расстояние. Излучение, отраженное от соседних атомных плоскостей, будет в результате интерференции усиливаться, если эта разность хода окажется равной целому числу m длин волн λ :

$$2d \sin \theta = m\lambda. \quad (7.32)$$

Это соотношение называют условием Вульфа – Брэгга.

Таким образом, хотя отражение от каждой плоскости происходит зеркально, только при определенных углах θ отраженные от всех параллельных плоскостей лучи складываются синфазно, что и приводит к появлению сильного отраженного (дифрагированного) пучка. Если бы каждая плоскость полностью отражала падающее излучение, то только первая плоскость из семейства параллельных плоскостей участвовала бы в формировании отраженного от кристалла пучка и зеркальное отражение происходило бы при всех длинах волн.

Формула Вульфа – Брэгга является следствием периодичности пространственной решетки. При этом, как следует из соотношения (7.32), дифракция имеет место только для таких волн, длина волны которых удовлетворяет условию $\lambda \leq 2d$.

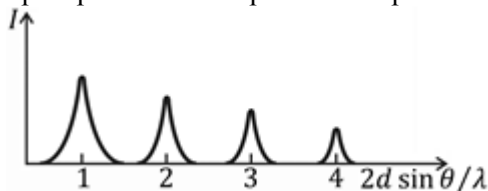


Рис. 7.19

Системы плоскостей, соответствующие $m = 1, 2, 3, \dots$, определяют брэгговское отражение соответственно первого, второго, третьего и т.д. порядков.

Можно показать, что формула Вульфа – Брэгга является следствием уравнений Лауэ.

Интенсивность дифрагированного пучка как функция длины волны при фиксированном угле θ имеет ряд дифракционных максимумов, показанных на рис. 7.19. Эти максимумы и образуют пятна Лауэ. Как видно на рис. 7.19, интенсивность в максимумах убывает с ростом отношения $2d \sin \theta / \lambda$, т.е. с увеличением порядка отражения m .

Интенсивность дифрагированного пучка как функция длины волны при фиксированном угле θ имеет ряд дифракционных максимумов, показанных на рис. 7.19. Эти максимумы и образуют пятна Лауэ. Как видно на рис. 7.19, интенсивность в максимумах убывает с ростом отношения $2d \sin \theta / \lambda$, т.е. с увеличением порядка отражения m .

Формула Вульфа – Брэгга лежит в основе рентгеновской спектроскопии и рентгеноструктурного анализа твердых тел. Если рентгеновское излучение неизвестной длины волны заставить отражаться от кристалла с известной структурой, то, измерив угол скольжения θ , при котором наблюдается первый дифракционный

максимум, по формуле (7.32) можно определить длину волны λ . С другой стороны, зная длину волны падающего на исследуемый кристалл излучения и измерив угол скольжения θ , соответствующий первому дифракционному максимуму, по той же формуле (7.32) можно определить межплоскостное расстояние d в кристалле. В настоящее время разработан ряд методов исследования кристаллической структуры, базирующихся на применении формулы (7.32).

Следует отметить, что между принципами построения изображения кристаллической структуры рентгеновскими лучами и формированием обычного оптического изображения существует близкая аналогия. Использование результатов дифракции рентгеновских лучей для определения расположения атомов в кристаллах оказывается, по существу, приложением теории формирования оптического изображения Аббе – Портера.

В заключение следует отметить, что одним из важнейших результатов проведенного в этой и двух предыдущих главах рассмотрения дифракции Френеля и дифракции Фраунгофера на различных апертурах является то, что явление дифракции приводит к искажениям изображения. Подробно этот вопрос будет рассмотрен в главе 11.

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

1. Аблеков В.К., Зубков П.И., Фролов А.В. Оптическая и оптоэлектронная обработка информации. –М.: Машиностроение, 1976.
2. Аблеков В.К., Колядин С.А., Фролов А.В. Высокора разрешающие оптические системы. –М.: Машиностроение, 1985.
3. Айзинов М.М. Радиотехнические цепи и сигналы. –М.: Транспорт, 1966.
4. Ахиезер А.И., Ахиезер И.А. Электромагнетизм и электромагнитные волны. –М.: Высшая школа, 1985.
5. Баскаков С.И. Радиотехнические цепи и сигналы. –М.: Высшая школа, 1988.
6. Борн М., Вольф Э. Основы оптики. –М.: Мир, 973.
7. Бутиков Е.И. Оптика. СПб., 1203.
8. Волосов Д.С. Фотографическая оптика. –М.: Искусство, 1978.
9. Гершензон Е.М., Малов Н.Н., Мансуров А.Н. Курс общей физики. Оптика и атомная физика. – М.: Просвещение, 1992.
10. Гоноровский И.С., Демин М.П. Радиотехнические цепи и сигналы. –М.: Радио и связь, 1994.
11. Гребенников О.Ф. Основы записи и воспроизведения изображения. –М.: Искусство, 1982.
12. Гребенников О.Ф., Тихомирова Г.В. Основы записи и воспроизведения информации. –СПб.: СПбГУКиТ, 2005.
13. Дитчберн Р. Физическая оптика. –М.: Наука, 1965.
14. Дьяков В.А., Тарасов Л.В., Щукин И.В. Введение в радиооптику. –М.: МИЭМ, 1975.
15. Гудмен Дж. Введение в фурье-оптику. –М.: Мир, 1970.
16. Гудмен Дж. Статистическая оптика. –М.: Мир, 1988.
17. Зверев В.А. Радиооптика. Преобразование сигналов в радиотехнике и оптике. –М.: Сов. Радио, 1975.
18. Калитиевский Н.И. Волновая оптика. –М.: Высшая школа, 1978.

19. Кингсеп А.С., Локшин Г.Р., Ольхов О.А. Основы физики. Т. 1. -М.: Физматлит, 2001. – 560 с.
20. Козане А., Флере Ж. и др. Оптика и связь. –М.: Мир, 1984.
21. Козел С.М., Локшин Г.Р. Волновая оптика. -М.: МФТИ, 1976.
22. Литвиненко О.Н. Основы радиооптики. –Киев: Техника, 1974.
23. Макс Ж. Методы и техника обработки сигналов. –М.: Мир, 1983.
24. Микаэлян А.Л. Оптические методы в информатике. –М.: Наука, 1990.
25. Москалев В.А., Нагибина Н.М., Полушкина Н.А., Рудин В.Л. Прикладная физическая оптика. –СПб.: Политехника, 1995.
26. Папулис А. Теория систем и преобразований в оптике. –М.: Мир, 1971.
27. Применение методов фурье-оптики// под ред. Г. Старка. –М.: Радио и связь, 1988.
28. Сайкин А.С., Морозова Н.Н. Фурье-анализ в оптике. – Чебоксары: ЧГУ, 1997.
29. Сивухин Д.В. Курс общей физики. Оптика. –М.: Наука, 1980.
30. Соколов В.В. Физическая оптика. –Томск: ТГУ, 1989.
31. Стюард И.Г. Введение в фурье-оптику. –М.: Мир, 1985.
32. Троицкий И.Н. Когерентная оптика и голография. –М.: МФТИ, 1982.
33. Фризер Х. Фотографическая регистрация информации. –М.: Мир, 1978.
34. Черемисин А.И. Дифракционные явления: Учебное пособие. – СПб.: СПбГУКиТ, 1997.
35. Черемисин А.И. Волновые свойства излучения: Учебное пособие. –СПб.: СПбГУКиТ, 1988
36. Юу Ф.Т.С. Введение в теорию дифракции, обработку информации и голографию. –М.: Сов. радио, 1979.

Министерство культуры Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ КИНО И ТЕЛЕВИДЕНИЯ»

А.И. ЧЕРЕМИСИН

Ф У Р Ь Е – О П Т И К А

Часть 2

Формирование оптического изображения

Учебное пособие для студентов технических
факультетов, обучающихся по специальностям
210400 и 210100

Санкт-Петербург
2012

УДК 535

ББК 22.34

Черемисин А.И. Фурье-оптика, Ч. 2. Формирование
оптического изображения. Учебное пособие. – СПб.: Изд.
СПбГУКиТ, 2012. – 329 с.

Во второй части учебного пособия «Фурье-оптика» дано изложение вопросов применения методов фурье-анализа в прикладной оптике (дифракционная теория оптического изображения, оптическая обработка информации, фильтрация, методы голографии и их применения). Рассмотрены все используемые в оптике линейные интегральные преобразования и такие основные линейные системы, как слой свободного пространства и тонкая линза.

Учебное пособие рекомендовано студентам, изучающим дисциплины «Основы записи и воспроизведения изображения», «Линейные интегральные преобразования в оптике», прикладную оптику и светотехнику, радиотехнику и телевидение, а также аспирантам и специалистам, использующим методы Фурье-оптики.

Рецензент д.ф.-м.н., профессор С.В. Ульянов (СПбТЭИ)

Рекомендовано к изданию в качестве учебного пособия
Советом факультета мультимедиа аппаратуры
Протокол № 12 от 21 марта .2012 г.

© Санкт-Петербургский государственный университет
кино и телевидения, 2012

ПРЕДИСЛОВИЕ

Во второй части учебного пособия «Фурье-оптика» изучаются вопросы линейных оптических систем, формирования оптического изображения, передаточные функции и функции рассеяния оптических систем, вопросы голографии. Вторая часть пособия включает в себя 6 глав – с 8 по 13.

В восьмой, девятой и десятой главах рассматриваются общие представления о линейных интегральных преобразованиях и их реализации в оптике. Излагаются широко используемые в оптике преобразования: Фурье, Фурье – Бесселя, Френеля, Абеля, Гильберта, Дирака и преобразования отсчетов. Здесь же вводится понятие широко используемой при описании формирования оптического изображения операции свертки. Вводится понятие линейной оптической системы, ее импульсной характеристики и передаточной функции. Рассматриваются импульсная характеристика и передаточная функция слоя свободного пространства, являющегося простейшей оптической системой и в то же время одним из основных элементов оптической системы формирования изображения. Рассматривается линейность оптической системы при различных видах освещения. Вводятся основные понятия геометрической теории оптического изображения, центрированных оптических систем формирования изображения, а также понятия функции пропускания тонкой линзы и ее частотной характеристики. Рассматриваются фурье-преобразующее и фокусирующее свойства тонкой линзы.

Одиннадцатая глава посвящена формированию оптического изображения. Рассматриваются изображения точечного и протяженного предметов, даваемых элементарной оптической системой. Показано, что образование изображения при когерентном освещении представляет собой двойную дифракцию Фраунгофера (двойное преобразование Фурье). Рассматривается влияние конечных размеров линзы на изображение и спектр пространственных частот. Рассматривается формирование

оптического изображения по Аббе – Портеру, формирование изображения при некогерентном и частично-когерентном освещении. Проводится сравнение формирования изображения при когерентном и некогерентном способах освещения.

В двенадцатой главе вводятся понятия оптических передаточных функций и функций рассеяния оптических систем, подробно рассматриваются эти функции для дифракционно-ограниченных оптических систем. Здесь же рассмотрена разрешающая способность оптических систем, оптические характеристики фоторегистрирующих материалов (коэффициент пропускания по амплитуде и интенсивности, оптическая плотность, экспозиция, частотно-контрастная характеристика и др.) и оптическая обработка пространственной информации (пространственная фильтрация, восстановление искаженных изображений).

Последняя, тринадцатая глава посвящена некоторым важным вопросам голографии. Рассматриваются физические принципы голографии, голография простейших объектов, голографическая схема Денисюка, геометрия формирования голограмм и восстановления изображений, голография Фурье с использованием линз и безлинзовая, применение голографии для обработки пространственной информации.

Пособие предназначено для студентов, изучающих прикладную оптику и светотехнику, радиотехнику, звукотехнику, а так же для магистрантов и аспирантов, всех, кто использует методы фурье-оптики. Материал, изложенный в пособии, может служить базой для изучаемых в СПбГУКиТ дисциплин «Линейные интегральные преобразования в оптике» и «Основы записи и воспроизведения изображения (ОЗВИ)».

Автор выражает благодарность студентам СПбГУКиТ Куксову В., Михайлову Л., Морозову А., Сморжевскому А. за выполненные рисунки, а также доценту кафедры светотехники Н.Н. Тульевой за интерес к работе и полезные советы и редактору издательства СПбГУКиТ Н.Н. Калининой за большой труд по выявлению и исправлению допущенных автором погрешностей.

При этом автор отмечает высокий профессионализм Н.Н. Калининой.

Глава 8

ЛИНЕЙНЫЕ ИНТЕГРАЛЬНЫЕ ПРЕОБРАЗОВАНИЯ В ОПТИКЕ

8.1. Общие представления о линейных интегральных преобразованиях и их применении в оптике

Интегральным преобразованием функции одной переменной $f(x)$ (одномерным интегральным преобразованием) называют функцию $F(\xi)$, определяемую интегралом вида

$$F(\xi) = \int_a^b f(x)K(x, \xi)dx ,$$

где $K(x, \xi)$ – некоторая функция x и ξ , называемая ядром одномерного преобразования. Аналогично можно определить интегральное преобразование и функции двух переменных (двумерное интегральное преобразование):

$$F(\xi, \eta) = \int_a^b \int_c^d f(x, y)K(x, y; \xi, \eta)dxdy ,$$

где $K(x, y; \xi, \eta)$ – ядро двумерного преобразования.

Все линейные интегральные преобразования могут рассматриваться как линейные операторы, действующие на функцию $f(x)$ (или на функцию $f(x, y)$) с целью получения

функции $F(\xi)$ (или функции $F(\xi, \eta)$). Оператор L называется линейным, если он обладает следующим свойством:

$$L\{c_1 f_1 + c_2 f_2\} = c_1 L\{f_1\} + c_2 L\{f_2\},$$

где f_1 и f_2 – две произвольные функции, c_1 и c_2 – произвольные постоянные.

Интегральные преобразования играют большую роль в оптике. При этом функции $f(x, y)$ и $F(x, y)$ обычно описывают оптические сигналы соответственно во входной и выходной плоскостях. При помощи оптических систем можно получить любое интегральное преобразование. Применяя различные комбинации оптических элементов, можно задать функцию отклика оптической системы, являющуюся ядром интегрального преобразования $K(x, y; \xi, \eta)$. Интегральные преобразования позволяют получить результаты более просто и с более высокой точностью даже в тех случаях, когда систему можно адекватно исследовать прямым методом. Все используемые в оптике интегральные преобразования являются линейными. Кроме того, в оптике используются только такие интегральные преобразования, которые имеют обратное преобразование вида

$$f(x) = \int_{a_1}^{b_1} F(\xi) G(x, \xi) d\xi$$

или, в случае функции двух переменных,

$$f(x, y) = \int_{a_1}^{b_1} \int_{c_1}^{d_1} F(\xi, \eta) G(x, y; \xi, \eta) d\xi d\eta,$$

где $G(x, \xi)$ и $G(x, y; \xi, \eta)$ – ядра обратного одномерного и двумерного преобразований соответственно. В некоторых случаях ядра прямого и обратного преобразований могут быть одинаковыми. Тогда функция и ее образ будут связаны симметричными соотношениями.

В оптике находят применение следующие интегральные преобразования: преобразование Фурье, преобразование Фурье – Бесселя, преобразование свертки, корреляция, преобразование Френеля, преобразование Абеля, преобразование Гильберта, преобразование Мелина, преобразование Дирака и преобразование отсчетов. Они отвечают вполне определенным физическим процессам, происходящим в оптико-информационных системах, в частности, в когерентных оптических системах.

Ниже рассматриваются все перечисленные преобразования и их реализация в когерентной оптике.

8.2. Преобразование Фурье

Преобразование Фурье переводит сигнал из пространства координат в пространство частот (фурье-пространство). Это преобразование и его применение к определению частотного (временного и пространственного) спектра (фурье-спектра) светового излучения и к описанию распределения амплитуды светового поля в фраунгоферовской дифракционной картине рассматривалось нами в главах 2, 4, 6, 7. В п. 10.5 мы покажем, что преобразование Фурье оптического сигнала может производить простая собирающая линза. В главах 12 и 13 будет рассматриваться применение преобразования Фурье в оптической фильтрации и в голографии Фурье.

В общем случае в преобразованиях Фурье $E(x, y)$ – любая непериодическая функция $f(x, y)$. Тогда это преобразование будет иметь вид

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-i2\pi(ux + vy)} dx dy,$$

где $F(u, v)$ – фурье-образ функции $f(x, y)$. Обратное преобразование

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) e^{i2\pi(ux + vy)} dudv.$$

Функция $f(x, y)$ при этом должна удовлетворять некоторым условиям (называемыми условиями Дирихле). В частности, она должна быть абсолютно интегрируема, т.е. интеграл по бесконечным пределам от функции $|f(x, y)|$ должен быть конечной величиной:

$$\int_{-\infty}^{\infty} |f(x, y)| dx dy < \infty,$$

где x, y – координаты в обычном физическом пространстве. Переменные u и v в преобразованиях Фурье представляют собой пространственные частоты. Они являются координатами в так называемом обратном или фурье-пространстве. Функцию $f(x, y)$ и ее фурье-спектр $F(u, v)$ называют парой преобразования Фурье. Ядром преобразования Фурье служит комплексная экспонента $\exp[-i2\pi(ux + vy)]$, а ядром обратного преобразования – комплексная экспонента $\exp[i2\pi(ux + vy)]$ (ядром временного преобразования Фурье является функция $\exp(i\omega t)$, а временного обратного преобразования – функция $\exp(-i\omega t)$). Различие в знаках, содержащихся в экспонентах и отличает прямое преобразование Фурье от обратного (инверсного) преобразования.

В соответствии с определениями, данными выше, применение прямого, а затем обратного преобразований Фурье дает исходную функцию:

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

Следовательно, обратное преобразование Фурье позволяет восстановить исходную функцию по ее фурье-образу. Это свойство фурье-преобразования называют его обратимостью. В преобразовании Фурье исходная функция представляется в виде

бесконечной суммы гармоник последовательно возрастающих частот. При этом фурье-образ исходной функции значительно отличается от нее самой, хотя ядра прямого и обратного преобразований отличаются только знаком перед i в показателях экспонент. Однако преобразование Фурье изображает исходную функцию однозначно, так что при повторном преобразовании Фурье исходная функция полностью восстанавливается.

Покажем теперь, что двойное преобразование Фурье возвращает исходную функцию с инверсными переменными, т.е. с заменой x на $-x$ и y на $-y$:

$$F\{F\{f(x, y)\}\} = f(-x, -y).$$

Имеем

$$\begin{aligned} F\{F\{f(x, y)\}\} &= F\{F(u, v)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) e^{-i2\pi(ux + vy)} dudv = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) e^{i2\pi[u(-x) + v(-y)]} dudv = f(-x, -y). \end{aligned}$$

Как реализуется двойное преобразование Фурье в оптике, будет показано в п. 10.5.

Операция преобразования Фурье, связывающая координатную и частотную области, отражает физическую сущность действия оптических систем, поэтому это преобразование является наиболее широко используемым в оптике математическим аппаратом. Обратное преобразование Фурье можно рассматривать как разложение сложной световой волны на множество плоских волн, направляющие косинусы которых определяются пространственными частотами волн ($\cos \alpha = \lambda u$, $\cos \beta = \lambda v$). Оно применяется не только для описания поля при дифракции Фраунгофера, но и в когерентной оптической обработке информации и всюду, где требуется частотный анализ, фильтрация, корреляция и распознавание сигналов. При определенных условиях свойства когерентной оптической

системы естественным образом описываются оператором фурье-образа, что в общем случае представляет собой двумерное преобразование Фурье.

Часто при анализе или синтезе линейных оптических систем с целью применения математического аппарата Фурье оптический сигнал рассматривают как распределение поля в плоскости XU с комплексной амплитудой, зависящей как от координат x и y , так и от времени t , т.е. считают, что сигнал имеет пространственно-временную природу. Такой сигнал (его иногда называют радиооптическим сигналом) должен описываться комплексной функцией $f(x, y, t)$. Если использовать ту же процедуру, которая применяется к временным и пространственным сигналам, то для пространственно-временного сигнала формулу преобразования Фурье следует использовать в виде

$$F(u, v, \nu) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y, t) \exp[-i2\pi(ux + vy - \nu t)] dx dy dt,$$

а обратное преобразование Фурье – в виде

$$f(x, y, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v, \nu) \exp[i2\pi(ux + vy - \nu t)] du dv d\nu,$$

где u, v – пространственные частоты, ν – временная частота. Как видим, в случае пространственно-временного сигнала речь идет о пространственно-временном частотном спектре.

Если комплексную амплитуду пространственно-временного сигнала можно представить в виде произведения двух функций $f(x, y, t) = f_1(x, y) \cdot f_2(t)$, то переменные в подынтегральном выражении разделяются и все операции фурье-анализа можно проводить раздельно, осуществив сначала временной фурье-анализ, а затем пространственный, или в обратной последовательности.

8.3. Преобразование Фурье – Бесселя

Это преобразование применимо к сигналам, обладающим круговой симметрией. Функции, выражающие такие сигналы, записываются в виде $f(x, y) = f(r)$, где $r = \sqrt{x^2 + y^2}$ – полярный радиус. Оптические системы должны при этом содержать только такие элементы, двумерная функция пропускания которых сводится к функции, зависящей от одной переменной r . К таким элементам относятся круговые диафрагмы, круглые оправы линз, зонные пластинки Френеля и т.д.

Можно показать, что образы двумерных распределений, являющихся функциями только полярного радиуса r , имеют также круговую симметрию (и, следовательно, представляют собой функции только радиальной пространственной частоты $\rho = \sqrt{u^2 + v^2}$).

Преобразование Фурье – Бесселя имеет вид

$$F(\rho) = 2\pi \int_0^{\infty} f(r) J_0(2\pi r\rho) r dr, \quad (8.1)$$

где J_0 – функция Бесселя первого рода нулевого порядка. Соответствующее обратное преобразование записывается в виде

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

Видим, что преобразование Фурье – Бесселя является одномерным интегральным преобразованием с одинаковыми ядрами $2\pi J_0(2\pi r\rho)$ прямого и обратного преобразований.

Преобразование Фурье – Бесселя, называемое также преобразованием Ганкеля нулевого порядка, получается из двумерного преобразования Фурье при переходе в нем к полярной системе координат, как в координатной, так и в частотной областях:

$$x = r \cos \theta, y = r \sin \theta, u = \rho \cos \varphi, \\ v = \rho \sin \varphi, dx dy = r dr d\theta.$$

Подставив эти соотношения в преобразование Фурье, получим

$$F(\rho) = \int_0^{\infty} r f(r) \left(\int_0^{2\pi} \exp[-ir\rho(\cos \theta \cos \varphi + \sin \theta \sin \varphi)] d\theta \right) dr = \\ = \int_0^{\infty} r f(r) \left(\int_0^{2\pi} \exp[-ir\rho \cos(\theta - \varphi)] d\theta \right) dr.$$

Учитывая определение функции Бесселя первого рода нулевого порядка

$$J_0(z) = \frac{1}{2\pi} \int_0^{2\pi} \exp[-iz \cos(\theta - \varphi)] d\theta,$$

придем к выражению (8.1).

С преобразованием Ганкеля мы уже встречались при рассмотрении дифракции Фраунгофера на круглом отверстии (интеграл (6.7)), апертурная функция $t(x, y)$ которого обладает круговой симметрией: $t(x, y) = t(r) = \text{circ}(r/R)$, где R – радиус отверстия.

8.4. Преобразование свертки

Понятие операции свертки имеет большое значение в оптике. Рассмотрим эту операцию сначала на примере функций одной переменной. Сверткой $s(x)$ двух функций $f(x)$ и $g(x)$ называют интеграл вида

$$s(x) = \int_{-\infty}^{\infty} f(\xi) g(x - \xi) d\xi, \quad (8.2)$$

где ξ является вспомогательной переменной интегрирования, x представляет ряд последовательных значений сдвига функции $g(x)$ относительно функции $f(x)$. Интеграл от произведения перекрывающихся частей функций $g(x)$ и $f(x)$ определяется при одном таком сдвиге x . Функция $s(x)$ получается путем последовательного определения интеграла от произведения двух функций $f(x)$ и $g(x)$, когда эти две функции последовательно сдвигаются относительно друг друга. Значение интеграла находится для каждого сдвига x функции $g(x)$ относительно $f(x)$, включая значение $f(0)$, которое соответствует значению $x = 0$, когда эти две функции $f(x)$ и $g(x)$ «совпадают», т.е. имеют одинаковую абсциссу.

Кратко свертку записывают в виде $s(x) = f(x) * g(x)$ или просто $s(x) = f * g$. Как видим, эта операция предполагает умножение ординаты функции $f(x)$ при каждом значении ξ целиком на другую функцию $g(x)$, но смещенную по оси $x > 0$ на отрезок ξ , т.е. умножение на функцию $g(x - \xi)$, и суммирование результатов, полученных при всех возможных значениях ξ .

Таким образом, для получения функции $s(x)$ следует построить зеркальное отражение функции $g(\xi)$ относительно оси ординат. Полученную функцию $g(-\xi)$ надо сдвинуть вправо вдоль оси X

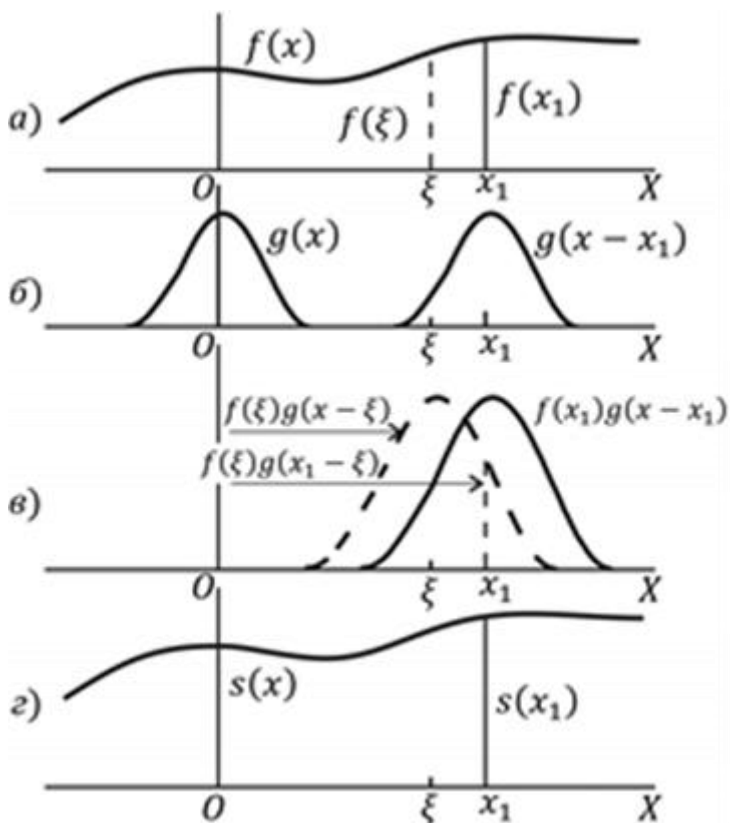


Рис. 8.1

на отрезок x , произвести поточечное умножение сдвинутой функции $g(x - \xi)$ на $f(\xi)$ и вычислить площадь под кривой $f(\xi)g(x - \xi)$. В результате мы получим значение функции $s(x)$ в одной конкретной точке x . Повторяя указанные действия для различных значений сдвига x , можно построить функцию $s(x)$ в соответствии с формулой (8.2).

Интегрирование в правой части соотношения (8.2) распространено на бесконечные пределы, так как только

относительно близкие к точке x точки будут оказывать влияние на величину $s(x)$, ибо кривая $g(x)$ довольно узкая

Функция $g(x)$ должна быть инвариантной относительно сдвига. Это означает, что форма кривой $g(x)$ не должна изменяться при смещении ее вдоль оси X на любое значение $x = \xi$. Это означает что кривые $g(x)$ и $g(x - \xi)$ должны иметь одинаковый вид при любом значении ξ (функции $g(x)$ и $g(x - \xi)$ должны иметь одинаковую зависимость от x). Поскольку функция $g(x)$ отлична от нуля только на ограниченном промежутке (a, b) , то интегрирование в (8.2) по бесконечным пределам следует заменить на интегрирование в пределах (a, b) .

Рассмотрим физическую сторону операции свертки (рис. 8.1). Ордината $f(x)$ кривой $f(x)$ при некотором значении x_1 (рис.8.1, *a*) после умножения на функцию $g(x - x_1)$ (рис. 8.1, *б*) размывается, становясь кривой $f(x_1)g(x - x_1)$ (сплошная кривая на рис. 8.1, *в*). Ордината этой кривой при $x = x_1$ дает основной вклад в значение функции $s(x)$ в точке x_1 . Имеются и другие вклады, которые порождаются размыванием других ординат кривой $f(x)$. Один из таких вкладов, возникающий из-за размывания ординаты кривой при $x = \xi$ и равный ординате кривой $f(\xi)g(x - \xi)$ в точке x_1 , т.е. величина $f(\xi)g(x_1 - \xi)$, представлен на рис. 8.1, *в* пунктирным вертикальным отрезком. Результирующее значение функции $s(x)$ в точке x_1 будет определяться суммой указанных вкладов, т.е. интегралом (8.2).

Проведенное рассмотрение показывает, что свертка двух функций представляет собой результат сканирования одной функции с помощью другой функции.

Аналогично определяется свертка функций двух переменных:

$$s(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi, \eta) g(x - \xi; y - \eta) d\xi d\eta. \quad (8.3)$$

Этот интеграл называют двумерной сверткой. Обратного преобразования свертка не имеет, и поэтому ее, строго говоря, нельзя считать интегральным преобразованием.

Из проведенного рассмотрения видно, что применение операции свертки к функции $f(x)$ приводит к изменению вида этой функции, а значит, и к искажению сигнала (например, к размытию изображения), описываемого этой функцией.

В теории линейных систем, в том числе и в оптике, широкое применение находит так называемая теорема свертки, согласно которой преобразование Фурье свертки двух функций равно произведению их собственных фурье-преобразований, т.е.

$$F\{s(x)\} = F\{f(x)*g(x)\} = F\{f(x)\} \cdot F\{g(x)\}$$

– в одномерном случае и

$$F\{s(x, y)\} = F\{f(x, y)*g(x, y)\} = F\{f(x, y)\} \cdot F\{g(x, y)\}$$

– в двумерном случае. Докажем эту теорему для одномерного случая. Запишем, используя определение преобразования Фурье:

$$F\{s(x)\} = \int_{-\infty}^{\infty} s(x) e^{-i2\pi x} dx.$$

Заменяя $s(x)$ сверткой функций $f(x)$ и $g(x)$ и изменяя затем порядок интегрирования, получим

$$\begin{aligned} F\{s(x)\} &= \int_{-\infty}^{\infty} e^{-i2\pi x} \left(\int_{-\infty}^{\infty} f(\xi) g(x - \xi) d\xi \right) dx = \\ &= \int_{-\infty}^{\infty} f(\xi) \left(\int_{-\infty}^{\infty} e^{-i2\pi x} g(x - \xi) dx \right) d\xi. \end{aligned}$$

Во внутреннем интеграле, в котором ξ является постоянной величиной, произведем замену переменной $x - \xi = \eta$. Будем иметь

$$F\{s(x)\} = \int_{-\infty}^{\infty} f(\xi) \left(\int_{-\infty}^{\infty} e^{-i2\pi u(\eta+\xi)} g(\eta) d\eta \right) d\xi. \quad (8.4)$$

Множитель $\exp(-i2\pi u\xi)$ является постоянным в пределах внутреннего интеграла, и поскольку остальная часть подынтегрального выражения внутреннего интеграла не содержит переменную ξ , она является постоянной для внешнего интегрирования по отношению к ξ . Таким образом, полное выражение (8.4) разделяется на произведение двух отдельных интегралов:

$$F\{s(x)\} = \int_{-\infty}^{\infty} f(\xi) e^{-i2\pi u\xi} d\xi \cdot \int_{-\infty}^{\infty} g(\eta) e^{-i2\pi u\eta} d\eta =$$

$$= F\{f(x)\} \cdot F\{g(x)\}.$$

Теорема свертки доказана.

Из теоремы свертки вытекает, что свертке в обычном (физическом) пространстве соответствует умножение в фурье-пространстве (в пространстве частот u , v). Этот очень важный вывод, как мы увидим ниже, позволяет не только наглядно объяснить процесс формирования оптического изображения, но и служит для получения так называемых передаточных функций оптической системы.

Справедлива и обратная теорема (называемая второй теоремой свертки, или теоремой Парсеваля): фурье-преобразование произведения двух функций равно свертке фурье-образов этих функций, т.е.

$$F\{f(x, y) \cdot g(x, y)\} = F\{f(x, y)\} * F\{g(x, y)\}.$$

Свертка обладает свойством коммутативности, согласно которому

$$s(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi, \eta) g(x - \xi; y - \eta) d\xi d\eta =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\xi, \eta) f(x - \xi; y - \eta) d\xi d\eta,$$

или в символическом виде

$$f(x, y) * g(x, y) = g(x, y) * f(x, y).$$

В одномерном случае

$$s(x) = \int_{-\infty}^{\infty} f(\xi) g(x - \xi) d\xi = \int_{-\infty}^{\infty} g(\xi) f(x_1 - \xi) d\xi,$$

или

$$f(x) * g(x) = g(x) * f(x).$$

Интегральная операция свертки широко используется при описании кратных процессов дифракции или последовательного наложения нескольких взаимно подобных процессов; связывает сигналы на входе и выходе оптической системы.

Линейные оптические системы по своей природе обладают двумя степенями свободы. Поэтому они могут осуществлять преобразование двух типов: преобразование

$$s(x, \eta) = \int_{-\infty}^{\infty} f(\xi, \eta) g(x - \xi; \eta) d\xi$$

и преобразование (8.3). Наибольшее значение свертка имеет в теории формирования и обработки оптического изображения. Одной из сворачиваемых функций ($f(x, y)$) здесь является распределение либо амплитуды, либо интенсивности поля на предмете, а второй ($g(x, y)$) – либо импульсная функция, либо сглаживающая функция (называемая часто размытием), либо

функция рассеяния (изображения) линии (ФРЛ, в одномерном случае) или функция рассеяния точки (ФРТ). Функция $s(x, y)$ в этом случае представляет собой распределение либо амплитуды, либо интенсивности поля на изображении. Термины «функция рассеяния», «размытие» отражают тот факт, что операция свертки, производимая оптической системой, приводит к размытию, нерезкости изображения.

Возможно образование свертки и более чем двух функций, которая также имеет физический смысл, в частности, в оптике (в интерферометрии, спектроскопии, голографии и т.д.).

Покажем, что апертурную функцию щелевой решетки можно представить сверткой апертурной функции одиночной щели и последовательности δ -функций, определяющей структуру решетки. В основе такого представления лежит сформулированная выше теорема свертки. Поскольку, как было показано в пп. 7.3.1

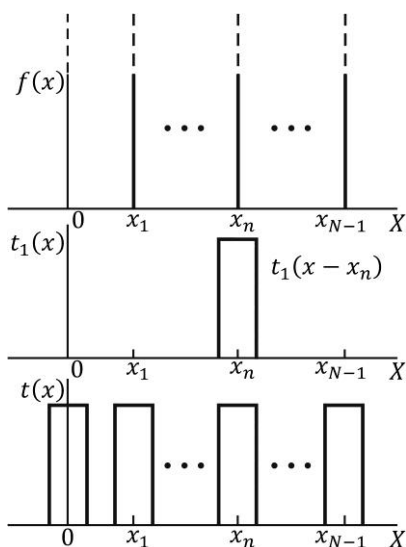


Рис. 8.2

апертурная функция решетки $t(x)$ может быть описана как свертка апертурной функции $t_1(x)$ одиночной щели с

и 7.3.3, распределение амплитуды светового поля в дифракционной картине решетки является преобразованием Фурье от ее апертурной функции $t(x)$, а, с другой стороны, это распределение представляет собой произведение фурье-преобразований апертурной функции $t_1(x)$ одиночной щели и последовательности δ -функций, определяющей структуру решетки, то на основании теоремы свертки приходим к сформулированному выше утверждению о том, что

последовательностью δ -функций, определяющей распределение этой щели в решетке, т.е. с функцией $f(x)$. На основании свойства коммутативности свертки функцию $t(x)$ можно представить и как свертку функции $f(x)$ с функцией $t_1(x)$:

$$t(x) = \int_{-\infty}^{\infty} f(\xi) t_1(x - \xi) d\xi.$$

Образование такой свертки показано на рис. 8.2. Обратим внимание на следующий факт. В примере на рис. 8.1 при образовании свертки происходит перекрытие ординат сворачиваемых функций; в случае же с решеткой такое перекрытие отсутствует.

8.5. Корреляция

Корреляция является широко распространенным методом обработки различных сигналов. Она находит применение и в оптике. С понятием корреляции мы уже встречались в четвертой главе при описании интерференции квазимонохроматического света и временной когерентности световых волн; функция корреляции используется также в общей теории голографических процессов. Корреляция является, по существу, методом оценки и определения взаимных связей, имеющих форму подобий или совпадений.

Пусть имеем одномерный, зависящий от времени детерминированный сигнал $f(t)$ конечной длительности. Интеграл

$$c(\tau) = \int_{-\infty}^{\infty} f(t) f(t - \tau) dt,$$

где τ – величина временного сдвига сигнала, называется автокорреляционной функцией сигнала $f(t)$. Она характеризует

степень связи (корреляции) сигнала $f(t)$ со своей копией, сдвинутой на величину τ по оси времени. Так как любой сигнал полностью коррелирован с самим собой, то ясно, что при $\tau = 0$ корреляционная функция достигает максимума. При этом величина

$$c(0) = \int_{-\infty}^{\infty} f^2(t) dt$$

определяет энергию сигнала. Как видим, максимальное значение автокорреляционной функции равно энергии сигнала.

С увеличением τ корреляционная функция убывает и при сдвиге τ , превышающем длительность сигнала, автокорреляционная функция обращается в нуль.

Заменяя время t на пространственную переменную x , получим автокорреляционную функцию одномерного пространственного сигнала $f(x)$:

$$c(a) = \int_{-\infty}^{\infty} f(x)f(x-a) dx, \quad (8.5)$$

где a – величина сдвига сигнала $f(x)$ вдоль положительного направления оси X . Сигнал $f(x)$ имеет отличные от нуля значения только в конечном интервале значений переменной x . Функция $c(a)$ имеет максимум при $a = 0$ и обращается в нуль при тех значениях a , при которых функции $f(x)$ и $f(x-a)$ перестают перекрываться. Произведение функций $f(x)$ и $f(x-a)$ отлично от нуля только в области их перекрытия. Интеграл (8.5) определяет площадь под графиком функции $f(x)f(x-a)$, т.е. площадь, ограниченную перекрытием при данной величине сдвига a . Каждому значению сдвига a соответствует свое произведение $f(x)f(x-a)$ и своя площадь под его графиком. Численные значения таких площадей

для соответствующих значений a и дают ординаты функции $c(a)$. Таким образом, значение $c(a)$ для любых выбранных значений a получается смещением функции относительно себя самой на величину a и определением площади перекрытия.

Для оценки степени связи между двумя различными сигналами $f(x)$ и $g(x)$ используется взаимная корреляционная (или кросскорреляционная) функция, определяемая как

$$c_{12}(a) = \int_{-\infty}^{\infty} f(x)g(x-a) dx.$$

Автокорреляционная функция $c(a)$ является частным случаем функции $c_{12}(a)$, когда сигналы $f(x)$ и $g(x)$ одинаковы.

Чтобы кросскорреляционную функцию можно было описать сдвигом a и в отрицательном направлении оси X , следует ее переписать следующим образом:

$$c_{12}(a) = \int_{-\infty}^{\infty} f(x)g(x+a) dx$$

и аналогично для автокорреляции

$$c(a) = \int_{-\infty}^{\infty} f(x)f(x+a) dx.$$

Обе корреляционные функции могут быть определены и для двумерных сигналов. В частности, кросскорреляционная функция двумерных сигналов $f(x, y)$ и $g(x, y)$

$$c_{12}(a, b) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)g(x-a, y-b) dx dy =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) g(x + a, y + b) dx dy .$$

Здесь a и b – смещения функции $g(x, y)$ вдоль координатных осей X и Y соответственно.

Для обозначения корреляции будем использовать символ \otimes . Тогда

$$c(a, b) = f(x, y) \otimes f(x, y), \quad c_{12}(a, b) = f(x, y) \otimes g(x, y).$$

Определение функции одномерной взаимной корреляции часто выражается в несколько иной форме, а именно:

$$c_{12}(x) = f(x) \otimes g(x) = \int_{-\infty}^{\infty} f(\xi) g(\xi - x) d\xi ,$$

и аналогично для автокорреляции. В этом выражении x является смещением, а ξ – вспомогательной переменной интегрирования. Отсюда видно, что корреляция подобна свертке, но отличается тем, что функция $g(x)$ сдвигается, а не переворачивается (перегибается). В двумерном случае

$$c_{12}(a, b) = f(x, y) \otimes g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi, \eta) g(\xi - x, \eta - y) d\xi d\eta.$$

Если функции $f(x)$ и $g(x)$ комплексные, то принято определять комплексную автокорреляционную функцию

$$c(x) = f(x) \otimes f^*(x) = \int_{-\infty}^{\infty} f(\xi) f^*(\xi - x) d\xi$$

и комплексную кросскорреляционную функцию

$$c_{12}(x) = f(x) \otimes g^*(x) = \int_{-\infty}^{\infty} f(\xi) g^*(\xi - x) d\xi,$$

где звездочка в обоих выражениях обозначает комплексное сопряжение. При этом если звездочка связана с другой функцией, то получается сопряженный результат. Отметим также, что

$$f(x) \otimes g^*(x) = f(x) * g^*(-x). \quad (8.6)$$

Часто бывает удобно нормировать корреляцию. Для этого корреляционную функцию делят на центральное значение корреляции, т.е. на ее значение при $x = 0$ (нулевой сдвиг). В итоге получаем нормированную автокорреляционную функцию

$$\gamma(x) = \left(\int_{-\infty}^{\infty} |f(\xi)|^2 d\xi \right)^{-1} \int_{-\infty}^{\infty} f(\xi) f^*(\xi - x) d\xi$$

и нормированную кросскорреляционную функцию

$$\gamma_{12}(x) = \left(\int_{-\infty}^{\infty} f(\xi) g^*(\xi) d\xi \right)^{-1} \int_{-\infty}^{\infty} f(\xi) g^*(\xi - x) d\xi.$$

С нормированной автокорреляцией мы встречались в четвертой главе.

8.6. Преобразование Френеля

Значение преобразования Френеля в оптике тесным образом связано с явлением дифракции света в ближней зоне (дифракции Френеля). Примером преобразования Френеля является зонная пластинка Френеля.

В своем основном виде двумерное преобразование Френеля определяется следующим образом:

$$\Phi(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi, \eta) \exp \left\{ i \frac{\pi S}{\lambda} [(x - \xi)^2 + (y - \eta)^2] \right\} d\xi d\eta, \quad (8.7)$$

где $f(x, y)$ – исходная функция, $\Phi(x, y)$ – ее френелевский образ. Мы видим, что двумерное преобразование Френеля представляет собой просто двумерную свертку функции $f(x, y)$ с экспоненциальной фазовой функцией

$$Z(x, y) = \exp \left[i \frac{\pi S}{\lambda} (x^2 + y^2) \right],$$

представляющей собой двумерную функцию Гаусса с мнимым комплексно сопряженным аргументом; ее называют функцией Френеля. Интегральное преобразование Френеля какой-либо функции $f(x, y)$ представляет собой свертку этой функции с функцией Френеля $Z(x, y)$. Его можно рассматривать и как свертку функции $f(x, y)$ с зонной пластинкой Френеля. Каждая точка области задания функции $f(x, y)$ размывается при этом в картину зон Френеля. Можно считать, что при преобразовании Френеля исходная функция представляет собой сумму большого числа последовательно смещенных френелевских картин. Величина каждой составляющей картины находится посредством умножения функции на соответствующую френелевскую картину и последующего интегрирования. Таким образом, при преобразовании Френеля каждое значение функции $f(x, y)$ в точке (x, y) размывается функцией Френеля, а затем все непрерывно смещенные картины Френеля суммируются.

Френелевский образ функции $f(x, y)$ можно найти и по формуле

$$\Phi(x, y) = f(x, y) * Z(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) Z(u, v) e^{i2\pi(ux+vy)} dudv,$$

где $F(u, v)$ – фурье-образ функции $f(x, y)$, $Z(u, v)$ – фурье-образ функции Френеля:

$$Z(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[i\pi s(x^2 + y^2)/\lambda] \exp[-i2\pi(ux + vy)] dx dy = \\ = \frac{i\lambda}{s} \exp[i\pi\lambda(u^2 + v^2)/s].$$

При получении этого результата была использована формула (5.24) для $\alpha = \pi s/i\lambda$, $\beta_1 = -i2\pi u$, $\beta_2 = -i2\pi v$. Как видим, фурье-образ функции Френеля в координатах пространственных частот $\sqrt{\pi\lambda/s} u$ и $\sqrt{\pi\lambda/s} v$ равен функции Френеля в координатах $\sqrt{\pi s/\lambda} x$ и $\sqrt{\pi s/\lambda} y$, умноженной на комплексный коэффициент $i\lambda/s = (\lambda/s) \exp(i\pi/2)$. Аналогичным свойством обладает лишь двумерная функция Гаусса $f(x, y) = \exp[-\alpha(x^2 + y^2)]$, фурье-образ которой равен самой функции Гаусса.

Преобразование Френеля, как и преобразование Фурье, дает однозначное представление функции, являясь операцией, допускающей восстановление исходной функции при повторном или обратном преобразовании Френеля. Обратное преобразование также имеет вид свертки

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(\xi, \eta) \exp \left\{ -i \frac{\pi s}{\lambda} [(x - \xi)^2 + (y - \eta)^2] \right\} d\xi d\eta.$$

В обратном преобразовании аргумент фазовой функции Френеля заменяется на комплексно сопряженный. Заметим, что при преобразовании Френеля, которое является сверткой, все операции производятся в обычном координатном пространстве. Этим оно качественно отличается от преобразования Фурье, в результате которого преобразуемая функция превращается в ее частотно-спектральный образ, заданный в пространстве фурье-координат, или пространственных частот. В преобразовании

Френеля параметр s в большинстве случаев интерпретируется как кривизна сферических волновых фронтов.

Сравнивая интеграл суперпозиции в приближении Френеля (5.20), записанный в виде

$$E'(x', y') = \frac{e^{ikz}}{i\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[i\pi \frac{(x' - x)^2 + (y' - y)^2}{\lambda z} \right] dx dy,$$

с выражением (8.5), видим, что этот интеграл представляет собой двумерное преобразование Френеля с ядром преобразования

$$h(x, y) = \frac{e^{ikz}}{i\lambda z} \exp \left(ik \frac{x^2 + y^2}{2z} \right),$$

являющимся, как мы увидим ниже, импульсным откликом слоя свободного пространства в приближении Френеля. Кривизна сферического волнового фронта в этом случае $s = 1/z$, где z – толщина слоя свободного пространства.

Таким образом, можно констатировать, что поле в ближней зоне при дифракции плоских волн можно определить с помощью двумерного преобразования Френеля. Если амплитуда E_0 падающей волны в плоскости апертуры постоянна, то согласно соотношению $E(x, y) = E_0 t(x, y)$ поле в плоскости $X'Y'$, расположенной в ближней зоне, можно определить с помощью преобразования Френеля от апертурной функции $t(x, y)$. Видим также, что в приближении Френеля интеграл суперпозиции представляет собой свертку функции распределения комплексной амплитуды поля в плоскости непосредственно за апертурой с импульсной характеристикой свободного пространства в приближении Френеля.

По известному распределению $E'(x', y')$ с помощью обратного преобразования Френеля

$$E(x, y) = \frac{e^{ikz}}{i\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E'(x', y') \exp \left[-ik \frac{(x' - x)^2 + (y' - y)^2}{2z} \right] dx' dy'$$

можно восстановить распределение поля в плоскости непосредственно за апертурой.

Между преобразованиями Френеля и Фурье существует тесная связь, обусловленная тем, что первое преобразование воспроизводит дифракцию Френеля, а второе – дифракцию Фраунгофера. Различие между обоими преобразованиями зависит только от расстояния, при котором проявляются оба типа дифракции. Связь между этими преобразованиями проще всего установить в одномерном случае. Используя преобразование Фурье, будем иметь

$$\begin{aligned} F(\xi) &= \int_{-\infty}^{\infty} f(x) e^{-i2\pi\xi x} dx = \int_{-\infty}^{\infty} f(x) e^{i\pi(\xi-x)^2} \cdot e^{-i\pi x^2} \cdot e^{-i\pi\xi^2} dx = \\ &= e^{-i\pi\xi^2} \int_{-\infty}^{\infty} f(x) e^{-i\pi x^2} e^{i\pi(\xi-x)^2} dx. \end{aligned}$$

Отсюда видно, что фурье-образ $F(\xi)$ функции $f(x)$ получается путем умножения функции $f(x)$ на функцию $\exp(-i\pi x^2)$, применения к полученному результату преобразования Френеля и последующего умножения на функцию $\exp(-i\pi\xi^2)$. Это соотношение позволяет представить преобразование Фурье с помощью операций умножения и свертки. При этом каждая из этих операций может быть осуществлено с помощью соответствующей оптической системы.

Преобразование Френеля играет важную роль не только при описании дифракции Френеля, но и при описании свободного распространения когерентных оптических полей, свойств голографии Френеля. Оно используется в оптике так же часто, как и преобразование Фурье. Использование обоих интегральных преобразований обусловлено тем, что пространство изменяет

распространяющиеся в нем световые волны согласно этим преобразованиям. Распространение света между элементами оптической системы (например, между линзами в оптической системе формирования изображения) также можно описать с помощью преобразования Френеля (или свертки с фазовым множителем, являющимся импульсной характеристикой слоя свободного пространства). Преобразование Френеля применяется и при исследовании голограмм Френеля и анализе систем воспроизведения с апертурами, кодированными зонной пластинкой.

8.7. Преобразование Абеля

Преобразование Абеля, как и преобразование Фурье – Бесселя, применимо к оптическим системам, обладающим вращательной симметрией. Это преобразование связывает функцию одной переменной x (или y) с функцией одной переменной

$r = \sqrt{x^2 + y^2}$ и имеет вид

$$g(x) = 2 \int_x^{\infty} \frac{rf(r)}{\sqrt{r^2 - x^2}} dr.$$

Обратное преобразование Абеля определяется соотношением

$$f(r) = -\frac{1}{\pi} \frac{\partial}{\partial r} \int_r^{\infty} \frac{rg(x)}{x\sqrt{x^2 - r^2}} dx. \quad (8.8)$$

Прямое и обратное преобразования Абеля по существу являются частным решением общей задачи восстановления многомерного объекта по известным проекциям.

Изменением переменной преобразование Абеля можно свести к интегралу свертки. В этом виде преобразование Абеля называется модифицированным. Вследствие своей пространственной инвариантности такое преобразование позволяет при анализе использовать методы Фурье и удобно для

вычислительных целей. Можно показать, что преобразования Абеля, Фурье – Бесселя и Фурье тесно связаны. Последовательное применение этих преобразований к некоторой функции дает исходную функцию. В оптике этот факт выражается в соотношениях между вводимой ниже функцией рассеяния точки и функцией рассеяния линии (преобразование Абеля), между линейной функцией рассеяния и (одномерной) передаточной функцией (преобразование Фурье) и между вводимой ниже оптической передаточной функцией и функцией рассеяния точки (преобразование Фурье – Бесселя).

8.8. Преобразование Гильберта

Прямое преобразование Гильберта определяется соотношением

$$g(x) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(\xi)}{\xi - x} d\xi,$$

а обратное преобразование – соотношением

$$f(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(\xi)}{\xi - x} d\xi.$$

Заметим, что в отличие от других функций и их преобразований, которые определяются в сопряженных областях, в преобразовании Гильберта функции g и f являются функциями одной переменной x . Между этими функциями существует несимметрично-обратное соотношение (обратное, если исключить знак минус). Говорят, что функции $g(x)$ и $f(x)$ сопряжены друг другу. Функцию $g(x)$ иногда называют функцией квадратуры, соответствующей функции $f(x)$.

Видим, что прямое и обратное преобразования Гильберта представляют собой операции свертки соответственно с функциями $-1/\pi x$ и $1/\pi x$. Это приводит к особенно простому

соотношению между $g(x)$ и $f(x)$ в пространстве координат преобразования Фурье.

Преобразование Гильберта связывает действительную и мнимую части аналитического сигнала. Для функции одной пространственной координаты аналитический сигнал можно определить в виде

$$z(x) = 2 \int_0^{\infty} F(u) e^{i2\pi ux} du = f(x) + i g(x),$$

где $g(x)$ и определяется преобразованием Гильберта функции $f(x)$. При этом если, например, $f(x) = \cos kx$, то $g(x) = \sin kx$, если $f(x) = \sin kx$, то $g(x) = -\cos kx$. Аналитический сигнал по

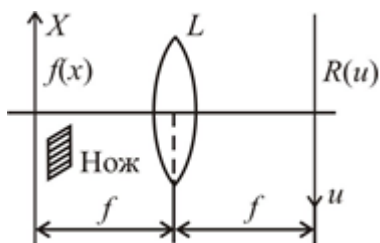


Рис. 8.3

одной из координат для объектов, характеризующихся комплексно-симметричной функцией пропускания, можно получить в оптической системе, показанной на рис. 8.3. На вход системы поступает сигнал $f(x) = f^*(-x)$, половина которого перекрывается

оптическим ножом. Действие ножа описывается ступенчатой функцией Хевисайда $1(x)$, т.е. распределение поля в плоскости пространственных частот (фокальной плоскости линзы) имеет вид $R(u) = F(u) \cdot 1(u)$,

где

$$1(u) = F\{1(x)\} = \frac{1}{2} \left[\delta(u) + \frac{1}{i\pi u} \right]$$

– фурье-образ ступенчатой функции Хевисайда (формула (6.21)). При этом фурье-образ функции $z(x)$

$$Z(u) = F\{z(x)\} \sim R(u).$$

Используя дальнейшее преобразование Фурье, получим непосредственно аналитический сигнал $z(x)$.

В оптических системах формирования изображения аналитический сигнал можно получить, перекрыв маской половину выходного зрачка.

Рассмотрим оптическое преобразование Гильберта. Для этого найдем фурье-образ аналитического сигнала:

$$Z(u) = F(u) + iG(u) = 2F(u)1(u) = 2F(u) \text{ при } u > 0 \\ \text{и } Z(u) = 0 \text{ при } u < 0,$$

Отсюда и получаем упомянутое выше простое соотношение между фурье-образами функций $g(x)$ и $f(x)$:

$$G(u) = -i [2F(u)1(u) - F(u)] = -iF(u) \cdot \text{sign}(u) = -iF(u) \\ \text{при } u > 0 \text{ и } G(u) = iF(u) \text{ при } u < 0,$$

где функция знака $\text{sign}(s)$ принимает значение, равное $+1$, при $s > 0$ и равное -1 , при $s < 0$. Из полученного соотношения видно, что преобразование Гильберта осуществляется в линейной системе с передаточной функцией

$$H(u) = -i \text{sign}(u) = -i = \exp(-i\pi/2) \text{ при } u > 0, \\ H(u) = i = \exp(i\pi/2) \text{ при } u < 0.$$

Оптическая схема, по которой осуществляется такое преобразование, показана на рис. 8.4. Первая линза L_1 формирует пространственный спектр сигнала $f(x)$ в плоскости u . В этой схеме установлен фазовый фильтр (π -фазовая ступенька), который изменяет фазу отрицательных компонент на π . На выходе системы наблюдается сигнал $g(x)$, соответствующий преобразованию Гильберта функции $f(x)$. Схема преобразования Гильберта двумерных сигналов отличается от вышеприведенной схемы только тем, что на вход первой линзы L_1

подается двумерный сигнал $f(x, y)$ и тем что в плоскости пространственных частот u, v устанавливается фазовая пластинка, которая вносит сдвиг фазы на π в четных или нечетных квадрантах. На выходе такой системы формируется сигнал $g(x, y)$, соответствующий двумерному преобразованию Гильберта.

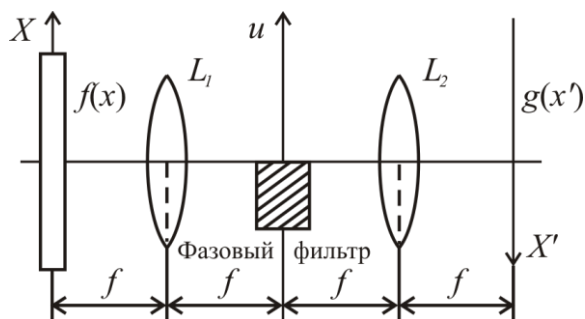


Рис. 8.4

Преобразование Гильберта используется для уменьшения диапазона пространственных частот системы формирования изображения, сохранив при этом разрешение конечной картины. Такая задача может возникнуть при передаче сигнала по каналу с ограниченной пропускаемой способностью. Преобразование Гильберта используется также при определении свойств сигнала в однополосной голограмме.

8.9. Преобразование Мелина

Одномерное преобразование Мелина определяется соотношением

$$g(\alpha) = \int_0^{\infty} f(x)x^{\alpha-1}dx,$$

где $\alpha = \omega + i\tau$ – комплексная переменная. При этом если для некоторого $k > 0$ имеет место соотношение

$$\int_0^{\infty} |f(x)| x^{k-1} dx < \infty,$$

то функция

$$f(x) = \frac{1}{i2\pi} \int_{c-i\infty}^{c+i\infty} g(\alpha) x^{-\alpha} d\alpha \quad (8.9)$$

является обратным преобразованием Мелина для любого $c > k$. Интегрирование в (8.9) ведется вдоль прямой, лежащей на комплексной плоскости $\alpha = \omega + i\tau$ и проходящей параллельно мнимой оси $i\tau$ на расстоянии c от нее. Величина постоянной c определяется характером подынтегральной функции $g(\alpha)x^{-\alpha}$: путь интегрирования должен проходить правее особых точек (полюсов) этой функции. Добавлением к прямой $c \pm i\infty$ дуги бесконечно большого радиуса можно образовать замкнутый контур и произвести контурное интегрирование.

Двумерное преобразование Мелина определяется аналогично:

$$g(\alpha, \beta) = \int_0^{\infty} \int_0^{\infty} f(x, y) x^{i\alpha-1} y^{i\beta-1} dx dy,$$

где $i\alpha$ и $i\beta$ – мнимые аргументы. Обратное двумерное преобразование Мелина

$$f(x, y) = \frac{1}{i(2\pi)^2} \int_{c_1-i\omega_1}^{c_1+i\omega_1} \int_{c_2-i\omega_2}^{c_2+i\omega_2} g(\alpha, \beta) x^{-i\alpha+1} y^{-i\beta+1} d\alpha d\beta.$$

Преобразование Мелина обладает важным для оптики свойством: модуль этого преобразования некоторой функции инвариантен по отношению к изменению масштаба данной функции точно так же, как фурье-образ некоторой функции инвариантен относительно ее сдвига. В обоих случаях в преобразование вводится постепенное линейное изменение фазы, или фазовый наклон. Это свойство позволяет использовать

преобразование Мелина в анализе линейных оптических систем, не являющихся пространственно-инвариантными⁶, и при восстановлении изображения после неоднородного смаза. Применяя преобразование Мелина к оптическим системам, функция рассеяния которых не изменяет своей формы, но меняется соответствующим образом лишь ее величина, мы получаем систему, которую можно анализировать с помощью линейного метода, инвариантного к сдвигу. Совместное применение преобразований Фурье и Мелина позволяет создать оптические корреляторы, нечувствительные не только к сдвигам, но также и к изменению масштаба между объектом и опорным сигналом.

8.10. Преобразование Дирака и преобразование отсчетов

Соотношение

$$f(x, y) = f(x, y) * \delta(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi, \eta) \delta(x - \xi; y - \eta) d\xi d\eta,$$

известное как фильтрующее свойство дельта-функции, называется интегральным преобразованием Дирака для свертки с ядром преобразования $\delta(x, y)$. Функцию $f(x, y)$ можно рассматривать как «дираковский образ». В результате преобразования Дирака исходной функции $f(x, y)$ получается значение этой функции для заданного значения аргумента. Совокупность таких значений определяет исходную функцию. Как видим, дираковский образ состоит из последовательности δ -функций, умноженных на значение исходной функции в соответствующих точках. Преобразование Дирака описывает идеализированную операцию

⁶ Определение линейных пространственно-инвариантных оптических систем будет дано в следующей главе.

сканирования или развертки сигнала вдоль осей координат X и Y с помощью бесконечно острого луча.

Преобразование Дирака обычно появляется в результате двух последовательных линейных операций с ядром типа Фурье.

В одномерном случае преобразование Дирака определяется как

$$f(x) = \int_{-\infty}^{\infty} f(\xi) \delta(x - \xi) d\xi.$$

Если вместо δ -функции Дирака ввести интегральное ядро вида

$$\frac{\sin \pi \alpha x}{\pi x} \cdot \frac{\sin \pi \beta y}{\pi y} = \alpha \operatorname{sinc}(\alpha x) \cdot \beta \operatorname{sinc}(\beta y), \quad (8.10)$$

то получим интегральное преобразование

$$S(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi, \eta) \frac{\sin \pi \alpha (x - \xi)}{\pi (x - \xi)} \cdot \frac{\sin \pi \beta (y - \eta)}{\pi (y - \eta)} d\xi d\eta,$$

называемое преобразованием отсчетов. Это название связано с тем, что ядро вида (8.10) в оптике называют двумерной функцией отсчетов. В одномерном случае преобразование отсчетов

$$S(x) = \int_{-\infty}^{\infty} f(\xi) \frac{\sin \pi \alpha (x - \xi)}{\pi (x - \xi)} d\xi,$$

т.е. является сверткой функции $f(x)$ с функцией $\alpha \operatorname{sinc}(\alpha x)$.

При $\alpha \rightarrow \infty$ и $\beta \rightarrow \infty$ функция отсчетов стремится к δ -функции $\left(\lim_{\substack{\alpha \rightarrow \infty, \\ \beta \rightarrow \infty}} \frac{\sin \pi \alpha x}{\pi x} \frac{\sin \pi \beta y}{\pi y} = \delta(x, y) \right)$, а преобразование

отсчетов переходит в преобразование Дирака, поэтому оба эти преобразования очень похожи друг на друга. Отличие преобразования отсчетов от преобразования Дирака состоит в том, что в этом преобразовании функция $f(x, y)$ сканируется не

бесконечно узким лучом, а лучом, размытым вдоль осей координат X и Y ; происходит процесс сглаживания исходной функции $f(x, y)$. При преобразовании отсчетов происходит отображение средних значений исходной функции на некотором интервале.

Преобразование отсчетов используется для обрезания или сужения спектра пространственных частот исходного сигнала, описывает свойства оптического сигнала, поступающего из системы с конечной апертурой или ограниченной полосой пространственных частот.

Преобразование отсчетов можно рассматривать как обобщение теоремы Котельникова, называемой также теоремой отсчетов. Согласно этой теореме, любую функцию $f(x)$, имеющую ограниченный спектр пространственных частот (от $u = 0$ до некоторого максимального значения $u = u_c$), можно разложить в ряд по ортогональным функциям $y_m(x)$, причем коэффициенты разложения c_m являются дискретными значениями функции $f(x)$, взятыми через интервал $\Delta x = 1 / 2u_c$, т.е.

$$f(x) = \sum_{m=-\infty}^{\infty} c_m y_m(x), \quad (8.11)$$

где $c_m = f(m\Delta x)$ – значения функции $f(x)$ в точках $x = m\Delta x = m / 2u_c$,

$$y_m(x) = \frac{\sin 2\pi u_c (x - m\Delta x)}{2\pi u_c (x - m\Delta x)} = \text{sinc}[2u_c(x - m\Delta x)],$$

$|m| = 0, 1, 2, \dots$

Для доказательства этой теоремы рассмотрим простейший одномерный сигнал с ограниченной полосой частот, у которого спектральная плотность $g(u)$ равна постоянной величине $1 / 2u_c$ в интервале частот от $-u_c$ до u_c и нулю вне этого интервала, т.е.

$$g(u) = \begin{cases} 1/2u_c, & |u| \leq u_c, \\ 0, & |u| > u_c. \end{cases}$$

Используя обратное преобразование Фурье, находим этот сигнал:

$$y_0(x) = \int_{-\infty}^{\infty} g(u) e^{i2\pi ux} du = \int_{-u_c}^{u_c} \frac{1}{2u_c} e^{i2\pi ux} du = \frac{\sin 2\pi u_c x}{2\pi u_c x} =$$

$$= \text{sinc}(2u_c x).$$

Функция $y_0(x) = \text{sinc}(2u_c x)$ имеет максимальное значение, равное единице; в точках $x = m / 2u_c$, $|m| = 1, 2, 3, \dots$ она равна нулю. При смещении этой функции вдоль оси X на целое число интервалов $\Delta x = 1 / 2u_c$ получаются функции

$$y_m(x) = y_0(x - m\Delta x) = \text{sinc}[2u_c(x - m\Delta x)],$$

ортогональные друг другу, т.е.

$$J = \int_{-\infty}^{\infty} y_m(x) y_n(x) dx = \begin{cases} 1/2u_c, & m = n, \\ 0, & m \neq n. \end{cases}$$

Это свойство функций $y_m(x)$ позволяет использовать их в качестве базиса. Поэтому любую функцию $f(x)$ со спектром, заключенным в полосе частот от 0 до u_c , можно представить в виде ряда Котельникова (8.11). Для определения коэффициентов разложения c_m положим $x = n\Delta x = n / 2u_c$, где n – целое число. Тогда получим

$$f(x) = \sum_{m=-\infty}^{\infty} c_m y_m(x) = \sum_{m=-\infty}^{\infty} c_m \text{sinc}(n - m).$$

При $m \neq n$ это выражение равно нулю, а при $m = n$ – единице. Отсюда получаем $c_m = f(m\Delta x)$. Теорема Котельникова доказана.

Обозначим период, соответствующий наивысшей частоте u_c через l_c , т.е. положим $l_c = 1/u_c$. В этих обозначениях коэффициенты разложения можно записать в виде $c_m = f(ml_c/2)$. Таким образом, функция с ограниченным спектром полностью определяется своими значениями, взятыми через интервал $\Delta x = l_c/2$.

Используя условие ортогональности, из (8.9) можно получить выражение для энергии сигнала:

$$\int_{-\infty}^{\infty} f^2(x) dx = \frac{1}{2u_c} \sum_{m=-\infty}^{\infty} y_m^2(x).$$

Хотя суммирование в ряде Котельникова проводится в бесконечных пределах, в действительности, однако, число значений функции $f(x)$ ограничено и определяется длиной сигнала L и шириной его полосы частот u_c :

$$N = L / \Delta x = 2L / l_c = 2Lu_c.$$

При $N \gg 1$ погрешность от перехода к конечному пределу суммирования невелика, т.е. в интервале $0 \leq x \leq L$ функция $f(x)$ полностью определяется N выборками из нее. От числа N зависит количество информации, содержащейся в данном сигнале $f(x)$ (точнее от $N + 1$, так как число точек на единицу больше числа интервалов). В оптике в качестве периода l_c принимают размер наиболее мелкой детали l_{min} , а в качестве длины сигнала L – длину транспаранта.

Для двумерного сигнала $f(x, y)$ ряд Котельникова имеет вид

$$f(x, y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(m\Delta x, n\Delta y) \text{sinc}[2u_c(x - m\Delta x)] \times$$

$$\times \text{sinc}[2v_c(y - n\Delta y)],$$

где u_c и v_c – граничные частоты по соответствующим переменным x и y .

Теорема Котельникова имеет важное значение в теории сигналов и в оптике. На ее основе любой непрерывный сигнал может быть со сколь угодно высокой степенью точности преобразован в дискретный.

8.11. Преобразование Лапласа

В оптике это преобразование не используется, но находит широкое применение в теории электрических сигналов. Преобразование Лапласа определяется обобщенным экспоненциальным ядром и представляет собой распространение принципа преобразования Фурье на функции, для которых не существует фурье-образов. Если для функции $f(x)$ условие

$$\int_{-\infty}^{\infty} |f(x)| dx < \infty$$

не выполняется, но интеграл $\int_{-\infty}^{\infty} |f(x)| \exp(-\sigma x) dx$ ограничен (для некоторого действительного числа σ), то преобразование Лапласа функции $f(x)$ по отношению к комплексной переменной p запишется в виде

$$L(p) = \int_{-\infty}^{\infty} f(x) \exp(-px) dx, \quad (8.12)$$

причем $\text{Re}(p) > \sigma$. Преобразование Лапласа, в котором интегрирование проводится в бесконечных пределах, называют двусторонним. Обратное преобразование Лапласа определяется формулой

$$f(x) = \frac{1}{i2\pi} \int_{c - i\infty}^{c + i\infty} L(p) \exp(px) dp,$$

где $s > \sigma$. Полагая в выражении (8.11) нижний предел интегрирования равным нулю, получим одностороннее преобразование Лапласа. Легко видеть, что одностороннее преобразование Лапласа и преобразование Фурье – это частные случаи двустороннего преобразования Лапласа. Преобразование Фурье получается из преобразования Лапласа формальной заменой переменной p на ik (или на $i\omega$), т.е. имеет место при мнимом p . Вообще говоря, преобразование Лапласа функции $f(x)$ эквивалентно преобразованию Фурье функции $f(x)\exp(-\alpha x)$, где α – вещественная часть комплексной величины p .

Используя замену $x = \exp(-i\xi)$, можно показать, что преобразование Мелина функции $f(x)$ эквивалентно двустороннему преобразованию Лапласа функции от ξ . В случае чисто мнимого α аналогичное соотношение имеет место между преобразованиями Мелина и Фурье.

Для функции двух переменных преобразование Лапласа определяется аналогично:

$$L(p, q) = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} f(x, y) \exp[-(px + qy)] dx dy.$$

Обратное двумерное преобразование Лапласа

$$f(x, y) = \frac{1}{i(2\pi)^2} \int_{c_1 - i\omega_1}^{c_1 + i\omega_1} \int_{c_2 - i\omega_2}^{c_2 + i\omega_2} L(p, q) \exp(px + qy) dp dq.$$

Свойства преобразования Лапласа в общем случае очень похожи на свойства преобразования Фурье.

Глава 9

ЛИНЕЙНЫЕ ОПТИЧЕСКИЕ СИСТЕМЫ

9.1. Понятие линейной оптической системы

Оптической системой называют прозрачные изотропные тела, преломляющие и отражающие световые лучи. Оптическая система представляет собой набор последовательно расположенных друг за другом оптических элементов. Такими оптическими элементами являются слой свободного пространства, диафрагмы (отверстия в непрозрачном экране), линзы, плоские пластинки, призмы, решетки и т.д. В дальнейшем будут рассматриваться центрированные относительно оптической оси (обычно оси Z схемы) оптические системы. Оптические элементы системы (за исключением слоя свободного пространства) будут представляться в виде тонкого экрана, модулирующего световой поток, проходящий через него в каждой точке экрана. Элементарный световой сигнал будет рассматриваться как луч света, проходящий в данной точке экрана. Необходимым условием применимости такого представления является выполнение неравенства $h \ll d^2/\lambda$, где h – толщина экрана, d – пространственный период сигнала. При этом условии луч света проходит через экран, не расширяясь (зона Френеля).

Оптическая система называется линейной, если ее реакция (или отклик) на сумму элементарных воздействий, называемую входным сигналом, равна сумме реакций на каждое элементарное воздействие, представляющее световой сигнал. Понятие линейной системы, как уже отмечалось, пришло в оптику из радиотехники и теории связи. Оказывается, что оптические системы и рассматриваемые в теории электрических цепей линейные четырехполюсники сходны между собой. Например, линзу можно и удобно представить в виде четырехполюсника. Подобно четырехполюснику оптическую систему, создающую изображение, можно рассматривать как систему, преобразующую сигналы. Поведение такой системы, как и поведение

четырёхполюсника, описывается соотношением между возбуждением на входе и реакцией на выходе.

Рассматриваемые в радиотехнике сигналы представляют собой функции времени, т.е. зависят от одной переменной – времени t . В оптике, как уже отмечалось, сигналом называют пространственное распределение амплитуды (при использовании когерентного света) или интенсивности (при некогерентном освещении) светового поля, т.е. функции $E(x, y)$ или $I(x, y)$ в некоторой плоскости XU . Оптические сигналы обычно записываются на фотопленках или фотопластинках, которые в данном случае называются транспарантами (диапозитивами). Коэффициент пропускания транспаранта $t(x, y)$ по амплитуде пропорционален зарегистрированному сигналу $E(x, y)$. Этот коэффициент, как известно, связывает между собой комплексную амплитуду светового поля на входе оптической системы с комплексной амплитудой поля на выходе. При восстановлении сигнала транспарант освещают плоской или сферической волной. Распределение поля в плоскости XU непосредственно за предметом (транспарантом) называется входным (для оптической системы) сигналом, а саму плоскость X , расположенную непосредственно за предметом, – плоскостью предмета или входной плоскостью (иногда эту плоскость называют пространством входного сигнала).

Функция пропускания предмета в общем случае является комплексной:

$$t(x, y) = |t(x, y)| \exp[i\varphi(x, y)],$$

причем модуль этой функции характеризует прозрачность предмета, а ее аргумент – набег фазы световой волны в предмете, зависящий от его толщины. В соответствии с этим предмет преобразует поле падающей на него световой волны $E_0(x, y)$ по закону

$$E(x, y) = E_0(x, y)t(x, y) = E_0(x, y)|t(x, y)| \exp[i\varphi(x, y)],$$

т.е. пространственно модулирует это поле по амплитуде и фазе (или только по одной из этих характеристик). Амплитудная модуляция определяется модулем функции пропускания $|t(x, y)|$, а фазовая – его аргументом $\varphi(x, y)$. Промодулированная предметом световая волна распространяется вдоль оси Z и, пройдя сквозь оптическую систему, образует в некоторой плоскости $X'Y'$ распределение амплитуды $E'(x', y')$ (или интенсивности $I(x', y')$). Распределение, получившееся в плоскости $X'Y'$ после преобразования оптической системой распределения $E(x, y)$ (или $I(x, y)$), называют выходным сигналом или реакцией оптической системы на входной сигнал, а плоскость $X'Y'$ – выходной плоскостью или плоскостью изображения (иногда пространством выходного сигнала). Характеристики выходного сигнала определяются, таким образом, как характеристиками входного сигнала, так и характеристиками оптической системы.

Рассмотрим необходимые условия линейности оптической системы. Обозначим исходный (входной) сигнал через $f(x, y)$, а конечный (выходной) сигнал – через $F(x', y')$, где x, y, x', y' – координаты соответственно во входной (XY) и выходной ($X'Y'$) плоскостях (предмета и его изображения). По аналогии с радиотехникой входной сигнал часто называют возбуждением, а выходной сигнал – откликом или реакцией на входное воздействие. Предположим, что входной сигнал $f_1(x, y)$ дает на выходе реакцию $F_1(x', y')$, а входной сигнал $f_2(x, y)$ – реакцию $F_2(x', y')$. Если оптическая система является линейной, то возбуждение на входе

$$f_1(x, y) + f_2(x, y)$$

вызывает реакцию

$$F_1(x', y') + F_2(x', y')$$

на выходе. Таким образом, необходимым условием линейности оптической системы является выполнение принципа суперпозиции, согласно которому если на систему действует несколько сигналов, то каждый отдельный сигнал не влияет на реакцию, вызванную другим сигналом, а выходной сигнал от суммы входных сигналов равен сумме сигналов от каждого входного сигнала, действующего поодиночке. Следовательно, в основе деления систем на линейные и нелинейные лежит условие применимости принципа суперпозиции, сформулированное относительно операции суммирования сигналов на выходе системы. Пусть далее $cf(x, y)$ – сигнал на входе оптической системы, где c – произвольная постоянная. Тогда для линейной оптической системы этому возбуждению на входе будет соответствовать реакция $cF(x, y)$ на выходе. Это второе необходимое условие линейности оптической системы описывает свойство гомогенности системы. Оно выражает собой свойство сохранения масштаба линейной системой. Следовательно, оптическая система линейна тогда и только тогда, когда она обладает свойствами аддитивности и гомогенности.

Выходной сигнал является результатом действия оптической системы на входной сигнал. Это действие можно выразить

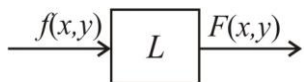


Рис. 9.1

математически как действие некоторого оператора L , присущего данной оптической системе. Оператор показывает, как нужно подействовать на входную функцию

$f(x, y)$, чтобы получить выходную функцию $F(x, y)$. Введя понятие оператора, преобразование входного сигнала в выходной сигнал можно записать в символической форме

$$F(x, y) = L\{f(x, y)\}.$$

Действие оператора L показано на рис. 9.1. Линейная оптическая система описывается линейным оператором L . Линейная оптическая система обладает тем свойством, что если,

например, для двух входных сигналов $f_1(x, y)$ и $f_2(x, y)$ и для двух постоянных c_1 и c_2 входной сигнал $c_1 f_1(x, y) + c_2 f_2(x, y)$ преобразуется в выходной сигнал $c_1 F_1(x, y) + c_2 F_2(x, y)$ на выходе, т.е. выполняется условие

$$L\{c_1 f_1(x, y) + c_2 f_2(x, y)\} = c_1 F_1(x, y) + c_2 F_2(x, y).$$

Для произвольного числа n воздействий это условие можно записать в виде:

$$L\left\{\sum_{i=1}^n c_i f_i(x, y)\right\} = \sum_{i=1}^n c_i F_i(x, y). \quad (9.1)$$

Соотношение (9.1) описывает одновременно и свойство аддитивности и свойство гомогенности линейной оптической системы.

Среди всех линейных оптических систем наибольшее значение в фурье-оптике имеют пространственно-инвариантные оптические системы. Пространственно-инвариантными называют оптические системы, инвариантные относительно сдвигов в координатной области, а частотно-инвариантными – оптические системы, инвариантные относительно сдвигов в частотной области. Первые называются фильтрами, а вторые – модуляторами. Фильтром является, например, слой свободного пространства, заключенный между двумя параллельными плоскостями. К модуляторам относятся апертуры, дифракционные решетки, зеркала, призмы, линзы.

Для линейных пространственно-инвариантных оптических систем имеет место соотношение

$$F(x - x_0; y - y_0) = L\{f(x - x_0; y - y_0)\}.$$

Это означает, что если входное воздействие $f(x, y)$ вызывает выходную реакцию $F(x, y)$, то входное воздействие $f(x - x_0; y - y_0)$ вызывает реакцию $F(x - x_0; y - y_0)$. Здесь x_0 и y_0 –

постоянные; масштаб системы координат в плоскости XU выбран так, что коэффициент увеличения равен единице.

Частотная инвариантность означает, что если спектральные плотности входного и выходного сигналов соответственно $g(u, v)$ и $G(u, v)$, то при спектральной плотности входного сигнала $g(u - u_0; v - v_0)$ спектральная плотность выходного сигнала будет $G(u - u_0; v - v_0)$, где u_0 и v_0 – постоянные. В операторном виде это утверждение можно записать так:

$$G(u - u_0; v - v_0) = L\{g(u - u_0; v - v_0)\}.$$

Пространственно-инвариантные линейные оптические системы (фильтры) обладают важным свойством: они не создают новых пространственных частот. Это означает, что если входное воздействие $f(x, y)$, приложенное к такой системе, содержит в своем спектре пространственные частоты u и v , то и вызываемая им выходная реакция $F(x, y)$ будет содержать те же самые частоты u и v .

Фильтры можно характеризовать импульсной характеристикой $h(x, y)$ и передаточной функцией $H(u, v)$, так как эти величины не зависят от сдвигов в координатной области⁷.

Модуляторы характеризуются функцией пропускания $t(x, y)$. Эта величина не зависит от сдвигов в частотной области. Убедимся в этом. Запишем обратное фурье-преобразование функции $f(x, y)$ в предположении, что ее пространственно-частотный спектр смещен на величины u_0, v_0 . Обозначив через $f_1(x, y)$ функцию, описывающую входной сигнал со смещенным спектром, будем иметь

⁷ Термины «импульсный отклик», «импульсная характеристика» и «передаточная функция» так же, как и термин «линейная система», пришли в фурье-оптику из теории связи.

$$\begin{aligned}
f_1(x, y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\alpha - u_0; \beta - v_0) e^{i2\pi(\alpha x + \beta y)} d\alpha d\beta = \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u, v) e^{i2\pi[(u+u_0)x + (v+v_0)y]} dudv = \\
&= \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u, v) e^{i2\pi(ux + vy)} dudv \right) \cdot e^{i2\pi(u_0x + v_0y)} = \\
&= f(x, y) \exp[i2\pi(u_0x + v_0y)].
\end{aligned}$$

Аналогично можно получить и для выходного сигнала:

$$F_1(x, y) = F(x, y) \exp[i2\pi(u_0x + v_0y)].$$

С учетом этого для функции пропускания получим

$$\begin{aligned}
t_1(x, y) &= \frac{F_1(x, y)}{f_1(x, y)} = \frac{F(x, y) \exp[i2\pi(u_0x + v_0y)]}{f(x, y) \exp[i2\pi(u_0x + v_0y)]} = \\
&= \frac{F(x, y)}{f(x, y)} = t(x, y).
\end{aligned}$$

Отсюда видно, что, хотя при сдвиге частот входной и выходной сигналы приобретают фазовый множитель $\exp[i2\pi(u_0x + v_0y)]$, функция пропускания модулятора, определяемая отношением этих сигналов, от сдвигов в частотной области не зависит (не зависит от u_0, v_0).

9.2. Импульсная характеристика оптической системы

Свойство линейности оптических систем существенно облегчает анализ их отклика на входной сигнал сложной формы, так как для этого нужно знать только отклик системы на сигнал простой формы и уметь представить входной сигнал как

суперпозицию таких простых сигналов. Выходной сигнал тогда будет равен суперпозиции таких откликов. Если в качестве входного сигнала $E(x, y)$ взять элементарный источник волн в виде точечного источника, описываемого дельта-функцией ($E(x, y) = \delta(x, y)$), то реакция (отклик) на такое воздействие

$$h(x, y; x', y') = L\{\delta(x' - x; y' - y)\} \quad (9.2)$$

называется импульсной характеристикой (или импульсным откликом) оптической системы. Ее называют также функцией пропускания оптической системы. Таким образом, под импульсной характеристикой оптической системы подразумевается отклик (реакция) оптической системы на воздействие, имеющее вид единичного импульса (дельта-функции). Как видим, понятие импульсной характеристики вводится, когда распределение комплексных амплитуд поля в плоскости предмета определяется δ -функцией. Можно сказать, что импульсная характеристика (функция пропускания) – это распределение амплитуды светового поля в плоскости изображения, которое получается при изображении системой точечного источника света. Физический смысл импульсной характеристики состоит в том, что он представляет собой поле точечного источника, моделью которого может служить маленькое прямоугольное отверстие в плоском экране с большой интенсивностью поля в отверстии. Другими словами, импульсная характеристика – это поле в произвольной точке (x, y, z) , создаваемое двумерным точечным источником

$$E(x, y, 0) = \delta(x)\delta(y)$$

Импульсная характеристика полностью определяет оптическую систему. Это означает, что отклик $E'(x', y')$ на произвольный сигнал $E(x, y)$ определяется функцией $h(x, y; x', y')$. Чтобы убедиться в этом, представим входной сигнал в виде суперпозиции бесконечно большого числа элементарных точечных источников – двумерных δ -функций:

$$E(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(\xi, \eta) \delta(x - \xi; y - \eta) d\xi d\eta \quad (9.3)$$

(предполагается, что функция $E(\xi, \eta)$ непрерывна в точке $x = \xi$, $y = \eta$). Соотношение (9.3) записано на основании фильтрующего свойства δ -функции. Поскольку поле в выходной плоскости $X'Y'$ связано с полем во входной плоскости XY соотношением $E'(x', y') = L\{E(x, y)\}$, то с учетом (9.2) и (9.3) получим

$$E'(x', y') = L \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(\xi, \eta) \delta(x - \xi; y - \eta) d\xi d\eta \right\}.$$

Если рассматривать $E(\xi, \eta)$ как весовой коэффициент для элементарной функции $\delta(x - \xi; y - \eta)$ (он определяет «вес» элементарного источника в разложении (9.3)) и воспользоваться свойством линейности (9.1), будем иметь

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(\xi, \eta) L\{\delta(x - \xi; y - \eta)\} d\xi d\eta.$$

Учитывая (9.2) и заменив переменные x и y на x' и y' , а ξ и η на x и y , окончательно получим

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) h(x, y; x', y') dx dy. \quad (9.4)$$

(Напомним, что x' и y' принимают те же значения, что и x и y). Таким образом, выходной сигнал $E'(x', y')$ оказывается представленным в виде суммы (интеграла) элементарных откликов $h(x, y; x', y')$ с весовыми коэффициентами $E(x, y)$. Это выражение можно трактовать и иначе, а именно: каждый элементарный сигнал в точке (x, y) на входе передается на выход

с весовой функцией $h(x, y; \xi, \eta)$ и выходной сигнал является суммой всех элементарных сигналов.

Получим соотношение (9.4) другим способом. Будем характеризовать оптическую систему функцией пропускания $h(x, y; x', y')$. Учтем, что функция пропускания представляет собой комплексную амплитуду поля (регистрируемого на единицу площади в плоскости $X'Y'$) в точке (x', y') выходной плоскости $X'Y'$, обусловленную полем с единичной амплитудой и нулевой фазой в точке (x, y) входной плоскости XY . Тогда если $E(x, y)$ – амплитуда поля в точке (x, y) входной плоскости, то элемент поверхности, содержащий точку (x, y) , создает поле

$$dE'(x', y') = E(x, y)h(x, y; x', y') dx dy$$

в точке (x', y') выходной плоскости. Полное поле в точке (x', y') найдем, проинтегрировав это выражение по площади источника. Интегрирование можно провести и по бесконечным пределам, так как для всех точек входной плоскости, от которых свет не попадает в выходную плоскость, величина $E(x, y)$ равна нулю. Это и приводит к соотношению (9.4).

Дельта-функция модулирует узкий и высокий «импульс» в точке (x, y) (точечный источник света). По этой причине функцию $h = L\{\delta\}$ называют импульсным откликом системы. Выражение (9.4) называется интегралом суперпозиции. Он показывает, что линейная оптическая система полностью характеризуется ее откликом на точечный сигнал. Таким образом, для линейной оптической системы сигнал на выходе определяется откликом системы на δ -функцию, которая располагается последовательно во всех точках плоскости входного сигнала (плоскости предмета XY). Отсюда следует, что для линейных оптических систем формирования изображения влияние отдельных элементов (линз, диафрагм и т.д.) можно полностью учесть, если располагать последовательно в предметной плоскости точечный источник, определяя при этом отклик системы, т.е. структуру изображения

точки. Зная импульсную характеристику линейной оптической системы, можно решить любую задачу о прохождении оптического сигнала через такую систему.

Поскольку импульсную функцию можно рассматривать как комплексную амплитуду в точке изображения точечного источника, то она (рассматриваемая как функция от x', y') имеет резкий максимум в точке изображения (в точке $x' = x, y' = y$) или вблизи от нее и быстро спадает с увеличением расстояния до этой точки. В скоррегированной оптической системе (в системе с достаточно ослабленными аберрациями) функция $h(x', y')$ имеет заметную величину лишь в области, размер которой порядка диаметра светового кружка в дифракционной картине Эри. Как функция координат x, y функция $h(x, y)$ изменяется очень слабо при перемещении точки по поверхности предмета. Она не зависит и от самой точки изображения. В скоррегированной системе функция $h(x, y; x', y')$ описывает с точностью до постоянного множителя картину Эри, центр которой находится в точке изображения.

Как уже отмечалось, важный класс линейных оптических систем составляют так называемые пространственно-инвариантные (или изопланарные) системы. Эти системы инвариантны относительно сдвигов в координатной области. Следствием этого является то, что импульсная характеристика h зависит не от координат x' и x, y' и y порознь, а от их разностей $x' - x$ и $y' - y$:

$$h(x, y; x', y') = h(x' - x, y' - y). \quad (9.5)$$

Для систем формирования изображения условие (9.5) означает, что при изменении положения точечного источника его изображение меняет только расположение, но не свою функциональную форму, форма выходного сигнала не зависит от смещения входного сигнала. Заметим, что оптические системы, как правило, не являются пространственно-инвариантными по всем входной и выходной плоскостям. Однако они обычно

пространственно-инвариантны внутри достаточно малых областей (обычно вблизи оптической оси системы), которые называют изопланарными участками. Тогда для любого изопланарного участка систему считают линейной и пространственно-инвариантной. Условие изопланарности хорошо выполняется при распространении света в свободном пространстве, а также при использовании в оптической системе линз с достаточно плавным изменением аберраций по полю и небольшим относительным отверстием.

Для линейных пространственно-инвариантных систем интеграл суперпозиции (9.4) представляет собой свертку распределения поля в плоскости предмета (входного сигнала) с импульсной характеристикой оптической системы:

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) h(x' - x; y' - y) dx dy \quad (9.6)$$

или при использовании обозначение свертки,

$$E'(x', y') = E(x, y) * h(x, y; x', y').$$

Это соотношение устанавливает связь между входным и выходным сигналами линейной пространственно-инвариантной системы с помощью импульсной характеристики. Выражение (9.6) показывает важное свойство выходной реакции линейной пространственно-инвариантной системы: выходная реакция является сверткой ее импульсной характеристики с входным воздействием. Возможность представления выходного сигнала в виде интеграла суперпозиции является отличительным признаком линейности системы.

Если в выражение (9.6) вместо $E(x, y)$ подставить $\delta(x, y)$, то, в соответствии с фильтрующим свойством δ -функции, получим

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x, y) h(x' - x; y' - y) dx dy = h(x', y').$$

Это подтверждает сказанное выше о том, что импульсная характеристика $h(x', y')$ оптической системы характеризует выходной сигнал системы при точечном источнике света в качестве входного сигнала.

Часто импульсная характеристика оптической системы обладает свойством зеркальной симметрии, а именно

$$h(x, y) = h^*(-x, -y);$$

$$h(x' - x; y' - y) = h^*(x - x'; y - y').$$

Для таких систем соотношение (9.6) можно переписать в виде

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) h^*(x - x'; y - y') dx dy.$$

Это выражение определяет корреляцию двух функций: $E(x, y)$ и $h^*(x, y)$. Следовательно, для линейных систем с импульсной

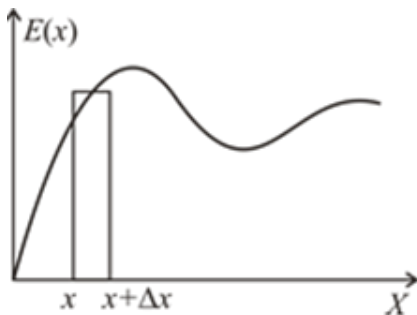


Рис. 9.2

функцией, обладающей зеркальной симметрией, выходной сигнал можно представить в виде функции корреляции входного сигнала и функции, комплексно-сопряженной с импульсной реакцией.

К интегралу суперпозиции (9.6), называемому также интегралом Дюамеля,

можно придти и иначе. Покажем это в случае одномерного сигнала. Пусть входной сигнал $E(x)$ имеет вид, показанный на рис. 9.2. Разобьем этот сигнал на элементарные прямоугольные импульсы (один из таких импульсов показан на рис. 9.2) и найдем отклик оптической системы в точке x' на элементарный импульс, действующий на входе в точке x . Если бы площадь этого импульса равнялась единице, то его можно было бы

рассматривать как дельта-функцию, возникающую в точке x . При импульсной характеристике оптической системы $h(x)$ отклик в точке x' был бы, очевидно, равен $h(x' - x)$. Поскольку, однако, площадь этого импульса равна $E(x)\Delta x$, а не единице, отклик в точке x' будет $E(x)\Delta x h(x' - x)$. Для определения полного значения выходного сигнала в точке x' нужно просуммировать действие всех импульсов в интервале от $x = 0$ до $x = x'$. При $\Delta x \rightarrow 0$ суммирование сводится к интегрированию. При этом, поскольку функция $h(x)$ узкая, интегрирование можно распространить на бесконечные пределы. Тогда в случае одномерного сигнала можно записать

$$E'(x') = \int_{-\infty}^{\infty} E(x)h(x' - x)dx .$$

Введем теперь понятие импульсного отклика линейного спектрометра. Пусть $R(\omega)$ – функция спектрального распределения энергии на входе линейной системы. Соответствующая функция на выходе будет $R'(\omega) = L\{R(\omega)\}$. Функция $R(\omega)$ может быть либо спектром, наблюдаемым с помощью дифракционной решетки в зависимости от циклической частоты ω или длины волны λ , либо откликом в оптическом спектральном анализаторе. Пусть, далее, функция на входе $R(\omega)$ состоит из суммы единичных импульсов

$$R(\omega) = \int_{-\infty}^{\infty} R(\omega_1)\delta(\omega - \omega_1)d\omega_1 . .$$

Функция на выходе

$$R'(\omega) = L\{R(\omega)\} = \int_{-\infty}^{\infty} R(\omega_1)L\{\delta(\omega - \omega_1)\}d\omega_1 =$$

$$= \int_{-\infty}^{\infty} R(\omega_1)h(\omega - \omega_1)d\omega_1$$

или

$$R'(\omega) = R(\omega) * h(\omega),$$

где $h(\omega) = L\{\delta(\omega)\}$. Определенную таким образом функцию $h(\omega)$ и называют импульсным откликом спектрометра. Полученное соотношение определяет связь между функциями спектрального распределения энергии на входе и выходе спектрального прибора и представляет собой основное уравнение, определяющее характерные свойства спектрометров.

9.3. Передаточная функция оптической системы

Понятие пространственной инвариантности позволяет ввести в рассмотрение и другую важнейшую характеристику оптической системы, называемую частотной (имеются в виду пространственные частоты) или передаточной функцией. Чтобы прийти к этой характеристике, покажем, что линейная пространственно-инвариантная оптическая система преобразует гармонический сигнал на входе в гармонический сигнал тех же пространственных частот на выходе. Действительно, пусть сигнал

$$E(x, y) = \exp [i 2\pi (ux + vy)].$$

На основании коммутативного свойства свертки ($f * g = g * f$) интеграл суперпозиции (9.6) перепишем в виде

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x' - x; y' - y)h(x, y) dx dy. \quad (9.7)$$

Подставив в (9.7) выражение гармонического сигнала, получим

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) e^{i2\pi[u(x'-x)+v(y'-y)]} dx dy =$$

$$= \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) e^{-i2\pi(ux+vy)} dx dy \right] e^{i2\pi(ux'+vy')}.$$

Или

$$E'(x', y') = H(u, v) e^{i2\pi(ux'+vy')},$$

где

$$H(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) e^{-i2\pi(ux+vy)} dx dy . \quad (9.8)$$

Тем самым мы показали, что гармонический сигнал на входе линейной пространственно-инвариантной оптической системы преобразуется в гармонический сигнал тех же пространственных частот на выходе:

$$L\{e^{i2\pi(ux+vy)}\} = H(u, v) e^{i2\pi(ux'+vy')}.$$

Комплексную функцию H пространственных частот u и v , определяемую выражением (9.8), называют частотной или передаточной функцией оптической системы. Как видно из (9.8), передаточная функция оптической системы представляет собой фурье-образ импульсной характеристики этой системы. В этом смысле передаточная функция является сопряженной характеристикой по отношению к импульсной характеристике. Передаточная функция полностью характеризует фильтрующие свойства оптической системы. В этом смысле любую пространственно-инвариантную оптическую систему можно рассматривать как линейный фильтр пространственных частот, а прохождение сигнала через оптическую систему – как пространственную фильтрацию. Зная передаточную функцию $H(u, v)$, с помощью обратного фурье-преобразования можно определить импульсную характеристику $h(x, y)$:

$$h(x, y) = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} H(u, v) e^{i2\pi(ux+vy)} dudv. \quad (9.9)$$

Частотная характеристика $H(u, v)$ является комплексной функцией, поэтому ее можно представить в виде

$$H(u, v) = |H(u, v)| \exp [i\varphi(u, v)]. \quad (9.10)$$

Вещественная функция $|H(u, v)|$ при этом называется амплитудно-частотной характеристикой (АЧХ), а $\varphi(u, v)$ – фазо-частотной характеристикой (ФЧХ) оптической системы. Часто их называют просто амплитудной и фазовой характеристиками линейной оптической системы. Соотношение (9.10) показывает, что гармонический сигнал после прохождения через линейную пространственно-инвариантную оптическую систему, не изменяя своих пространственных частот, изменяет амплитуду в $|H(u, v)|$ раз и фазу на $\varphi(u, v)$.

Взяв преобразование Фурье от обеих частей соотношения (9.6) и воспользовавшись теоремой свертки, получим

$$E'(u, v) = E(u, v) H(u, v), \quad (9.11)$$

где $E(u, v)$ и $E'(u, v)$ – фурье-образы входного и выходного сигналов соответственно. Таким образом, преобразование Фурье реакции линейной пространственно-инвариантной оптической системы равно произведению фурье-преобразования входного воздействия и фурье-преобразования импульсной характеристики линейной системы. Другими словами, фурье-образ (спектр) выходного сигнала равен произведению передаточной функции оптической системы на фурье-образ (спектр) входного сигнала. Следует подчеркнуть тот факт, что если известна импульсная характеристика оптической системы, то ее преобразование Фурье является передаточной функцией этой системы.

Соотношение (9.11) позволяет заменить сложную операцию свертки простой операцией умножения пространственных спектров. Это соотношение показывает, что если сигналы в плоскости предмета и плоскости изображения представить в виде суперпозиции пространственных гармоник всех возможных пространственных частот u , v , то каждая компонента сигнала в плоскости изображения будет определяться соответствующей компонентой в плоскости предмета. Учитывая, что $E(u, v)$ представляет собой разложение входного сигнала $E(x, y)$ в виде линейной комбинации гармонических функций (гармоник) с различными пространственными частотами u и v , на основании соотношения (9.11) получаем, что при умножении на передаточную функцию $H(u, v)$ учитывается действие оптической системы на каждую гармонику отдельно. Это действие, в соответствии с (9.10), сводится к изменению амплитуды и фазы каждой составляющей спектра. Согласно (9.11), сигнал передается оптической системой без искажений, т.е. $E'(u, v) = E(u, v)$, если

$$H(u, v) = 1 \quad (9.12)$$

для любых u и v . Так как фурье-образ δ -функции равен единице, то равенство (9.12) имеет место, если импульсной характеристикой системы является δ -функция. Равенство (9.12) может выполняться лишь для некоторых частот. Тогда эти составляющие сигнала передаются без искажений. Частоты же, для которых передаточная функция $H(u, v) = 0$, подавляются (отфильтровываются) оптической системой полностью.

Согласно определению прямого и обратного фурье-преобразований, из (9.11) следует

$$E(x, y) = F^{-1}\{E'(u, v)\} = F^{-1}\{E(u, v)H(u, v)\}.$$

Это соотношение позволяет восстановить исходную волновую информацию. Для этого следует либо зарегистрировать в произвольной плоскости пространственный спектр выходного

сигнала $E'(u, v)$, либо рассчитать его по результатам измерения. Затем, зная отклик или передаточную характеристику оптической системы, найти произведение $E(u, v)H(u, v)$, обратное фурье-преобразование которого является исходным сигналом.

Так как

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E'(u, v) e^{i2\pi(ux'+vy')} dudv,$$

то на основании соотношения (9.11) для выходного сигнала можно записать

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v)H(u, v) e^{i2\pi(ux'+vy')} dudv. \quad (9.13)$$

Выражение (9.13) показывает, что сигнал на выходе линейной оптической системы можно получить суммированием составляющих спектра $E(u, v)$ входного сигнала, взятых с весом $H(u, v)$. Отсюда следует, что передаточная функция линейной оптической системы $H(u, v)$ является весовым множителем, определяющим относительный вклад различных составляющих спектра $E(u, v)$ в сигнал $E'(x', y')$. Формула (9.13) лежит в основе так называемого спектрального метода исследования линейных оптических систем.

Таким образом, любую линейную пространственно-инвариантную оптическую систему можно рассматривать либо в пространственной области с помощью ее импульсной характеристики, либо в области пространственных частот с помощью частотной или передаточной функции. В первом случае на входе системы (в предметной плоскости) формируют сигнал от точечного излучателя, описываемого δ -функцией, и анализируют функцию $h(x', y')$ в плоскости изображения. Во втором случае на вход системы подают гармонический сигнал единичной амплитуды и наблюдают передаточную функцию системы

$H(u, v)$. Оба подхода равноценны и выбор одного из них определяется лишь удобствами получения исходных данных о системе и простотой вычисления.

Метод спектрального разложения (спектральный метод) и метод интеграла суперпозиции используют для описания процесса прохождения сигнала через линейные оптические системы.

Покажем в заключение этого параграфа, что передаточная функция не зависит от сдвигов в координатной области. Спектральная плотность входного сигнала

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(\xi - x_0; \eta - y_0) e^{-i2\pi(u\xi + v\eta)} d\xi d\eta =$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) e^{-i2\pi[u(x+x_0) + v(y+y_0)]} dx dy = E(u, v) e^{-i2\pi(ux_0 + vy_0)}.$$

Подобным образом находим спектральную плотность выходного сигнала: $E'(u, v) e^{-i2\pi(ux_0 + vy_0)}$. При определении отношения этих спектральных плотностей фазовый множитель $\exp[-i2\pi(ux_0 + vy_0)]$ сокращается и мы получаем величину $H(u, v)$, не зависящую от x_0 и y_0 .

9.4. Слой свободного пространства

9.4.1. Импульсная характеристика и передаточная функция слоя свободного пространства

Простейшей линейной пространственно-инвариантной оптической системой, преобразующей входной сигнал $E(x, y)$ в выходной сигнал $E'(x', y')$, является слой свободного пространства, заключенный между двумя параллельными плоскостями XU и $X'U'$. Обычно в этих плоскостях размещаются различные оптические элементы (линзы, диафрагмы, транспаранты и т.п.). Так как слой пространства является

фильтром, то его свойства описываются импульсной характеристикой и передаточной функцией. Найдем эти характеристики.

Пусть в координатной плоскости XY задано распределение поля $E(x, y)$. Тогда распределение поля $E'(x', y') = E'(x, y, z)$ в плоскости $X'Y'$, находящейся на расстоянии z от входной плоскости, найдется с помощью интеграла Кирхгофа – Френеля – Зоммерфельда (5.3). Представим амплитудный весовой множитель $\cos \alpha / r$ в виде $\cos \alpha / r = z / r^2$. Тогда с учетом (5.4) для поля $E'(x', y')$ будем иметь

$$E'(x', y') = \frac{1}{i\lambda} \int_{-\infty}^{\infty} E(x, y) \frac{z}{(x' - x)^2 + (y' - y)^2 + z^2} \times \\ \times \exp(ik \sqrt{(x' - x)^2 + (y' - y)^2 + z^2}) dx dy. \quad (9.14)$$

Полученный интеграл представляет собой интеграл свертки входного сигнала $E(x, y)$ с функцией

$$h(x, y) = \frac{1}{i\lambda} \frac{z}{x^2 + y^2 + z^2} \exp(ik \sqrt{x^2 + y^2 + z^2}). \quad (9.15)$$

Эта функция и является импульсной характеристикой слоя свободного пространства толщиной z . Следовательно, интеграл суперпозиции Кирхгофа – Френеля – Зоммерфельда можно записать в виде

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) h(x - x'; y - y') dx dy.$$

Это соотношение показывает, что распространение оптического сигнала в свободном пространстве описывается линейным преобразованием или, другими словами, распространение световой волны в однородной среде можно рассматривать как преобразование сигнала в линейной пассивной

системе. Это соотношение показывает также, что поле в плоскости $X'Y'$ определяется сверткой поля в плоскости XY с импульсной характеристикой слоя свободного пространства.

Физический смысл полученной импульсной характеристики состоит в том, что точечный источник света, расположенный в начале координат ($x = 0, y = 0$), создает сферическую волну, которая в плоскости $X'Y'$ вызывает изменение фазы сигнала по гиперболическому закону при медленном уменьшении амплитуды с увеличением расстояния от оси Z . Как и следовало ожидать, импульсная характеристика симметрична относительно оси Z .

Для определения передаточной функции свободного пространства, в соответствии с (9.8), следует найти фурье-образ функции (9.15):

$$H(u, v) = \frac{1}{i\lambda} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{z}{x^2 + y^2 + z^2} \exp[ik(\sqrt{x^2 + y^2 + z^2} - ux - vy)] dx dy .$$

Вычисление этого интеграла методом стационарной фазы дает следующее простое выражение для передаточной функции свободного пространства:

$$H(u, v) = \exp(i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}), \quad (9.16)$$

где учтено, что волновое число $k = 2\pi / \lambda$.

К выражению (9.16) для передаточной характеристики свободного пространства можно придти и иначе. Во входной плоскости XY поле характеризуется частотным спектром $E(u, v)$ (или $E(u, v; 0)$), в выходной плоскости $X'Y'$ – спектром $E'(u, v)$ (или $E(u, v; z)$, где координату z можно рассматривать в качестве параметра). В результате распространения поля на расстояние z (от входной плоскости XY до выходной плоскости $X'Y'$) его фурье-спектр претерпевает изменение по фазе, определяемое

экспоненциальным

множителем

$\exp(ik_z z) = \exp(i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2})$. Таким образом,

$$E'(u, v) = E(u, v) \exp(i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}).$$

Учитывая соотношение (9.11), приходим к выражению (9.16).

Заметим, что, в соответствии с этим выражением, распространение волн в свободном пространстве на расстояние z проявляется только в изменении относительных фаз различных составляющих углового спектра. Фазовые сдвиги возникают вследствие того, что плоские волны, распространяясь под различными углами, проходят различные расстояния, пока достигнут рассматриваемой точки.

Представляет интерес также получить выражение (9.16) и из решения волнового уравнения Гельмгольца для свободного пространства с помощью двумерного преобразования Фурье. Умножим волновое уравнение Гельмгольца

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} + k^2 E = 0$$

на выражение $\exp[-i2\pi(ux + vy)]$ и проинтегрируем по x и y в бесконечных пределах. Вычисление первых двух интегралов дает соответственно выражения $-u^2 E(u, v; z)$ и $-v^2 E(u, v; z)$. Третье слагаемое приводится к виду $d^2 E(u, v, z)/dz^2$, а четвертое – к виду $(2\pi/\lambda)^2 E(u, v, z)$. С учетом этого приходим к уравнению

$$\frac{d^2 E(u, v, z)}{dz^2} + 4\pi^2(1/\lambda^2 - u^2 - v^2)E(u, v, z) = 0.$$

Решением этого обыкновенного дифференциального уравнения второго порядка является функция вида

$$E(u, v, z) = C_1(u, v) \exp(i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}) +$$

$$+ C_2(u, v) \exp(-i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}),$$

где C_1 и C_2 – постоянные интегрирования (они не зависят от координаты z , но могут зависеть от пространственных частот u и v). Первый член в этом выражении соответствует прямой волне, а второй – обратной. В свободном пространстве обратная волна отсутствует, поэтому второй член должен быть равен нулю. Учитывая далее, что в плоскости $z = 0$ пространственный спектр распределения поля есть $E(u, v; 0)$, найдем, что $C_1(u, v) = E(u, v; 0)$. Тогда получим

$$E(u, v; z) = E(u, v; 0) \exp(i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}).$$

Следовательно, для определения спектра выходного сигнала $E(u, v; z)$ нужно спектр входного сигнала $E(u, v; 0)$ умножить на функцию частот u и v :

$$H(u, v) = \exp(i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}),$$

которой и является передаточная функция свободного пространства.

Заметим, что передаточная функция слоя свободного пространства представляет собой фазовый множитель $\exp(ik_z z) = \exp(i2\pi w z)$ в выражении плоской волны.

Выражение (9.16) показывает, что слой свободного пространства играет роль оптической системы, фильтрующей пространственные частоты. Область пропускания этого фильтра определяется условием

$$u^2 + v^2 < 1/\lambda^2. \quad (9.17)$$

В плоскости пространственных частот это соотношение определяет круг радиуса $1/\lambda$. Соотношение (9.17) представляет собой условие распространения световых волн с частотами u и v . Если это условие не выполняется, квадратный корень в

экспоненте (9.16) становится мнимым, а сама экспонента – действительной величиной с отрицательным показателем степени. Это приводит к появлению в амплитуде плоской волны (1.30) экспоненциально убывающего множителя $\exp(\alpha z)$, где

$$\alpha = 2\pi\sqrt{u^2 + v^2 - 1/\lambda^2}.$$

Такие волны (их называют неоднородными или поверхностными волнами) могут распространяться вдоль плоскости XU и быстро затухают (практически на расстоянии нескольких длин волн λ) при удалении от этой плоскости (с ростом z). Передаточная функция $H(u, v)$ при $z \gg \lambda$ и невыполнении условия (9.17) близка к нулю. Поэтому пропускание свободного пространства вне круга радиуса $1/\lambda$ близко к нулю. Отсюда следует, что волны, частоты u и v которых не удовлетворяют условию (9.17), на расстоянии $z \gg \lambda$ от входной плоскости не содержат никакой информации о пространственных частотах оптического поля $E(x, y)$. Чем больше длина участка свободного пространства z , тем меньше сквозь него проходит пространственных частот.

Поскольку частотная характеристика слоя свободного пространства является величиной комплексной, она может быть представлена в виде

$$H(u, v) = |H(u, v)| \exp[i\varphi(u, v)].$$

При этом модуль этой функции $|H(u, v)|$ определяет амплитудно-частотную, а аргумент

$$\varphi(u, v) = \exp(i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2})$$

– фазо-частотную характеристики слоя свободного пространства. Внутри круга $u^2 + v^2 < 1/\lambda^2$ АЧХ

$$|H(u, v)| = 1,$$

а ФЧХ

$$\varphi(u, v) = \exp(i2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}),$$

а вне этого круга АЧХ

$$|H(u, v)| = \exp(-2\pi z \sqrt{u^2 + v^2 - 1/\lambda^2}),$$

ФЧХ

$$\varphi(u, v) = 0.$$

В приближении Френеля, когда $x^2 + y^2 \ll z^2$, импульсную характеристику свободного пространства можно получить, учтя, что в приближении Френеля $\sqrt{x^2 + y^2 + z^2} \approx z + \frac{x^2 + y^2}{2z}$, и положив в знаменателе предэкспоненциального множителя в (9.15) $x^2 + y^2 \approx z^2$. Будем иметь

$$\begin{aligned} h(x, y) &= \frac{e^{ikz}}{i\lambda z} \exp\left(ik \frac{x^2 + y^2}{2z}\right) = \\ &= \frac{\exp(i2\pi z/\lambda)}{i\lambda z} \exp\left[i \frac{\pi}{\lambda z} (x^2 + y^2)\right]. \end{aligned} \quad (9.18)$$

Это соотношение можно записать и в такой форме:

$$h(x, y) = C_1 \exp\left(ik \frac{x^2 + y^2}{2z}\right), \quad (9.19)$$

где коэффициент $C_1 = \frac{\exp(ikz)}{i\lambda z}$ не зависит от координат x и y .

Как видим, фаза импульсной характеристики слоя свободного пространства изменяется по квадратичному закону относительно переменных x и y , а ее амплитуда уменьшается обратно пропорционально толщине слоя z . От координат x и y амплитуда импульсной характеристики слоя свободного пространства не

зависит. Заметим, что импульсная характеристика слоя свободного пространства представляет собой расходящуюся сферическую волну с радиусом кривизны, равным толщине слоя z . Отметим также, что свойства слоя свободного пространства определяются только одним параметром – его глубиной z .

Передаточную функцию слоя свободного пространства в приближении Френеля найдем, подставив (9.18) в (9.8):

$$\begin{aligned}
 H(u, v) &= \frac{\exp(i2\pi z / \lambda)}{i\lambda z} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} \exp\left[i \frac{\pi}{\lambda z} (x^2 + y^2)\right] \times \\
 &\quad \times \exp[-i2\pi(ux + vy)] dx dy = \\
 &= \exp\left(i \frac{2\pi z}{\lambda}\right) \exp[-i\pi\lambda z(u^2 + v^2)]
 \end{aligned}$$

или

$$H(u, v) = C_2 \exp[-i\pi\lambda z(u^2 + v^2)], \quad (9.20)$$

где $C_2 = \exp(i2\pi z / \lambda)$ – коэффициент, не зависящий от переменных u и v . При получении конечного выражения мы воспользовались значением интеграла

$$\begin{aligned}
 &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[-\alpha (x^2 + y^2) \pm \beta_1 x \pm \beta_2 y] dx dy = \\
 &= \frac{\pi}{\alpha} \exp\left(\frac{\beta_1^2 + \beta_2^2}{4\alpha}\right) \quad (9.21)
 \end{aligned}$$

при $\alpha = \pi / (i\lambda z)$, $\beta_1 = i2\pi u$, $\beta_2 = i2\pi v$.

Выражение (9.19) для $H(u, v)$ в приближении Френеля можно получить и непосредственно из общего выражения (9.16), полагая, что в этом приближении

$$\sqrt{1/\lambda^2 - u^2 - v^2} \approx \frac{1}{\lambda} - \frac{\lambda(u^2 + v^2)}{2}.$$

Первый экспоненциальный множитель $\exp(i2\pi z/\lambda)$ в передаточной функции (9.20) описывает общее изменение фазы для каждой компоненты спектра, обусловленное прохождением ее от входной плоскости до выходной плоскости, расстояние между которыми равно z . Этот множитель – постоянный в плоскости XU , поэтому в выражениях (9.18) и (9.19) его можно опустить. Второй экспоненциальный множитель $\exp[-i\pi\lambda z(u^2 + v^2)]$ описывает фазовую зависимость волны от пространственных частот (фазовую дисперсию) по квадратичному (параболическому) закону. Как видим, в приближении Френеля для всех частот амплитудно-частотная характеристика слоя свободного пространства $|H(u, v)|$ постоянна и равна единице, а его фазо-частотная характеристика $\varphi(u, v)$ изменяется по закону $\exp[-i\pi\lambda z(u^2 + v^2)]$.

Частотную характеристику слоя свободного пространства можно также представить в виде

$$H(u, v) = C_2 \exp\left(-i4\pi^2 z \frac{u^2 + v^2}{2k}\right) \quad (9.22)$$

с тем же коэффициентом $C_2 = \exp(i2\pi z/\lambda)$, не зависящим от частот u и v . Заметим, что обе характеристики слоя свободного пространства – импульсная и частотная определяются только одним параметром – его толщиной z .

9.4.2. Решение интегрального уравнения распространения световой волны в свободном пространстве

Покажем здесь, что интегральное уравнение распространения волн (5.10)

$$E(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y, 0) \exp(i 2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}) dudv,$$

как и интегральное уравнение (преобразование) Фурье, обладает свойством обратимости, т.е. зная распределение поля $E(x, y, z)$ в плоскости $z = \text{const} > 0$, можно найти спектр исходного распределения (в плоскости $z = 0$), а, следовательно, и само исходное распределение $E(x, y, 0)$. Введем оптический фильтр с передаточной функцией

$$H'(u, v) = \exp(-i 2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}), \quad (9.23)$$

являющейся комплексно сопряженной передаточной функции слоя свободного пространства. Пусть распределение $E(x, y, z)$ преобразуется указанным фильтром. Действие фильтра выразится в том, что функция спектральной плотности $E(u, v)$ умножится на частотную характеристику фильтра $H'(u, v)$ и тогда вместо спектральной функции $E(u, v)$ получаем спектральную функцию

$$E(u, v) H'(u, v) = E(u, v) \exp(-i 2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}).$$

Подставляя это в уравнение (5.8)

$$E(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) e^{i 2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}} e^{i 2\pi (ux + vy)} dudv,$$

получим

$$\begin{aligned} E(x, y, z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) e^{-i 2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}} e^{i 2\pi z \sqrt{1/\lambda^2 - u^2 - v^2}} e^{i 2\pi (ux + vy)} dudv = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) e^{i 2\pi (ux + vy)} dudv = E(x, y, 0). \end{aligned}$$

Следовательно, действие фильтра, имеющего частотную характеристику вида (9.23), переводит выход во вход. Этого и следовало ожидать, так как комплексно сопряженной передаточной функции (9.23) соответствует комплексно сопряженная амплитуда плоской волны, которая описывает распространение волны в обратном направлении. Полученный результат и позволяет решить интегральное уравнение распространения волн вида (5.10).

Комплексно сопряженной передаточной функции $H'(u, v)$ соответствует, очевидно, и комплексно сопряженная импульсная характеристика. Для ее получения в любой формуле, определяющей импульсную характеристику слоя свободного пространства, следует заменить параметр z на $-z$ (что соответствует замене i на $-i$). Тогда импульсная характеристика рассматриваемого фильтра будет иметь вид

$$h'(x, y) = \frac{1}{i\lambda} \cdot \frac{-z}{x^2 + y^2 + z^2} \exp(-ik\sqrt{x^2 + y^2 + z^2}) \quad (9.24)$$

– в общем случае или

$$h'(x, y) = \frac{\exp\left(-\frac{2\pi z}{\lambda}\right)}{-i\lambda z} \exp\left[-i\frac{\pi}{\lambda z}(x^2 + y^2)\right] \quad (9.25)$$

– в приближении Френеля. Принимая распределение $E(x, y, z)$ за входной сигнал, а распределение $E(x, y, 0)$ за выходной, на основании интеграла суперпозиции получим

$$E(x, y, 0) = \int_{-\infty-\infty}^{\infty \infty} E(x', y', z) h'(x - x'; y - y') dx' dy' .$$

Подставляя сюда выражение (9.15), найдем $E(x, y, 0)$:

$$E(x, y, 0) = \frac{1}{\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y, z) \times \\ \times \frac{z \exp(-ik \sqrt{(x-x')^2 + (y-y')^2 + z^2})}{(x-x')^2 + (y-y')^2 + z^2} dx' dy'. \quad (9.26)$$

В приближении Френеля

$$E(x, y, 0) = \\ = \frac{e^{-ikz}}{i\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y, z) \exp \left[-ik \frac{(x-x')^2 + (y-y')^2}{2z} \right] dx' dy'. \quad (9.27)$$

Соотношения (9.26) и (9.27) и решают задачу обратимости интегрального уравнения в представлении Рэлея: (9.26) – в общем случае, а (9.27) – в приближении Френеля.

9.5. Линейность оптической системы при различных видах освещения

Пусть в оптической системе, освещаемой когерентным светом, входным комплексным амплитудам $E_1(x, y)$ и $E_2(x, y)$ соответствуют выходные комплексные амплитуды $E'_1(x, y)$ и $E'_2(x, y)$. Если система линейна, то комплексной амплитуде $E_1(x, y) + E_2(x, y)$ на входе должна соответствовать комплексная амплитуда $E'_1(x, y) + E'_2(x, y)$ на выходе. Выходная интенсивность $I' = (E'_1 + E'_2) \cdot (E_1^* + E_2^*)$ будет при этом иметь вид многочлена:

$$I' = |E'_1|^2 + |E'_2|^2 + E'_1 E_2^* + E_1^* E'_2 =$$

$$= I'_1 + I'_2 + E'_1 E_2'^* + E_1'^* E', \quad (9.28)$$

где I'_1 – выходная интенсивность при действии на входе только сигнала E_1 , а I'_2 – выходная интенсивность при действии только сигнала E_2 . При когерентном освещении все четыре члена в (9.28), вообще говоря, не равны нулю. Следовательно, в этом случае система не линейна по интенсивности. Однако при некогерентном освещении выходная интенсивность складывается только из выходных интенсивностей: $I' = I'_1 + I'_2$, т.е. системы, освещаемые некогерентным светом, линейны по интенсивности.

Различие между этими случаями является фундаментальным и составляет основу обсуждения процесса образования изображения, рассматриваемого в главах 11 и 12.

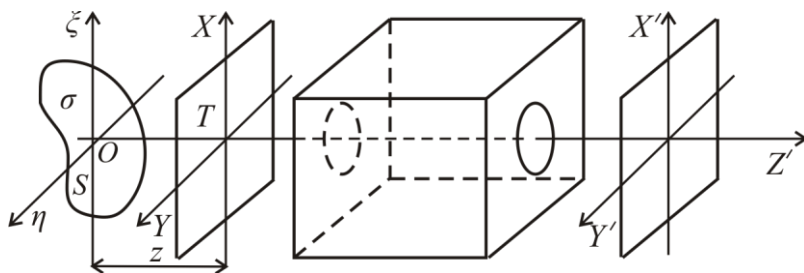


Рис. 9.3

Рассмотрим теперь оптическую систему (рис. 9.3), состоящую из протяженного источника света S , транспаранта T , расположенного в плоскости XY , оптической системы формирования изображения (на рисунке показана в виде прямоугольного параллелепипеда с входным и выходным отверстиями) и выходной плоскости (плоскости изображения) XY' . Разобьем мысленно источник света на элементарные излучатели площадью $d\sigma$. Пусть $E_0(x, y)$ – комплексная амплитуда поля в плоскости непосредственно перед транспарантом, обусловленного элементарными источниками

света. Если функция пропускания транспаранта по амплитуде $t(x, y)$, то комплексное световое поле в плоскости непосредственно за транспарантом будет $E(x, y) = E_0(x, y)t(x, y)$. Предположим, что оптическая система линейна, пространственно-инвариантна и имеет пространственную импульсную характеристику $h(x, y; x', y')$. Тогда комплексное световое поле в выходной плоскости $X'Y'$, обусловленное элементарным источником, можно описать уравнением свертки

$$E'(x', y') = E(x, y) * h(x, y; x', y') = E_0(x, y)t(x, y) \times h(x, y; x', y'). \quad (9.29)$$

Интенсивность в выходной плоскости от элементарного источника света

$$dI'(x', y') = |E'(x', y')|^2 d\sigma = E'(x', y')E'^*(x', y')d\sigma, \quad (9.30)$$

а полная интенсивность в выходной плоскости, создаваемая всем источником света, будет

$$I'(x', y') = \iint_{\sigma} |E'(x', y')|^2 d\sigma, \quad (9.31)$$

где интегрирование проводится по всей площади σ источника света S . Выражение (9.31) можно записать в виде интеграла свертки

$$I'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Gamma(x, y; x_1, y_1) h(x' - x; y' - y) \times h^*(x' - x_1; y' - y_1) t(x, y) t^*(x_1, y_1) dx dy dx_1 dy_1, \quad (9.32)$$

где

$$\Gamma(x, y; x_1, y_1) = \iint_{\sigma} E_0(x, y) E_0^*(x_1, y_1) d\sigma$$

представляет собой функцию взаимной пространственной когерентности. Эта функция характеризует статистическую связь полей в плоскости предмета. В параксиальном приближении функция $\Gamma(x, y)$ определяется соотношением (4.25).

$$\Gamma(x, y) = e^{i\psi_1} \iint_S I_S(\xi, \eta) \exp[-i2\pi(u\xi + v\eta)] d\xi d\eta,$$

где $I_S(\xi, \eta)$ – интенсивность источника света, приходящаяся на единицу площади его поверхности на единичном расстоянии от него, $\psi_{12} = k(x^2 + y^2)/2z$, $u = x/(\lambda z)$, $v = y/(\lambda z)$, ξ, η – координаты в плоскости источника света.

Рассмотрим два предельных случая.

1. Источник света пренебрежимо мал, т.е. $I_S(\xi, \eta) = \delta(\xi, \eta)$. В этом случае

$$\Gamma(x, y) = C_1, \tag{9.33}$$

где C_1 – положительная постоянная. Функция взаимной пространственной когерентности $\Gamma(x, y)$ отлична от нуля для любой пары точек в плоскости предмета. Этот случай соответствует полностью пространственно когерентному освещению (или, как говорят, полностью когерентной оптической системе), так как в этом случае имеет место статистическая связь между полями в плоскости предмета. Такое освещение можно получить от точечного источника света, расположенного в фокусе собирающей линзы без aberrаций, поставленной перед транспарантом. С этой целью можно использовать и пучок лазерного излучения, который, как известно, обладает хорошей пространственной когерентностью даже при искусственном расширении пучка. В этом случае транспарант освещается коллимированным пучком света, т.е. плоской волной. Распределение поля в плоскости непосредственно за транспарантом будет пропорционально его функции пропускания по амплитуде $t(x, y)$. При полностью когерентном освещении

$$\begin{aligned}
I'(x', y') &= \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} h(x' - x; y' - y) t(x, y) dx dy \times \\
&\times \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} h^*(x' - x_1; y' - y_1) t^*(x_1, y_1) dx_1 dy_1 = \\
&= E'(x', y') \cdot E^*(x', y').
\end{aligned} \tag{9.34}$$

Отсюда получаем

$$E'(x', y') = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) h(x' - x; y' - y) dx dy, \tag{9.35}$$

где $E(x, y)$ пропорционально $t(x, y)$ – распределение комплексной амплитуды в предметной плоскости. Следовательно, в случае полностью когерентного освещения оптическая система линейна по комплексной амплитуде. При этом комплексная амплитуда на выходе определяется сверткой комплексной амплитуды на входе с импульсной характеристикой системы.

2. Источник света бесконечно большой, а его излучение однородно, т.е. $I_S(\xi, \eta) = \text{const}$. В этом случае функция взаимной пространственной когерентности

$$\Gamma(x, y) = C_2 \delta(x, y), \tag{9.36}$$

где C_2 – комплексная постоянная. Это означает отсутствие статистической связи между полями точечных источников в плоскости предмета. Значение функции $\Gamma(x, y)$ отлично от нуля лишь в точках (x, y) , поэтому выражение (9.36) соответствует полностью некогерентному освещению (описывает полностью некогерентную оптическую систему формирования изображения). Такое освещение можно получить от протяженного источника с помощью подходящего коллиматорного устройства. Следовательно, в случае полностью некогерентного освещения распределение интенсивности в выходной плоскости записывается в виде

$$I'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x_1 - x; y_1 - y) h(x' - x; y' - y) \times \\ \times h^*(x' - x_1; y' - y_1) t(x, y) t^*(x_1, y_1) dx dy dx_1 dy_1$$

или

$$I'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |t(x, y)|^2 |h(x' - x; y' - y)|^2 dx dy. \quad (9.37)$$

Стоящая в подынтегральном выражении (9.37) функция $|t(x, y)|^2$ называется функцией пропускания транспаранта по интенсивности. Она будет пропорциональна распределению интенсивности поля в предметной плоскости $I(x, y)$. С учетом этого интеграл (9.37) можно переписать в виде

$$I'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) |h(x' - x; y' - y)|^2 dx dy,$$

или

$$I'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) F_0(x' - x; y' - y) dx dy. \quad (9.38)$$

Здесь $F_0(x, y) = |h(x, y)|^2$. Функцию $F_0(x, y)$ называют импульсной характеристикой оптической системы по интенсивности, некогерентной переходной функцией оптической системы, а также функцией рассеяния точки (сокращенно ФРТ). Эта функция описывает распределение интенсивности в изображении светящейся точки. Поскольку представляет интерес лишь относительное распределение интенсивности I , оказывается удобным нормировать ФРТ на единицу, т.е. положить

$$\int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} F_0(x, y) dx dy = 1. \quad (9.39)$$

Таким образом, в случае полностью некогерентного освещения интенсивность поля в выходной плоскости представляет собой свертку интенсивности входного сигнала с импульсной характеристикой системы по интенсивности. В случае полностью некогерентного освещения оптическая система является линейной относительно интенсивности.

В отличие от когерентного сигнала $E'(x', y')$ некогерентный сигнал $I'(x', y')$ описывается не комплексной, а действительной функцией. Эта функция несет информацию, заложенную в интенсивности световой волны. Следовательно, некогерентный сигнал несет только амплитудную информацию; фазовая информация утрачивается.

Таким образом, полностью когерентная оптическая система является линейной по амплитуде, а полностью некогерентная – по интенсивности. Общее выражение (9.32) соответствует частично-когерентному освещению.

Используя коммутативное свойство свертки, получим другое, более удобное, выражение для интенсивности поля в выходной плоскости:

$$I'(x', y') = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} F_0(x, y) I(x' - x; y' - y) dx dy. \quad (9.40)$$

9.6. Оптическая передаточная функция оптической системы

Применив теорему свертки к интегралу (9.39), получим для преобразования Фурье от распределения интенсивности в выходной плоскости:

$$I'(u, v) = A(u, v)I(u, v), \quad (9.41)$$

где $I(u, v)$ – преобразование Фурье от распределения интенсивности на объекте, $A(u, v)$ – преобразование Фурье от функции рассеяния точки:

$$A(u, v) = \int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} F_0(x, y) e^{-i2\pi(ux+vy)} dx dy. \quad (9.42)$$

Комплексная функция $A(u, v)$ пространственных частот u и v называется частотным откликом оптической системы, или оптической передаточной функцией (сокращенно ОПФ) оптической системы.

К понятию оптической передаточной функции можно прийти и рассматривая, как действует линейная по интенсивности пространственно-инвариантная оптическая система на гармонический сигнал $I(x, y) = \text{Re}\{\exp[i2\pi(ux + vy)]\}$. Подставляя это выражение в (9.39) и опуская символ Re , получим

$$\begin{aligned} I'(x', y') &= \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_0(x, y) e^{-i2\pi(ux+vy)} dx dy \right) e^{i2\pi(ux'+vy')} = \\ &= A(u, v) e^{i2\pi(ux'+vy')}. \end{aligned} \quad (9.43)$$

В полученном соотношении комплексно-амплитудный множитель $A(u, v)$ и представляет собой ОПФ оптической системы. Это соотношение показывает, что линейная по интенсивности оптическая система действует на гармонический сигнал так же, как и оптическая система, линейная по комплексной амплитуде: гармонический сигнал на входе такой системы преобразуется в гармонический же сигнал тех же пространственных частот на выходе.

Учитывая, что $F_0(x, y) = |h(x, y)|^2 = h(x, y)h^*(x, y)$ и используя теорему умножения для преобразования Фурье, выражение (9.42) можно переписать в виде

$$\begin{aligned}
A(u, v) &= \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} h(x, y) h^*(x, y) e^{-i2\pi(ux+vy)} dx dy = \\
&= \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} H(u_1, v_1) H^*(u_1 + u, v_1 + v) du_1 dv_1 = \\
&= H(u, v) * H(-u, -v),
\end{aligned}$$

так как $H^*(u, v) = H(-u, -v)$. Полученная величина есть автокорреляция функции $H(u, v)$. В символической форме ее можно записать так:

$$A(u, v) = H(u, v) \otimes H(u, v).$$

Следовательно, оптическая передаточная функция некогерентной оптической системы представляет собой функцию автокорреляции передаточной функции когерентной системы.

Согласно соотношению (9.41), частотный спектр распределения интенсивности в выходной плоскости оптической системы представляет собой произведение спектра частот распределения интенсивности по объекту на частотный отклик оптической системы. Мы видим, что каждой компоненте $I(u, v)$ соответствует коэффициент (комплексный), на который эта компонента умножается при переходе от объекта к изображению. Функцию $I(x, y)$ можно разложить в общем случае в интеграл Фурье, т.е. представить распределение интенсивности на объекте в виде совокупности бесконечного множества различных гармонических составляющих. Поэтому функция $A(u, v)$ определяет, каким образом каждая гармоническая составляющая спектра пространственных частот распределения интенсивности на объекте передается оптической системой, т.е. как действует оптическая система в частотной области.

Как комплексную функцию ОПФ можно представить в виде

$$A(u, v) = T(u, v) e^{-i\varphi(u, v)}. \quad (9.44)$$

Здесь модуль комплексной функции, т.е. величина $T(u, v) = |A(u, v)|$, характеризует изменение амплитуды каждой гармонической составляющей спектра объекта, а аргумент $\varphi(u, v)$ – изменение ее фазы. Функцию $T(u, v)$ пространственных частот называют частотно-контрастной характеристикой (ЧКХ) оптической системы, а функцию $\varphi(u, v)$ – ее частотно-фазовой характеристикой (ЧФХ), или фазовой передаточной функцией (ФПФ).

Соотношения (9.41) и (9.44) показывают, что действие линейной по интенсивности оптической системы на каждую гармоническую составляющую спектра объекта сводится лишь к изменению амплитуды и фазы этой составляющей.

Из выражения (9.42) непосредственно вытекают следующие свойства оптической передаточной функции.

1. $A(0, 0) = 1$. Это следует из того, что при $u = v = 0$ интеграл (9.39) превращается в интеграл нормировки функции $F_0(x, y)$ (равен единице, соотношение (9.39)).

2. $A(-u, -v) = A^*(u, v)$, где $A^*(u, v)$ – функция, комплексно сопряженная функции $A(u, v)$. Это следует из того, что мнимая единица i и пространственные частоты u и v входят в подынтегральную функцию (9.42) в виде произведений iu и iv , поэтому замена u на $-u$ и v на $-v$ эквивалентна замене i на $-i$, а значит, замене функции A на комплексно-сопряженную функцию A^* .

3. Можно показать также, что $|A(u, v)| \leq |A(0, 0)|$; ОПФ является убывающей функцией пространственных частот.

Зная ОПФ оптической системы, из (9.42) с помощью обратного преобразования Фурье можно найти функцию рассеяния точки:

$$F_0(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(u, v) e^{i2\pi(ux+vy)} dudv. \quad (9.45)$$

В главе 11 эта формула будет использована нами для вычисления ФРТ некоторых дифракционно-ограниченных систем (идеальных систем без aberrаций, в которых имеют место только дифракционные эффекты).

Таким образом, аналогично соотношению (9.8), в случае линейных по интенсивности изопланарных оптических систем вводится некогерентная передаточная функция – ОПФ. По этой причине передаточную функцию $H(u, v)$ часто называют когерентной передаточной функцией (КПФ) оптической системы. По аналогии с импульсной характеристикой когерентной оптической системы для систем, линейных по интенсивности, вводится функция рассеяния точки.

Существуют различные методы теоретической оценки и экспериментального определения ОПФ, но все они достаточно сложные. Только в случае дифракционно-ограниченных оптических систем со зрачками несложной формы и в пренебрежении aberrациями задача определения ОПФ системы может быть решена аналитически. Аналитические выражения ОПФ таких систем будут получены в следующей главе.

9.7. Функции рассеяния оптической системы

При оценке качества оптической системы, формирующей изображение, важно знать реакцию системы на простейшие воздействия в виде светящейся точки, линии (узкой щели) и края. Функции, описывающие распределение интенсивности на изображении точки, линии и края полуплоскости (т.е. края идеально прямолинейного ножа), называются функциями рассеяния соответственно точки, линии и края.

Введенную выше функцию, описывающую распределение интенсивности в изображении светящейся точки, – ФРТ – можно определить как интенсивность светового поля в точке (x', y') изображения, которое возникает от изображения точечного объекта, описываемого δ -образной интенсивностью в точке $(x,$

у). Действительно, полагая в интеграле суперпозиции (9.40) $I(x' - x; y' - y) = \delta(x' - x; y' - y)$ и учитывая фильтрующее свойство дельта-функции, получим для интенсивности изображения в точке (x', y') :

$$I'(x', y') = \int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} F_0(x, y) \delta(x' - x; y' - y) dx dy = F_0(x', y').$$

Функция рассеяния точки является двумерной функцией. Это усложняет ее графическое представление и анализ. Но в некоторых случаях она упрощается. Если оптическая система обладает симметрией вращения, т.е. если отклик на точечный источник зависит только от радиуса $r = \sqrt{x^2 + y^2}$, то ФРТ становится функцией одной переменной – радиуса r : $F_0(x, y) = F_0(r)$. Иногда ее можно представить в виде произведения функций одной переменной (ФРТ с разделяющимися переменными):

$$F_0(x, y) = F_{01}(x) \cdot F_{02}(y).$$

Рассмотрим важный частный случай, когда заданное распределение интенсивности не зависит от координаты y . В этом одномерном случае интенсивность $I(x, y) = I(x, 0)$; обозначим ее через $I(x)$. Распределение интенсивности в выходной плоскости при этом также будет зависеть только от одной координаты x' . Интеграл суперпозиции (9.40) в этом случае можно записать в виде

$$I'(x') = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} F_0(x, y) dy \right) I(x' - x) dx$$

или

$$I'(x') = \int_{-\infty}^{\infty} F_1(x) I(x' - x) dx, \quad (9.46)$$

где функция одной переменной x

$$F_1(x) = \int_{-\infty}^{\infty} F_0(x, y) dy \quad (9.47)$$

называется функцией рассеяния линии (ФРЛ). Чтобы установить физический смысл этой функции, рассмотрим отклик линейной оптической системы на входной сигнал, описываемый δ -функцией $I(x) = \delta(x)$, или, иначе, — на щелевой источник, параллельный оси Y . Подставив в интеграл суперпозиции (9.40) $I(x' - x; 0) = \delta(x' - x)$ и учитывая фильтрующее свойство δ -функции, получим

$$I'(x') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_0(x, y) \delta(x' - x) dx dy = \int_{-\infty}^{\infty} F_0(x', y) dy. \quad (9.48)$$

Сравнивая это с (9.47), видим, что

$$F_1(x') = I'(x').$$

К такому же соотношению можно прийти, если исходить сразу из выражения (9.46), положив в нем $I(x' - x) = \delta(x' - x)$:

$$I'(x') = \int_{-\infty}^{\infty} F_1(x) \delta(x' - x) dx = F_1(x').$$

Следовательно, функция рассеяния линии $F_1(x)$ описывает распределение интенсивности в изображении бесконечно длинной светящейся линии (щели). По аналогии с ФРТ ее можно определить как интенсивность в точке x' изображения, которое возникает от изображения линейного объекта, описываемого δ -образной интенсивностью в точке x . Выражение (9.47) показывает, что функция рассеяния линии получается суммированием интенсивностей от всех точек, расположенных вдоль всей линии.

Из условия нормировки ФРТ на единицу вытекает нормировка на единицу и ФРЛ. Действительно, используя (9.39), будем иметь

$$\int_{-\infty}^{\infty} F_1(x) dx = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} F_0(x, y) dy \right) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_0(x, y) dx dy = 1.$$

В оптике функция рассеяния линии $F_1(x)$ используется чаще, чем функция рассеяния точки $F_0(x, y)$, так как в оптической системе бывает проще и быстрее определить отклик $F_1(x)$ от длинной узкой щели, чем отклик $F_0(x, y)$ от точечного источника.

Если оптическая система имеет ФРТ с разделяющимися переменными, т.е. если $F_0(x, y)$ можно представить в виде произведения, $F_0(x, y) = F_{01}(x)F_{02}(y)$, то на основании формулы (9.47) получаем

$$F_1(x) = \int_{-\infty}^{\infty} F_{01}(x)F_{02}(y) dy = F_{01}(x) \int_{-\infty}^{\infty} F_{02}(y) dy = C_1 F_{01}(x),$$

где постоянная C_1 есть значение интеграла по бесконечным пределам от функции $F_{02}(y)$. Аналогично можно получить для распределения интенсивности в изображении линейного источника, расположенного вдоль оси X : $F_1(y) = C_2 F_{02}(y)$, где C_2 – значение интеграла по бесконечным пределам от функции $F_{01}(x)$. Постоянные C_1 и C_2 находятся из условия нормировки. Таким образом, функция рассеяния линии систем, имеющих ФРТ с разделяющимися переменными, с точностью до постоянного множителя равна одному из сомножителей ФРТ.

Для систем с ФРТ, обладающей круговой симметрией, отклик на щелевой источник, как и в общем случае (9.47), определится соотношением

$$F_1(x) = \int_{-\infty}^{\infty} F_0(r) dy, \quad (9.49)$$

т.е. находится простым интегрированием отклика $F_0(r)$ вдоль оси Y . Можно решить и обратную задачу: по известной ФРЛ $F_1(x)$ определить ФРТ $F_0(r)$. С этой целью приведем интеграл в (9.49) к другому виду. Поскольку подынтегральная функция

$$F_0(r) = F_0(\sqrt{x^2 + y^2})$$

четная, то можно записать:

$$\int_{-\infty}^{\infty} F_0(r) dy = 2 \int_0^{\infty} F_0(r) dy.$$

Учитывая теперь, что

$$y = \sqrt{r^2 - x^2}, \quad dy = \frac{r dr}{\sqrt{r^2 - x^2}}$$

и что при изменении y от 0 до ∞ величина r пробегает значения от x до ∞ , получим

$$F_1(x) = 2 \int_x^{\infty} \frac{r F_0(r)}{\sqrt{r^2 - x^2}} dr.$$

Тем самым мы пришли к преобразованию Абеля. Таким образом, функция рассеяния линии оптической системы, обладающей круговой симметрией, связана с функцией рассеяния точки этой системы преобразованием Абеля. С помощью этого соотношения и известной ФРТ можно найти ФРЛ. Обратную задачу – по известной ФРЛ найти ФРТ – тогда, очевидно, можно решить, используя обратное преобразование Абеля. Заменяя в формуле (8.7) $f(r)$ на $F_0(r)$ и $g(x)$ на $F_1(x)$, для ФРТ оптической системы с круговой симметрией, будем иметь

$$F_0(r) = -\frac{1}{\pi} \frac{\partial}{\partial r} \int_r^{\infty} \frac{r F_1(x)}{x \sqrt{x^2 - r^2}} dx.$$

Таким образом, для оптических систем, обладающих круговой симметрией, отклик $F_1(x)$ на щелевой источник (ФРЛ) и отклик $F_0(r)$ на точечный источник (ФРТ) однозначно связаны между собой: зная одну функцию, можно найти другую функцию.

Задачу определения ФРТ по известной ФРЛ можно решить и не обращаясь к преобразованию Абеля. Введем функцию

$$f(x) = \int_{-\infty}^{\infty} F_1(r) dy.$$

Тогда, как можно показать,

$$F_0(r) = -\frac{1}{2\pi r} \frac{df(r)}{dr}.$$

Наличие множителя $1/2\pi r$ связано с нормировкой функции $F_0(r)$.

Функция рассеяния края (ФРК) $F_2(x)$ описывает распределение интенсивности в изображении края светящейся полуплоскости. Предположим, что край полуплоскости совпадает с осью Y . Тогда распределение интенсивности на крае полуплоскости можно представить функцией единичного скачка (функцией Хевисайда):

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

(при $x = 0$ значение функции $1(x)$ полагается равным $1/2$). Предполагается, что интенсивность на полуплоскости постоянна и равна единице. Заменив в выражении (9.43) $I'(x')$ на $F_2(x)$ и $I(x' - \xi)$ на $1(x' - \xi)$, получим

$$F_2(x) = \int_{-\infty}^x F_1(\xi) d\xi, \quad (9.50)$$

причем с учетом нормировки $F_1(x)$ значение $F_2(0) = 1/2$. Графики распределения интенсивности $I'(x)$ на крае полуплоскости и на его изображении (т.е. ФРК $F_2(x)$) показаны на рис. 9.4.

Согласно (9.50), распределение интенсивности в направлении, перпендикулярном краю полуплоскости $I'(x) = F_2(x)$, описывается суммированием функций рассеяния линий в пределах полуплоскости для каждого значения координаты x :

$$I'(x) = \int_{-\infty}^x F_1(x) dx. \quad (9.51)$$

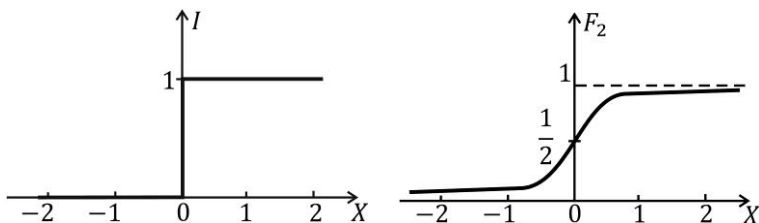


Рис. 9.4

Всю полуплоскость можно мысленно разделить на отдельные линейные элементы, параллельные краю полуплоскости. Каждой

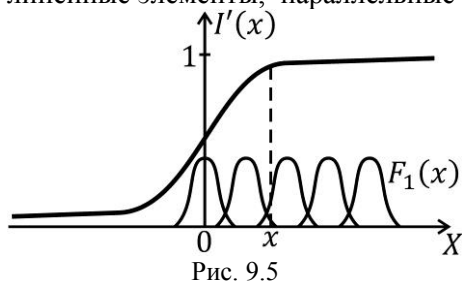


Рис. 9.5

такой светящийся линейный элемент дает свою функцию рассеяния $F_1(x)$. Полное изображение края образуется суммированием функций рассеяния, как это показано на рис. 9.5 для каждой точки x . На

рисунке можно видеть, что $I'(x)$, т.е. сумма ординат множества функций рассеяния в данной точке x , численно равна сумме всех ординат функции рассеяния, расположенной левее точки x – для абсцисс от ∞ (так как, в принципе, функция рассеяния простирается в обе стороны до бесконечности) до x . Это и выражается формулой (9.51)

Соотношение (9.50) позволяет найти ФРК по известной ФРЛ. Используя свойство несобственных интегралов, получим обратное соотношение:

$$F_1(x) = dF_2(x)/dx.$$

Следовательно, ФРЛ является первой производной от ФРК. С помощью этого соотношения можно найти ФРЛ по известной ФРК. Используя коммутативное свойство свертки, выражение (9.46) можно записать и в виде

$$I'(x') = \int_{-\infty}^{\infty} I(x)F_1(x' - x)dx. \quad (9.52)$$

Следует отметить, что суммирование ординат функций рассеяния

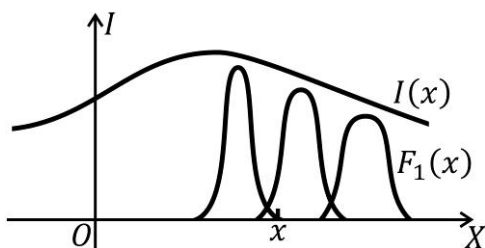


Рис. 9.6

при определении свертки $I'(x')$ отличается от такого же суммирования ФРЛ для определения $I(x) = F_2(x)$. В случае края полуплоскости интенсивность объекта постоянна, поэтому высоты всех суммируемых функций рассеяния одинаковы. В случае же произвольного объекта интенсивность разных его точек различна, и наибольшие высоты отдельных функций рассеяния не равны между собой, а пропорциональны интенсивностям соответствующих точек объекта $I(x)$. При этом, как это видно на рис. 9.6, «ширина» функций $F_1(x)$ уменьшается, так как из-за ее нормировки на единицу, площадь под этой кривой постоянна и равна постоянному значению – единице.

при определении свертки $I'(x')$ отличается от такого же суммирования ФРЛ для определения $I(x) = F_2(x)$. В случае края полуплоскости интенсивность объекта постоянна, поэтому высоты всех суммируемых функций рассеяния одинаковы. В случае же произвольного объекта интенсивность разных его точек различна, и наибольшие высоты отдельных функций рассеяния не равны между собой, а пропорциональны интенсивностям соответствующих точек объекта $I(x)$. При этом, как это видно на рис. 9.6, «ширина» функций $F_1(x)$ уменьшается, так как из-за ее нормировки на единицу, площадь под этой кривой постоянна и равна постоянному значению – единице.

Глава 10

ЦЕНТРИРОВАННЫЕ ОПТИЧЕСКИЕ СИСТЕМЫ ФОРМИРОВАНИЯ ИЗОБРАЖЕНИЯ

10.1. Основные понятия геометрической теории оптических изображений

Если световые лучи, выйдя из светящейся точки S , после прохождения через оптическую систему пересекаются в точке S' , то точку S' называют оптическим изображением или просто изображением точки S . Ее называют также фокусом геометрического схождения лучей. При этом расходящаяся из точки S сферическая волна, пройдя через оптическую систему, превращается в сферическую волну, сходящуюся в точке S' . Очевидно, что в точке S' будет находиться максимум интенсивности света. Каждое самосветящееся или освещенное тело можно представить как совокупность точечных источников света. Такое тело в оптике называют действительным предметом

(объектом) или просто предметом (объектом). Совокупность изображений всех точек светящегося объекта представляет собой изображение этого объекта. Изображение S' называют действительным, если световые лучи действительно пересекаются в точке S' , и мнимым, если в точке S' пересекаются продолжения лучей, проведенные в направлении, обратном распространению света. С волновой точки зрения это означает различие в знаках кривизны волновых поверхностей в направлении распространения света. Простейшим примером мнимого изображения является отражение источника S в плоском зеркале. Изображение S' в этом случае представляет собой точку, расположенную симметрично точке S . Действительное изображение характеризуется тем, что непосредственно освещает соответствующим образом расположенную поверхность; мнимое изображение такого освещения произвести не может. Однако при помощи оптических приспособлений мнимые изображения могут быть преобразованы в действительные. Это имеет место, например, в нашем глазу; полученное изображение освещает определенный участок сетчатки глаза.

Наряду с мнимым изображением вводится понятие и мнимого источника света. Точечный объект называется мнимым, если он является точкой пересечения продолжений действительных лучей, проведенных в обратных направлениях. Совокупность мнимых точечных источников образует мнимый предмет (мнимый объект). Если источник света S и его изображение S' поменять местами, то, как показывает опыт, форма всех лучей останется без изменения, но их направление изменится на противоположное. Останутся без изменения и оптические длины световых лучей. Эти утверждения носят название принципа обратимости световых лучей. Точка S' при этом будет играть роль источника света, а точка S – его изображения. Поэтому точки S и S' называют сопряженными. Аналогично, две линии или две плоскости называют сопряженными, если одна из них является оптическим изображением другой. Непрерывная совокупность точек, изображаемых оптической системой, называется

пространством предмета, а непрерывная совокупность точек, являющихся их изображениями, – пространством изображения.

Когда все лучи пересекаются строго в одной точке S , изображение называется стигматическим (точечным). Пучок лучей, выходящий из одной точки или сходящийся в одну точку, называется гомоцентрическим.

Изображение называется идеальным, если любые две сопряженные кривые C и C' подобны друг другу. Оптическая система, дающая стигматическое (но не обязательно идеальное) изображение трехмерного предмета, называется абсолютным оптическим прибором. Для абсолютного прибора справедлива теорема Максвелла, согласно которой в абсолютном приборе оптическая длина $L(C)$ любой кривой C в пространстве предмета равна оптической длине $L(C')$ изображения C' этой кривой: $L(C) = L(C')$. Из этой теоремы, в частности, следует, что любой треугольник и его изображение в абсолютном приборе подобны друг другу, или, что то же самое, их соответственные углы равны.

Никакие реальные оптические системы абсолютными приборами не являются. В любых оптических системах наблюдаются небольшие отклонения от идеальных условий получения изображения. Однако в теоретическом анализе часто ограничиваются рассмотрением идеальной оптической системы.

Отображения, сохраняющие неизменными углы, называют конформными. Существует общая теорема, согласно которой конформное отображение трехмерного пространства может быть только двух видов – инверсия и проективное преобразование. Поэтому отображение, создаваемое абсолютным прибором, представляет собой либо инверсию, либо проективное преобразование, либо их комбинацию. Инверсией относительно некоторой точки O называется преобразование, которое не изменяет произведение длин отрезков SO и OS' , т.е. если $SO \cdot OS' = \text{inv}$, где S и S' – пара сопряженных точек; точка O называется центром инверсии. В частности, точкой, сопряженной центру O , является бесконечно удаленная точка.

Проективное преобразование, или коллинеация, переводит точку S с координатами x, y, z в точку S' с координатами x', y', z' , определяемыми соотношениями

$$x' = F_1 / F_0, \quad y' = F_2 / F_0, \quad z' = F_3 / F_0, \quad (10.1)$$

где F_0 и F_i – линейные функции координат x, y, z , т.е. $F_0 = ax + by + cz + d$ и $F_i = a_i x + b_i y + c_i z + d_i$, где $a, b, c, d, a_i, b_i, c_i, d_i$ – постоянные коэффициенты; $i = 1, 2, 3$. Значения этих коэффициентов можно определить, если известны положения начал координат и параметры оптической системы. Соотношения (10.1) устанавливают соответствие между точками пространства предметов и пространства изображений. Здесь предполагается, что точки S и S' относятся к разным координатным системам $OXYZ$ и $O'X'Y'Z'$, начала O и O' которых лежат на оптической оси системы (с которой обычно совмещают ось Z) и могут не совпадать друг с другом. Если разрешить эти соотношения относительно x, y, z , получим полностью аналогичные соотношения:

$$x = F_1' / F_0', \quad y = F_2' / F_0', \quad z = F_3' / F_0', \quad (10.2)$$

где F_0' и F_i' – линейные функции переменных x', y', z' с коэффициентами $a', b', c', d', a_i', b_i', c_i', d_i'$, отличными от $a, b, c, d, a_i, b_i, c_i, d_i$. То, что правые части системы уравнений (10.1) являются частными от деления двух линейных функций от x, y, z , а (10.2) – функций от x', y', z' , и обуславливает, что заданным значениям x, y, z соответствует одно, а не несколько систем значений x', y', z' и наоборот, заданным значениям x', y', z' соответствует одна система значений x, y, z . Поскольку знаменателями дробей (10.1) является одна и та же линейная функция F_0 , а знаменателями дробей (10.2) – одна и та же функция F_0' , плоскости

$$ax + by + cz + d = 0 \quad (10.3)$$

соответствует плоскость же

$$a'x' + b'y' + c'z' + d' = 0, \quad (10.4)$$

и наоборот. В этом можно убедиться, если в уравнении плоскости (10.3) не штрихованные координаты заменить на штрихованные по формулам (10.2) или в уравнении плоскости (10.4) штрихованные координаты заменить на не штрихованные по формулам (10.1). Это справедливо и в отношении прямой линии и точки, так как прямую можно рассматривать как линию пересечения двух плоскостей, а точку – как точку пересечения трех плоскостей.

Таким образом, при проективном преобразовании каждая точка, прямая и плоскость пространства предметов изображается соответственно в виде точки, прямой и плоскости пространства изображений. Кроме того, из формулы (10.1) следует, что изображение любой точки, лежащей в плоскости $F_0 = 0$, находится на бесконечности ($x' = y' = z' = \infty$, а из формулы (10.2) – что бесконечно удаленные точки проецируются на плоскость $F_0' = 0$ (при $x = y = z = \infty$ получаем $F_0' = F_1' / x = F_2' / y = F_3' / z = 0$). Это означает, что все лучи, вышедшие из одной и той же точки плоскости $F_0 = 0$, после прохождения через оптическую систему становятся параллельными и наоборот, лучи, параллельные в пространстве предметов, после прохождения через оптическую систему пересекаются в одной и той же точке, расположенной в плоскости $F_0' = 0$. Плоскость $F_0 = 0$ называют фокальной плоскостью в пространстве предметов, или передней фокальной плоскостью оптической системы, а плоскость $F_0' = 0$ – фокальной плоскостью в пространстве изображений, или задней фокальной плоскостью оптической системы.

Если обе фокальные плоскости находятся на бесконечности, преобразования называют аффинными или телескопическими. В этом случае коэффициенты $a = b = c = 0$ и $a' = b' = c' = 0$. Действительно, уравнение фокальной плоскости пространства предметов имеет вид $ax + by + cz + d = 0$; x -координата этой

плоскости $x = -(by + cz + d)/a$. Эта координата будет равна бесконечности, если $a = 0$. Равенство бесконечности y - и z -координат требует, чтобы были равны нулю коэффициенты b и c . Аналогично доказывается равенство нулю и коэффициентов a', b' и c' . Оптические системы, не имеющие фокальных плоскостей на конечном расстоянии, называются афокальными или телескопическими системами. Они являются предельными случаями обычных систем, когда обе фокальные плоскости сдвинуты в бесконечность. Всякий параллельный пучок после прохождения через такую систему остается параллельным, могут изменяться лишь ширина и направление пучка.

С математической точки зрения задача геометрической теории оптических изображений сводится к определению положения изображения при любом заданном положении предмета. В основе этой теории лежит закон прямолинейного распространения света (в однородных средах), а также законы отражения и преломления света на границе раздела сред. Следует, однако, отметить, что полная физическая теория оптических изображений должна учитывать волновые свойства света. Волновая теория позволяет более глубоко и всесторонне изучить структуру оптического изображения, оценить качество изображения, даваемого оптической системой. Как будет показано в главе 11, с точки зрения волновой природы света формирование оптического изображения можно рассматривать как результат дифракционных и интерференционных явлений, возникающих при прохождении света через объект (или при отражении от объекта) и оптическую систему.

С волновой точки зрения под оптическим изображением понимается распределение интенсивности света на плоскости, где формируется изображение. Это непосредственно связано с процессом регистрации изображения, так как любой приемник светового излучения при фиксации изображения реагирует именно на интенсивность.

Общие свойства изображающих оптических систем в геометрической теории исследуются с помощью положения,

согласно которому оптические длины всех лучей, соединяющих сопряженные точки S и S' , одинаковы. Это вытекает из того, что сферическая волна, вышедшая из точки S , превращается в сферическую же волну, сходящуюся в точке S' , а оптические длины всех лучей от одного положения волнового фронта до другого одинаковы. Указанное положение справедливо как в отношении действительных, так и в отношении мнимых изображений, если во втором случае под лучом понимается прямолинейное продолжение его в сторону изображения S' . Оптическая система не только изменяет знак кривизны волновой поверхности сферической волны, исходящей из каждой точки S предмета, превращая ее в сферическую волну, сходящуюся в точке S' , но и ограничивает ее фронт. Изменение знака кривизны волновой поверхности волны соответствует фокусирующему свойству оптической системы, так как волна, выходящая из точки S , соберется в некоторой точке S' .

Ограничение размеров волновых поверхностей световых волн с помощью оправ линз и специальных диафрагм при прохождении света через оптическую систему неизбежно приводит к возникновению дифракции. Дифракция света накладывает основное ограничение на качество изображения предмета. Она же, как будет показано в главе 11, лежит и в основе самого процесса формирования оптического изображения. Процесс образования изображения сводится к двойной дифракции: когерентный пучок света от объекта сначала испытывает дифракцию на отверстии входного зрачка, а изображение является результатом второй дифракции на выходном зрачке или, точнее, – на краях апертурной диафрагмы.

Изображаемый предмет должен быть освещен. Из опыта известно, что свет может иметь определенную окраску, вызывать различные световые ощущения, распространяться в различных направлениях. Падая на предмет, свет отражается или проходит сквозь него, причем это взаимодействие излучения и предмета влияет на само излучение. Параметры световой волны изменяются таким образом, что в них содержится информация о форме

предмета, о состоянии его поверхности, о способности отдельных частей предмета поглощать и естественный белый свет и изменять его спектральный состав. Распространяясь, свет передает эту информацию объектам живой и неживой природы, которые в дальнейшем могут ее обрабатывать, регистрировать и т.п.

Глаз человека как орган зрения воспринимает световую информацию следующим образом. В выпуклой его части находится линза, которая строит изображение предмета. За глазной линзой находится система рецепторов, передающих изображение поэлементно в центральную нервную систему. Рецепторы способны быстро принять информацию об образах и передать ее для дальнейшей обработки в мозговой центр. В рецепторах информация об образе не может долго сохраняться; лишь центральная нервная система способна до определенной степени запомнить черты образа и затем воспроизвести их как представление о предмете. Эти факты лежат в основе создания оптических систем построения изображения. В этих системах используется оптический изображающий элемент, аналогичный глазной линзе. Оптическими изображающими элементами могут являться, прежде всего, линзы и зеркала.

10.2. Центрированные оптические системы

На практике для получения изображения используются так называемые центрированные оптические системы. В таких системах изображение расположено симметрично относительно оптической оси системы. Центрированными системами являются системы, у которых пограничные поверхности отражающих и преломляющих тел представляют собой поверхности вращения с общей осью, или в частности, сферические поверхности, центры которых лежат на одной прямой (ее называют оптической осью). Вследствие симметрии изображение S' точки S должно лежать в плоскости (меридиональной), проходящей через оптическую ось системы и через точку S , поэтому для изучения изображения достаточно знать соотношения между предметом и его

изображением в такой меридиональной плоскости. Если в качестве меридиональной плоскости выбрать плоскость XZ пространства предмета и плоскость XZ' пространства изображения, а оси OZ и OZ' совместить с осью симметрии системы (оптической осью), то из уравнений (10.1) выпадут координаты y и y' . Вследствие симметрии системы величина z' не должна менять своего значения при изменении знака x . Это возможно при выполнении условия $a_3 = a = 0$. Кроме того, из соображений симметрии следует также, что x' меняет знак одновременно с x . Это выражается условием $c_1 = d_1 = 0$. С учетом полученных соотношений уравнения (10.1) приводятся к виду

$$x' = \frac{a_1 x}{cz + d}, \quad (10.5)$$

$$z' = \frac{c_3 z + d_3}{cz + d}. \quad (10.6)$$

Хотя в уравнения (10.5) и (10.6) входят пять постоянных, центрированную систему можно определить четырьмя характеристическими параметрами, в качестве которых можно принять отношения четырех каких-либо параметров к пятому. Отметим, что в центрированных системах за положительное направление оси OZ принимается направление падающего светового луча (слева направо), а за положительное направление оси OZ' – наоборот, т.е. справа налево, противоположно направлению оси OZ . За положительные направления осей OX и OX' принимается направление снизу вверх.

Обращение уравнений (10.5) и (10.6) дает

$$x = \frac{c_3 d - c d_3}{a_1} \cdot \frac{x'}{c_3 - c z'}, \quad (10.7)$$

$$z = \frac{dz' - d_3}{c_3 - cz'}. \quad (10.8)$$

Плоскость FF , определяемая уравнением $cz + d = 0$, представляет собой фокальную плоскость пространства предмета, а плоскость FF' , определяемая уравнением $c_3 - cz' = 0$, – фокальную плоскость пространства изображения. Эти плоскости перпендикулярны оптической оси центрированной системы. Точки их пересечения с оптической осью системы, т.е. точки с z -координатами $z_F = -d/c$ и $z'_{F'} = c_3/c$, называются фокальными точками F и F' , или фокусами оптической системы пространства предмета и пространства изображения соответственно. Их называют также соответственно передним и задним главными фокусами оптической системы. Если за начало отсчета координат z и z' принять фокальные точки z_F и $z'_{F'}$, т.е. если произвести замену $z \rightarrow z - d/c$ и $z' \rightarrow z' + c_3/c$, то $cz + d$ можно заменить на cz , а $c_3 - cz'$ – на $-cz'$. Тогда из уравнений (10.5) и (10.6) получаем

$$zz' = \frac{cd_3 - c_3d}{c^2},$$

$$\frac{x'}{x} = \frac{a_1}{cz}. \quad (10.9)$$

Здесь z и z' представляют собой расстояния предмета и изображения от фокальных плоскостей FF и FF' . Теперь в уравнениях изображения остаются только две характеристические комбинации постоянных: a_1/c и $(cd_3 - c_3d)/c^2$. Введя для обеих комбинаций постоянных более простые обозначения

$$f = a_1/c, \quad f' = (cd_3 - c_3d)/a_1c,$$

для уравнений изображения будем иметь

$$zz' = ff', \quad (10.10)$$

$$x'/x = f/z = z'/f'.$$

Отношение x'/x называется линейным увеличением или просто увеличением оптической системы. Согласно формулам (10.5) или (10.9), оно не зависит от x . Это означает, что изображение плоского предмета, перпендикулярного оптической оси системы, подобно самому предмету. Если увеличение положительное, то изображение прямое; в противоположном случае изображение перевернутое. Увеличение равно единице при $z = f$ (и соответственно при $z' = f'$). Этими соотношениями определяются две плоскости HN и $H'N'$, нормальные оптической оси системы; они называются главными плоскостями системы. Уравнения главных плоскостей можно получить, полагая в соотношениях (10.5) и (10.7) $x = x'$. Это дает

$$cz + d - a_1 = 0; \quad c_3 - cz' - \frac{c_3d - cd_3}{a_1} = 0. \quad (10.11)$$

Первая плоскость называется главной плоскостью пространства предмета, или передней главной плоскостью, а вторая – главной плоскостью пространства изображения, или задней главной плоскостью. Точки их пересечения с оптической осью системы называются главными точками H и H' пространства предмета и пространства изображения соответственно, или соответственно передней и задней главными точками оптической системы. Координаты этих точек найдем, решив уравнения (10.11). Будем иметь

$$z_H = \frac{a_1 - d}{c};$$

$$z'_{H'} = \frac{cd_3 + (a_1 - d)c_3}{a_1c}.$$

Как видим, главные плоскости обладают тем свойством, что любая точка S , расположенная в одной из главных плоскостей, имеет сопряженную точку S' , расположенную в другой главной плоскости, отстоящую от оси системы на таком же расстоянии.

Введенные выше постоянные f и f' называются главными фокусными расстояниями соответственно пространства предмета и пространства изображения. Они определяют расстояния главных плоскостей HN и HN' от соответствующих фокальных плоскостей FF и FF' . Действительно,

$$z_H - z_F = \frac{a_1}{c} = f;$$

$$z'_{H'} - z'_{F'} = \frac{cd_3 - c_3d}{a_1c} = f'.$$

Главное фокусное расстояние f считается положительным, если световые лучи достигают сначала фокальной плоскости FF и затем главной плоскости HN ; для главного фокусного расстояния f' положительным направлением считается обратное направление. Если начала координат поместить в главные точки H и H' и обозначить через a расстояние предмета от главной плоскости HN , а через a' расстояние его изображения от главной плоскости HN' , то, учитывая, что $a = f + z$, $a' = f' + z'$, из уравнения (9.10) получаем

$$\frac{f}{a} + \frac{f'}{a'} = 1. \quad (10.12)$$

Это соотношение есть еще один вид уравнения изображения. Расстояния a и a' считаются положительными, если предмет лежит перед плоскостью HN , а изображение – за плоскостью HN' .

Уравнение (10.12) имеет место и в том случае, если a и a' представляют собой расстояния предмета и изображения от двух

любых сопряженных плоскостей, расположенных нормально к оптической оси, а f и f' – расстояния фокальных точек от этих плоскостей. Этот результат также получается из уравнения (10.9).

Главные и фокальные точки центрированной оптической системы называются ее кардинальными точками. Заметим, что число кардинальных точек равно четырем, т.е. числу характеристических параметров центрированной оптической системы. Кардинальные точки полностью характеризуют оптическую систему в том смысле, что, зная положение этих точек, можно найти изображение любого предмета, даваемое оптической системой.

10.3. Тонкая линза

Простейшей центрированной оптической системой, формирующей изображение, является тонкая линза. Линза – это прозрачное тело (обычно стеклянное), ограниченное двумя сферическими или плоской и сферической поверхностями. Линза является основным элементом любой оптической системы (фотоаппарат, кино- и телекамера, микроскоп, телескоп и др.). Линза называется тонкой, если ее максимальная толщина d_0 много меньше радиусов кривизны R_1 и R_2 сферических поверхностей, образующих линзу. В тонкой линзе поперечным смещением луча можно пренебречь и считать, что луч, входящий в точке с координатами x_0, y_0 на одной поверхности линзы выходит в точке приблизительно с такими же координатами на противоположной поверхности.

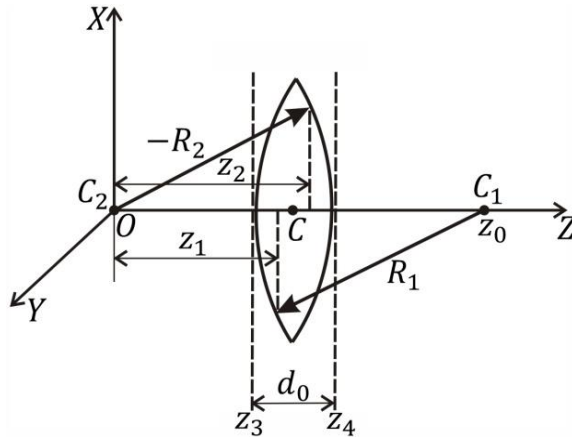


Рис. 10.1

Тонкая двояковыпуклая линза изображена на рис. 10.1. Здесь C – центр линзы, C_1 и C_2 – центры кривизны ограничивающих линзу сферических поверхностей. Прямая, проходящая через центр линзы C и центры кривизны C_1 и C_2 , называется оптической осью линзы. Любой луч, проходящий через центр линзы C , не преломляется, так как угол преломления здесь равен нулю. В тонкой линзе поглощением света, а значит, и изменением амплитуды проходящей световой волны можно пренебречь (отражением света от поверхности линзы также пренебрегаем). Тонкая линза просто задерживает фронт падающей волны на величину, пропорциональную толщине линзы. Поэтому действие тонкой линзы на проходящую через нее волну сводится только к изменению фазы волны. В этом смысле действие тонкой линзы на падающую на нее волну можно рассматривать как действие фазового экрана, модулирующего волну по фазе. Следовательно, поле E' в плоскости непосредственно за линзой будет отличаться от поля E в плоскости непосредственно перед линзой только фазовым множителем $\exp(i\delta)$, где δ – изменение фазы волны, вносимое линзой.

10.3.1. Функция пропускания тонкой линзы

Получим закон изменения фазы волны, прошедшей через тонкую двояковыпуклую собирающую линзу. Поместим в центре кривизны C_2 правой поверхности линзы начало O прямоугольной системы координат, направив ось Z вдоль оптической оси (рис.10.1). Пусть d – толщина линзы в точке (x, y) . При этом $d = z_1 - z_2$, где z_1 и z_2 – z -координаты точек входа луча света в линзу и выхода из нее соответственно. Так как линза тонкая, то d столь мало, что точки z_1 и z_2 почти совпадают. Тогда полное изменение фазы в этой точке после прохождения волны через линзу будет равно

$$\delta = knd + k[d_0 - d] = kd_0 + k(n - 1)d,$$

где n – показатель преломления материала линзы, knd – изменение фазы, вводимое линзой в точке (x, y) ; $k[d_0 - d]$ – изменение фазы за счет оставшейся части свободного пространства между двумя плоскостями $z = z_3$ и $z = z_4$, касательными к левой и правой поверхностям линзы соответственно в точках их пересечения с оптической осью линзы (на рис. 10.1 эти плоскости показаны пунктирами). Поскольку линза тонкая, то эти плоскости тоже почти совпадают.

Условимся считать радиус кривизны каждой выпуклой поверхности, встречающейся на пути луча, распространяющегося слева направо (вдоль оси Z), положительным, а радиус кривизны вогнутой поверхности – отрицательным. Следовательно, радиус кривизны R_1 левой поверхности будет выражаться положительным числом, а радиус кривизны R_2 правой поверхности – отрицательным числом.

Правая поверхность линзы описывается уравнением сферы радиуса $-R_2$: $x^2 + y^2 + z_2^2 = R_2^2$. Откуда

$$z_2 = \sqrt{R_2^2 - x^2 - y^2}.$$

Аналогично левая поверхность описывается уравнением сферы радиуса R_1 и с центром в точке $C_1(0, 0, z_0)$: $x^2 + y^2 + (z_0 - z_1)^2 = R_1^2$. Отсюда находим

$$z_1 = z_0 - \sqrt{R_1^2 - x^2 - y^2}.$$

Тогда получим

$$d = z_2 - z_1 = \sqrt{R_2^2 - x^2 - y^2} - z_0 + \sqrt{R_1^2 - x^2 - y^2}.$$

В приближении парааксиальной оптики, когда рассматривается часть волнового фронта вблизи оптической оси, $x^2 + y^2 \ll R_{1,2}^2$, где $R_{1,2}$ – радиусы кривизны левой и правой поверхностей линзы, будем иметь

$$\begin{aligned} d &\approx -R_2 \left(1 - \frac{x^2 + y^2}{2R_2^2} \right) - z_0 + R_1 \left(1 - \frac{x^2 + y^2}{2R_1^2} \right) = \\ &= R_1 - R_2 - z_0 - \left(\frac{1}{R_1} - \frac{1}{R_2} \right) \frac{x^2 + y^2}{2}. \end{aligned}$$

При получении этого выражения положительное число $-R_2$ было вынесено из-под знака первого радикала.

С учетом этого находим полное изменение фазы световой волны после прохождения ее через линзу

$$\begin{aligned} \delta &= kd_0 + k(n-1)d = \\ &= k[d_0 + (n-1)(R_1 - R_2 - z_0)] - (n-1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right) \frac{x^2 + y^2}{2}. \end{aligned}$$

Величина

$$\Phi = \frac{1}{f} = (n-1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right) \quad (10.13)$$

хорошо известна в геометрической оптике и представляет собой оптическую силу линзы, а f – фокусное расстояние линзы.

Следовательно, тонкую линзу можно описать следующим общим пространственным преобразованием фазы

$$t(x, y) = \exp[i\delta(x, y)].$$

Опуская множитель $\exp\{ik[d_0 + (n-1)(R_1 - R_2 - z_0)]\}$, обуславливающий постоянный фазовый сдвиг для всех x, y , получим

$$t(x, y) = \exp\left(-ik \frac{x^2 + y^2}{2f}\right). \quad (10.14)$$

Поэтому если поле непосредственно перед линзой $E(x, y)$, а поле непосредственно за линзой $E'(x, y)$, то можно записать

$$E'(x, y) = E(x, y) \exp\left(-ik \frac{x^2 + y^2}{2f}\right). \quad (10.15)$$

Функция $t(x, y)$ представляет собой функцию пропускания тонкой линзы. Она является важной характеристикой линзы и широко используется при теоретическом рассмотрении применения линзы для формирования оптического изображения и для фурье-преобразования. Видим, что действие тонкой линзы на поле проходящей световой волны сводится к умножению функции распределения этого поля в плоскости XU непосредственно перед линзой на функцию пропускания линзы.

Тем самым мы показали, что действие тонкой линзы на падающую на нее световую волну можно рассматривать как действие экрана, изменяющего фазу волны (фазового экрана) в соответствии с законом (10.14).

Аналогично можно показать, что для вогнутой линзы фазовое преобразование в пренебрежении несущественной комплексной постоянной равно

$$t(x, y) = \exp\left(ik \frac{x^2 + y^2}{2f}\right). \quad (10.16)$$

Заметим, что существенная разница в фазовых преобразованиях выпуклой и вогнутой линз выражается в положительном и отрицательном знаке квадратичной фазовой задержки. Функция (10.14) описывает линзу с положительным фокусным расстоянием ($f > 0$), а функция (10.16) – с отрицательным фокусным расстоянием ($f < 0$). Поэтому первую линзу называют положительной, а вторую – отрицательной. Положительной линзой является, например, двояковыпуклая линза, а отрицательной – двояковогнутая линза. Это следует из формулы (10.13). Действительно, для двояковыпуклой линзы кривизна левой ограничивающей сферической поверхности отрицательна, а правой – положительна, т.е. $1/R_2 < 0 < 1/R_1$, и тогда $\Phi > 0, f > 0$; для двояковогнутой линзы $1/R_1 < 0 < 1/R_2$ и $\Phi < 0, f < 0$.

При наложении двух линз свет проходит их последовательно: $E_2 = E_1 t_2 = E_0 t_1 t_2$, т.е. функции пропускания перемножаются. Следовательно, две линзы, расположенные вплотную друг к другу, имеют функцию пропускания

$$t(x, y) = \exp\left[-ik \left(\frac{1}{f_1} \pm \frac{1}{f_2}\right) \frac{x^2 + y^2}{2}\right],$$

где знак плюс относится к двум собирающим линзам, а минус – к одной собирающей, а другой рассеивающей линзам. Как видим, в системе из двух положительных линз оптические силы складываются, а в системе из положительной и отрицательной линз – вычитаются.

Таковыми же функциями пропускания описываются и параболические зеркала.

В прикладной оптике, кроме сферических линз, часто используют и цилиндрические линзы. Эти линзы преобразуют плоские волны в цилиндрические и наоборот. Подобно тому, как

фронт цилиндрической волны получается из фронта сферической волны формальной подстановкой в формуле сферического фронта $y = 0$, так и функцию пропускания цилиндрической линзы можно получить, положив в функции пропускания сферической линзы $y = 0$. В итоге для функции пропускания цилиндрической линзы будем иметь

$$t(x) = \exp\left(-ik \frac{x^2}{2f}\right).$$

Таким образом, линза осуществляет квадратичную фазовую модуляцию сигнала. Она сдвигает фазу падающего пучка на величину, пропорциональную квадрату расстояния ρ между осью и точкой падения пучка: $\rho^2 = x^2 + y^2$. Собирающая (положительная) линза создает отрицательный сдвиг фаз, а рассеивающая (отрицательная) – положительный. Поскольку линза – модулятор (фазовый), то функция $t(x, y)$ является функцией пропускания тонкой линзы. Следовательно, действие тонкой линзы на поле проходящей световой волны сводится к умножению функции распределения этого поля в плоскости XU непосредственно перед линзой на функцию пропускания линзы. Поэтому если тонкая выпуклая линза освещается монохроматическим источником света, у которого распределение комплексных амплитуд светового поля в плоскости непосредственно перед линзой описывается функцией $E(x, y)$, то аналогичное распределение в плоскости по другую сторону линзы можно записать так:

$$E'(x, y) = E(x, y)t(x, y). \quad (10.17)$$

В частности, для выпуклой положительной линзы эти распределения связаны соотношением (10.15).

Конечные поперечные размеры линзы учитываются введением в функцию пропускания в качестве множителя так называемой зрачковой функции $P(x, y)$, определяемой как

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

Под апертурой линзы понимают угол раскрытия пучка лучей (обозначается $2\sigma_A$), идущих из какой-либо точки S через линзу (или угол, под которым видно входное отверстие линзы из точки S). Иногда апертурой линзы называют область значений координат x и y , внутри которой $P(x, y) \neq 0$, т.е. внутри круга $x^2 + y^2 \leq R^2$, где R – радиус апертуры линзы. С учетом конечных размеров линзы соотношение для $E'(x, y)$ переписывается в виде

$$E'(x, y) = E(x, y)t(x, y)P(x, y).$$

10.3.2. Фокусирующие свойства линзы

Значение фазового преобразования, осуществляемого линзой, можно продемонстрировать на различных примерах волн, проходящих через линзу. Начнем с плоской волны, падающей на линзу параллельно ее оптической оси (оси Z). Покажем, что после прохождения через линзу плоская волна превращается в сферическую волну, сходящуюся в фокусе линзы. Пусть амплитуда плоской волны равна E_0 . Тогда поле волны, прошедшей через линзу в плоскости непосредственно за линзой найдем, умножив амплитуду плоской волны E_0 на функцию пропускания линзы $t(x, y)$. Получим

$$E(x, y, 0) = E_0 \exp\left(-ik \frac{x^2 + y^2}{2f}\right). \quad (10.18)$$

Соотношение (10.18) описывает в параксиальном приближении поле сферической волны, сходящейся в точку, расположенную на оптической оси линзы на расстоянии f от начала координат. Действительно, для поля сферической волны, исходящей из точки

$(x, y, 0)$ линзы и сходящейся в точку $(0, 0, f)$, в параксиальном приближении имеем

$$E(x, y, 0) = \frac{C \exp(-ikf)}{f} \exp\left(-ik \frac{x^2 + y^2}{2f}\right).$$

Следовательно, квадратичный фазовый множитель в выражении (10.18) указывает на сходящуюся в точке $F'(0, 0, f)$ сферическую волну.

В случае вогнутой линзы амплитуда E_0 плоской волны умножается на фазовый множитель $t(x, y) = \exp\left(ik \frac{x^2 + y^2}{2f}\right)$.

Получающееся при этом выражение описывает в параксиальном приближении фронт сферической волны, расходящейся из точки $(0, 0, -f)$.

Таким образом, лучи, идущие параллельно оптической оси, после преломления в положительной линзе собираются (пересекаются) в одной точке F' ,

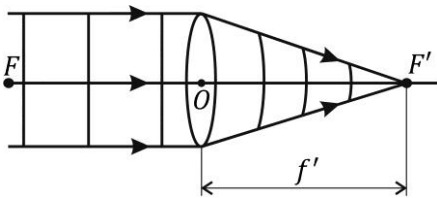


Рис. 10.2

расположенной на оптической оси линзы (рис. 10.2) на расстоянии f за линзой. Иными словами, тонкая положительная линза преобразует плоскую волну в сферическую волну, сходящуюся в точке $F'(0, 0, f')$. Точка F' называется фокусом линзы. Линза имеет два фокуса: передний F (со стороны падающего луча – слева от линзы) и задний F' (со стороны преломленного луча).

Отрицательная линза преобразует параллельный пучок лучей в расходящийся. Если эти лучи продолжить в обратном направлении, то они пересекутся в точке $F(0, 0, -f)$, называемой мнимым фокусом линзы (рис. 10.3). В связи с этим линзу с

положительным фокусным расстоянием называют собирающей, а линзу с отрицательным фокусным расстоянием – рассеивающей. Плоская волна, не обладающая кривизной, после прохождения через положительную линзу приобретает положительную кривизну $1/f$, а после прохождения через отрицательную линзу – отрицательную кривизну $-1/f$.

Мы видим, что действие тонкой линзы на плоскую волну сводится к преобразованию поля этой волны в поле сферической вол-

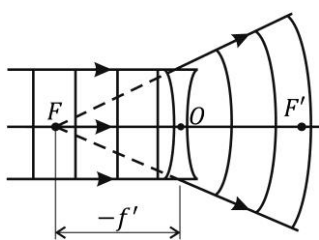


Рис. 10.3

ны, сходящейся (для положительной линзы) в некоторой точке $F'(0, 0, f)$ за линзой или расходящейся (для отрицательной линзы) из некоторой точки $F(0, 0, -f)$ перед линзой.

С другой стороны, линза преобразует световое поле, изображаемое точечным источником, расположенным в ее передней фокальной плоскости, в идеальную плоскую волну. Действительно, умножив поле

$$E(x, y, 0) = \frac{C \exp(ik f)}{f} \exp\left(ik \frac{x^2 + y^2}{2f}\right),$$

создаваемое в плоскости $z = 0$ и исходящее из точки S , которая находится на расстоянии f перед линзой, на функцию пропускания линзы (10.14), получим в той же плоскости $z = 0$ поле плоской волны с амплитудой $C \exp(ik f)/f$:

$$\begin{aligned} E(x, y, 0) &= \frac{C \exp(ik f)}{f} \exp\left(ik \frac{x^2 + y^2}{2f}\right) \exp\left(-ik \frac{x^2 + y^2}{2f}\right) = \\ &= C \exp(ik f)/f. \end{aligned}$$

Амплитуда осталась прежней, так как линза меняет только фазу волны, не изменяя амплитуды. Идеальность получающейся плоской волны ограничена только апертурой линзы. (Такая линза называется дифракционно-ограниченной.)

Заметим, что это утверждение можно было бы рассматривать и как опытный факт и, основываясь на нем, получить функцию пропускания $t(x, y)$ линзы как такую функцию, при умножении на которую комплексная амплитуда поля расходящейся сферической волны в плоскости $z = 0$ преобразуется в амплитуду поля плоской волны в той же плоскости.

Покажем теперь, что собирающая линза превращает сферическую волну в сферическую же волну, но с иным радиусом кривиз-

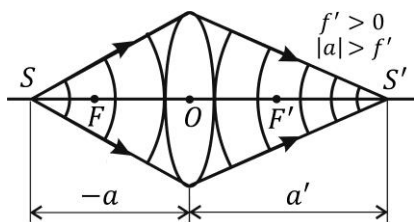


Рис. 10.4

ны. Центры кривизны этих волн представляют собой точечный объект S и его действительное изображение S' , трактуемые с точки зрения геометрической оптики (рис. 10.4).

Действительно, пусть точка $S(0, 0, -a)$ испускает

сферическую волну. Поле, создаваемое этой расходящейся сферической волной в плоскости непосредственно перед

линзой, будет $\frac{C}{a} \exp\left(ik \frac{x^2 + y^2}{2a}\right)$. Поле за линзой найдется

умножением этой функции на функцию пропускания линзы:

$$\frac{C}{a} \exp\left(ik \frac{x^2 + y^2}{2a}\right) \cdot \exp\left(-ik \frac{x^2 + y^2}{2f}\right) = \frac{C}{a} \exp\left(-ik \frac{x^2 + y^2}{2a'}\right),$$

где

$$\frac{1}{a'} = -\frac{1}{a} + \frac{1}{f}. \quad (10.19)$$

Следовательно, за линзой будет распространяться сферическая волна с кривизной, определяемой соотношением (10.19). При этом, если $|a| > f$, сферическая волна, распространяющаяся за линзой, будет сходиться в точке $S'(0, 0, a')$. Если же $|a| < f$, кривизна $1/a'$ изменит свой знак на противоположный; волна станет расходящейся из точки $S'(0, 0, -a')$. В первом случае (рис. 10.4) точка S' является действительным изображением точки S , а во втором – мнимым изображением. Второй случай показан на рис. 10.5.

В случае рассеивающей линзы поле в плоскости непосредственно за линзой будет

$$\frac{C}{a} \exp\left(ik \frac{x^2 + y^2}{2a}\right) \cdot \exp\left(ik \frac{x^2 + y^2}{2f}\right) = \frac{C}{a} \exp\left(ik \frac{x^2 + y^2}{2a'}\right).$$

Следовательно, рассеивающая линза превращает расходящуюся сферическую волну в расходящуюся же волну с кривизной

$$\frac{1}{a'} = \frac{1}{a} + \frac{1}{f}. \quad (10.20)$$

Расходится эта волна не из точки $S(0, 0, -a)$, а из точки $S'(0, 0, -a')$, являющейся мнимым изображением точки S (рис. 10.6).

Мы видим, что линзу можно рассматривать как оптический элемент, преобразующий сферическую волну одного радиуса в

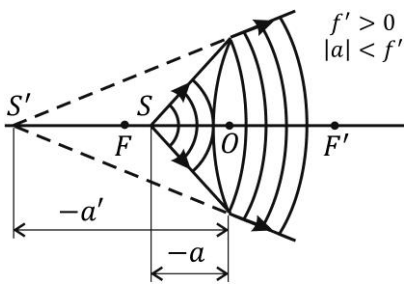


Рис. 10.5

сферическую волну другого радиуса, или, иными словами, гомоцентрический (исходящий из одной точки или сходящийся в одну точку) пучок одной кривизны в гомоцентрический пучок другой кривизны. Следовательно, линза добавляет к фазе приходящего сигнала

дополнительный член, соответствующий сферическому фронту волны с радиусом кривизны, равным фокусному расстоянию линзы.

Преобразование плоской волны в сходящуюся или расходящуюся сферическую волну радиуса $a' = f$ есть частный случай преобразования сферической волны радиуса a в сферическую волну радиуса a' , когда $a = \infty$.

Справедливо и обратное утверждение (этот случай рассмотрен нами выше): если на линзу падает сферическая волна с радиусом, равным фокусному расстоянию линзы ($a = f$), то на выходе получим плоскую волну ($a' = \infty$).

Представляет интерес рассмотреть также случай, когда точечный источник света S расположен в передней фокальной плоскос-

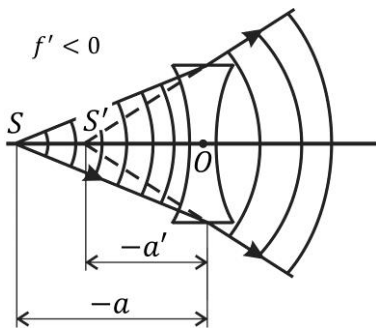


Рис. 10.6

ти линзы на расстоянии x_0 от оптической оси (рис. 10.7), т.е. имеет координаты $x_0, 0, -f$. Сферическая волна, исходящая из этого источника, в плоскости $z = 0$, расположенной непосредственно перед тонкой собирающей линзой, создает поле

$$E(x, y) = \frac{C}{r} \exp(ikr),$$

где

$$r = \sqrt{(x-x_0)^2 + y^2 + f^2} = \sqrt{x^2 + y^2 - 2xx_0 + R^2},$$

$R^2 = f^2 + x_0^2$. В параксиальном приближении

$$r \approx R + \frac{x^2 + y^2 - 2xx_0}{2R}.$$

Предполагая, что $x_0 \ll R$, а значит, $R \approx f$ и так как $x_0 / R = \sin \varphi$, получим

$$r \approx f + \frac{x^2 + y^2}{2f} - \frac{xx_0}{R} = f + \frac{x^2 + y^2}{2f} - x \sin \varphi.$$

С учетом этого для поля непосредственно перед линзой будем иметь

$$E(x, y) = \frac{C \exp(ikf)}{f} \exp\left(ik \frac{x^2 + y^2}{2f}\right) \exp(-ikx \sin \varphi).$$

Умножив это выражение на функцию пропускания линзы, найдем поле непосредственно за линзой:

$$E(x, y) = \frac{C \exp(ikf)}{f} \exp(-ikx \sin \varphi).$$

Полученное выражение показывает, что если в передней фокальной плоскости линзы в точке S , смещенной на расстояние x_0

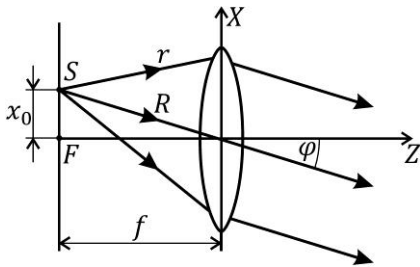


Рис. 10.7

относительно оптической оси линзы, расположить точечный источник света, то за линзой будет распространяться плоская волна в направлении под углом φ к оптической оси линзы (к оси Z), определяемым соотношением $\sin\varphi = x_0/R$.

Справедливо и обратное утверждение: плоская волна, падающая на собирающую линзу под углом φ к оптической оси линзы (т.е. плоская волна с пространственной частотой $u = \sin\varphi/\lambda$), преобразуется этой линзой в сферическую волну, сходящуюся в точку S , расположенную на расстоянии x_0 от оптической оси, т.е. в направлении, определяемом углом φ . Этим свойством собирающей линзы мы пользовались в главе 5 при изучении дифракции Фраунгофера. Вследствие этого свойства каждая пространственная гармоника, входящая в пространственный спектр сложного излучения, падающего на линзу, дает в задней фокальной плоскости линзы свою светящуюся точку. Совокупность этих точек и определяет пространственный фурье-спектр сложного излучения.

Собирающее и рассеивающее действия линзы можно объяснить так. Входящие в линзу участки фронта волны замедляются (так как скорость света в линзе $v = c/n < c$), происходит их задержка на величину, пропорциональную толщине линзы в данном месте. В результате этого участки фронта волны, проходящие у краев линзы, выйдут наружу раньше (для выпуклой линзы) или позже (для вогнутой линзы), чем участки, идущие посередине, и фронт волны станет вогнутым или

выпуклым – волна станет сходящейся в точке F' или расходящейся из точки F . Точка F' , как и точка S' , есть точка дифракционного максимума; в эту точку волны, проходящие через разные части линзы, приходят в одинаковой фазе. Условием максимума является равенство оптических путей всех лучей, идущих через линзу в точку F' (или точку S'). По современным представлениям в основе фокусирующего действия линзы лежит процесс преобразования сигнала в его спектр. Как было показано в п. 6.2, амплитуда поля в фокальной плоскости линзы пропорциональна спектральной плотности поля на входе линзы⁸. Пусть l_1 – длина входного сигнала. Согласно соотношению неопределенностей (называемому также соотношением масштабов) пространственная частота – длина, ширина Δu спектра входного сигнала обратно пропорциональна его длине l_1 : $\Delta u \approx 1 / l_1$. Так как $u = x / (\lambda f)$, а $x = l_2 / 2$, то $\Delta u \approx l_2 / (2\lambda f)$. Сравнивая оба выражения для Δu , получаем $l_2 = 2\lambda f / l_1$. Следовательно, длина выходного сигнала l_2 обратно пропорциональна длине входного сигнала l_1 . Поэтому если l_1 велико, то l_2 – мало. В этом и состоит фокусирующее действие линзы. Следует, однако, отметить, что линза не фокусирует, а модулирует по фазе падающую на нее световую волну. Фокусирующим действием обладает линза вместе с примыкающим к ней слоем свободного пространства. Таким образом, фокусирование включает в себя пространственную модуляцию (линзой) и фильтрацию (слоем свободного пространства).

Формула (10.19) называется формулой собирающей линзы, а формула (10.20) – формулой рассеивающей линзы. Эти формулы определяют положение изображения (a') по известному положению предмета (a) и фокусному расстоянию линзы (f). При этом следует иметь в виду принятое в оптике правило знаков, согласно которому все расстояния отсчитываются от центра

⁸ Подробно об этом будет говориться в п. 10.6.

линзы и если эти расстояния проходятся в направлении противоположном направлению распространения светового луча (справа налево, в отрицательном направлении оси Z) они считаются отрицательными, а, если в направлении, совпадающем с направлением луча, – положительными.

Линзу по своему действию на плоскую волну можно рассматривать как зонную пластинку Френеля, поэтому отношение интенсивности в фокусе линзы к интенсивности падающей волны можно оценить по той же формуле, что и для зонной пластинки: $I/I_0 = R^4 / (\lambda a')^2$, положив в ней $a' = f$ и принимая в качестве R радиус линзы. Тогда при $R = 0,5$ см, $f = 10$ см и $\lambda = 0.5$ мкм будем иметь $I/I_0 \sim 10^6$.

Слой свободного пространства, линза и транспаранти-диапозитив являются основными элементами когерентной оптической системы.

Фокусирующее свойство линзы определяет ее способность формировать оптическое изображение.

Относительно изображающих свойств линз необходимо отметить следующее. Если точечный источник S находится на расстоянии a_1 , не удовлетворяющем формуле линзы, то его изображение получается не в виде точки S' , а в виде светового кружка, называемого кружком рассеяния (рис. 10.8). В случае, когда строится изображение объемного предмета, его отдельные элементы в соответствии с формулой линзы изображаются в различных плоскостях, и изображение всего предмета будет иметь пространственный характер. Однако на практике обычно изобраа-

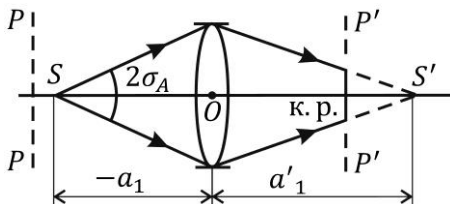


Рис. 10.8

жение объемного предмета получается на плоскости (двумерном экране или на фотопластинке), поэтому только для точек предмета, которые лежат в плоскости PP , называемой плоскостью установки или плоскостью наводки, будет удовлетворяться формула линзы, а их изображение будет резким. Изображения же остальных точек предмета будут получаться в виде кружков рассеяния. Одновременно четко передать все элементы изображения объемного предмета на плоскости оказывается невозможным. Элементы предмета, расположенные на расстоянии, не удовлетворяющем формуле линзы, будут более или менее размытыми в зависимости от их удаления от плоскости изображения.

Глаз человека, как будет показано ниже, обладает определенной разрешающей способностью, или элементом разрешения. Это объясняет существование некоторой глубины пространства, резко отображаемого на плоскости, определяющего продольное расстояние между двумя точками предмета, размеры изображений которых на фотопленке не превышают элемента разрешения глаза. Зарегистрированное таким образом на фотопленке изображение уже нельзя превратить в трехмерное. Третье измерение можно воспринимать только за счет размытия удаленных точек предмета, за счет законов перспективы, изменения цвета и т.п.

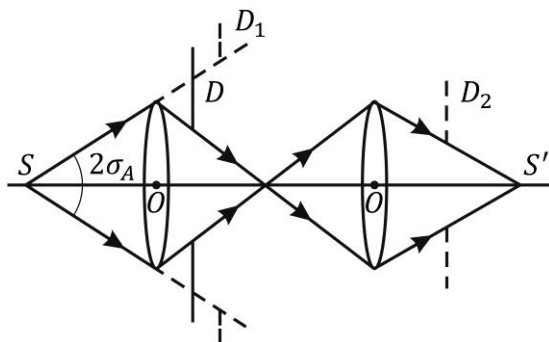


Рис. 10.9

Как видно из рис. 10.8, размер кружка рассеяния, соответствующего изображению точки S в плоскости PP' , будет тем больше, чем больше смещение S из плоскости установки PP (или из сопряженной ей плоскости $P'P'$) и чем больше угловая ширина $2\sigma_A$ реального пучка лучей, формирующих изображение.

Для получения удовлетворительного изображения размер кружка рассеяния не должен превышать определенного предела. Глубина резкости возрастает с уменьшением угловой ширины пучка. Поэтому для уменьшения угловой ширины светового пучка (или, как говорят, для ограничения световых пучков) в оптических системах формирования изображения используются специальные диафрагмы (на рис. 10.9 диафрагма обозначена буквой D), называемые апертурными диафрагмами. Они располагаются либо между линзами, либо перед системой, либо после нее. Роль диафрагмы может играть оправа линзы (на рис. 10.9 изображена двумя вертикальными черточками). Изображение апертурной диафрагмы (действительное или мнимое), создаваемое находящейся перед ней частью оптической системы, называется входным зрачком (или входным отверстием системы), а изображение, создаваемое расположенными за диафрагмой оптическими элементами – выходным зрачком (на рис. 10.9 входной и выходной зрачки обозначены буквами D_1 и D_2). Входным зрачком определяется уже упоминавшаяся в п. 10.3

апертура – угол $2\sigma_A$ раскрытия пучка лучей, идущих из точки S предмета через систему.

В заключение этого параграфа рассмотрим вопрос о таутохронизме линз. Под таутохронизмом понимают свойство линзы, согласно которому оптические пути всех лучей, вышедших из

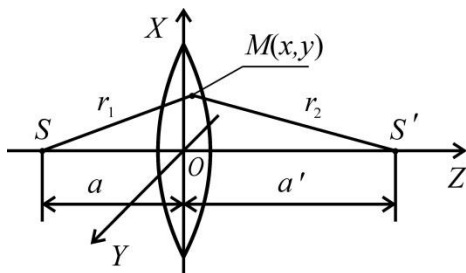


Рис. 10.10

точки S и пришедших в точку изображения S' , одинаковы. Так как геометрические пути различны, выравнивание оптических путей происходит в линзе. Разность геометрических путей должна быть скомпенсирована

фазовым сдвигом, обусловленным линзой. Пусть точка S лежит на оптической оси положительной линзы на расстоянии a от нее, точка изображения S' тоже лежит на оптической оси, но на расстоянии a' от линзы (рис. 10.10). Геометрический путь лучей, идущих через центр линзы, будет

$$L_1 = a + a'.$$

Геометрический же путь лучей, прошедших через произвольную точку $M(x, y)$ линзы, находящуюся на расстоянии $\rho = \sqrt{x^2 + y^2}$ от оптической оси,

$$L_2 = r_1 + r_2 = \sqrt{\rho^2 + a^2} + \sqrt{\rho^2 + a'^2}.$$

В параболическом приближении

$$L_2 = a + \frac{\rho^2}{2a} + a' + \frac{\rho^2}{2a'}.$$

Разность геометрических путей

$$\Delta = L_2 - L_1 \cong \frac{x^2 + y^2}{2a} + \frac{x^2 + y^2}{2a'} = \frac{x^2 + y^2}{2f}.$$

При получении этого выражения была использована формула линзы (10.19). Вследствие этой разности хода волны, идущие по второму (более длинному) пути, будут отставать по фазе от волн, идущих по первому пути, на величину

$$\Delta\varphi = k\Delta = k \frac{x^2 + y^2}{2f}.$$

Но знак минус перед показателем степени экспоненты (10.14) показывает, что волны, идущие через точку M , опережают по фазе волны, идущие через центр линзы. Следовательно, эти разности фаз компенсируют друг друга, поэтому и оптические пути L_2 и L_1 должны быть равны.

Тот факт, что оптические пути всех лучей, вышедших из точки S и пришедших в точку изображения S' , одинаковы, является следствием известного в оптике принципа Ферма. Взяв за основу этот факт, можно придти к формуле (10.14). Действительно, указанное выше отставание по фазе может быть скомпенсировано

фазовым сдвигом $\Delta\varphi = -k \frac{x^2 + y^2}{2f}$, вносимым линзой.

Следовательно, фазовая характеристика линзы должна иметь вид

$$t(x, y) = \exp(i\Delta\phi) = \exp\left(-ik \frac{x^2 + y^2}{2f}\right),$$

что совпадает с формулой (10.14).

10.3.3. Тонкая линза как элемент, выполняющий преобразование Фурье

Как уже отмечалось при рассмотрении дифракции Фраунгофера, тонкая линза обладает свойством фурье-преобразования, т.е. производит физическое разложение падающего на нее излучения в пространственный фурье-спектр. При этом каждая гармоническая составляющая, т.е. плоская волна определенного направления, в виде совокупности которых можно представить излучение, падающее на линзу, фокусируется линзой в «свою» точку задней фокальной плоскости, в которой возникает, таким образом, картина пространственного спектра. Чтобы убедиться в этом, найдем распределение амплитуды светового поля в задней фокальной плоскости линзы, расположенной за транспарантом, освещаемым плоской волной. Для этого будем последовательно вычислять поле в различных плоскостях оптической системы по ходу распространения световой волны.

Рассмотрим сначала частный случай, когда транспарант расположен непосредственно перед собирающей линзой (рис. 10.11).

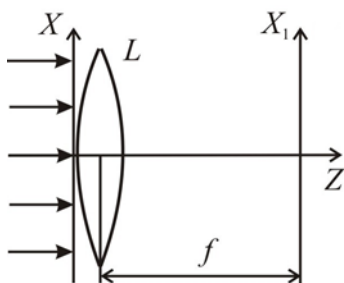


Рис. 10.11

С плоскостью транспаранта свяжем координатную плоскость XU , функцию пропускания транспаранта обозначим $t(x, y)$.

Предположим, что транспарант равномерно освещается нормально падающей плоской монохроматической волной с амплитудой E_0 . Тогда распределение амплитуды

светового поля непосредственно перед линзой (входной сигнал) будет $E(x, y) = E_0 t(x, y)$. В плоскости непосредственно за линзой распределение амплитуды поля $E'(x, y)$ найдется умножением поля $E(x, y)$ на функцию пропускания линзы (10.14):

$$E'(x, y) = E(x, y) \exp\left(-ik \frac{x^2 + y^2}{2f}\right). \quad (10.21)$$

Считаем, что поле $E(x, y)$ локализовано вблизи оси Z и занимает небольшую часть пространства. Поэтому конечными размерами линзы пренебрегаем, т.е. полагаем, что зрачковая функция линзы $P(x, y) = 1$ при любых значениях x и y .

Распределение поля $E_1(x_1, y_1)$ в задней фокальной плоскости X_1Y_1 линзы найдем как сигнал на выходе слоя свободного пространства толщиной f , непосредственно примыкающего к этой плоскости, т.е. с помощью преобразования Френеля

$$E_1(x_1, y_1) = \frac{e^{ikz}}{i\lambda z} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp\left[ik \frac{(x_1 - x)^2 + (y_1 - y)^2}{2z}\right] dx dy, \quad (10.22)$$

положив в нем $z = f$. Сигналом на входе слоя свободного пространства в данном случае является поле $E'(x, y)$. Подставив выражение (10.21) в формулу (10.22) вместо $E(x, y)$ и разложив квадратичные члены в подынтегральной функции, мы видим, что квадратичный фазовый множитель $\exp\left(-ik \frac{x^2 + y^2}{2f}\right)$ в подынтегральной функции сокращается. В результате приходим к выражению

$$E_1(x_1, y_1) = \frac{\exp(ikf)}{i\lambda f} \exp\left(ik \frac{x_1^2 + y_1^2}{2f}\right) \times \\ \times \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp\left(-ik \frac{xx_1 + yy_1}{f}\right) dx dy,$$

где x_1, y_1 – координаты в фокальной плоскости X_1Y_1 . Обозначив $u = x_1 / (\lambda f)$, $v = y_1 / (\lambda f)$, выражение для $E_1(x_1, y_1)$ можно записать в виде

$$E_1(x_1, y_1) = \frac{\exp(ikf)}{i\lambda f} \exp\left(ik \frac{x_1^2 + y_1^2}{2f}\right) \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp[-i2\pi(ux + vy)] dx dy. \quad (10.23)$$

Двойной интеграл в (10.23) представляет собой двумерное преобразование Фурье функции пропускания транспаранта $t(x, y)$ на пространственных частотах $u = x_1 / (\lambda f)$ и $v = y_1 / (\lambda f)$, т.е. спектральную плотность входного сигнала $E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right)$.

Следовательно,

$$E_1(x_1, y_1) = \frac{\exp(ikf)}{i\lambda f} \exp\left(ik \frac{x_1^2 + y_1^2}{2f}\right) E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right).$$

Опуская не зависящий от координат x_1 и y_1 множитель $\frac{\exp(ikf)}{i} = -i \exp(ikf) = \exp[i(kf - \pi/2)]$, для поля в фокальной плоскости линзы будем иметь

$$E_1(x_1, y_1) = \frac{1}{\lambda f} E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right) \exp\left(ik \frac{x_1^2 + y_1^2}{2f}\right). \quad (10.24)$$

Для модуля выходного сигнала получим более простое выражение

$$|E_1(x_1, y_1)| = \frac{1}{\lambda f} \left| E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right) \right|.$$

Соотношение (10.24) показывает, что поле в фокальной плоскости линзы с точностью до множителя $\frac{e^{ikf}}{i\lambda f} \exp\left(ik \frac{x_1^2 + y_1^2}{2f}\right)$ представляет собой фурье-образ распределения комплексных амплитуд в плоскости предмета. Поскольку поле $E(x, y)$ и функция пропускания линзы $t(x, y)$ отличаются друг от друга несущественным (не зависящим от координат x_1 и y_1) множителем E_0 , то соотношение (10.24) можно записать также в виде

$$E_1(x_1, y_1) = \frac{1}{\lambda f} T\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right) \exp\left(ik \frac{x_1^2 + y_1^2}{2f}\right),$$

где

$$T\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y) \exp\left[-ik2\pi\left(\frac{x_1}{\lambda f}x + \frac{y_1}{\lambda f}y\right)\right] dx dy$$

– фурье-образ функции пропускания транспаранта.

Согласно (10.24) каждой точке (x_1, y_1) фокальной плоскости соответствует пара пространственных частот $u = x_1/(\lambda f)$ и $v = y_1/(\lambda f)$ спектра предметных волн. Наличие квадратичного фазового множителя показывает, что фазовое распределение в фокальной плоскости не совпадает с фазовым распределением в спектре предмета (имеется в виду фурье-спектр функции пропускания предмета $t(x, y)$, так как E и t отличаются постоянным множителем E_0). Различие между ними представляет простое фазовое искажение. Однако это искажение не играет существенной роли, так как оно не влияет на пространственное распределение света. Квадратичный фазовый множитель

$\exp\left(ik \frac{x_1^2 + y_1^2}{2f}\right)$ можно устранить, поместив в плоскости X_1Y_1 еще одну линзу с фокусным расстоянием f ; эта вторая линза, модулируя сигнал по закону $\exp\left(-ik \frac{x_1^2 + y_1^2}{2f}\right)$, даст на выходе поле $\frac{1}{\lambda f} E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right)$. Кроме того,

Поскольку фотоэмульсия (или фотопластинка), используемая для регистрации сигнала, реагирует только на интенсивность падающего на него света, т.е. на величину, зависящую от квадрата модуля сигнала, то в большинстве случаев интересуются распределением интенсивности в фокальной плоскости линзы, т.е. величиной $|E_1(x_1, y_1)|^2$, для которой фазовые множители не существенны, так как при определении интенсивности все фазовые множители исчезают. Вследствие чего исчезнут и фазовые искажения.

Таким образом, амплитуда поля в фокальной плоскости линзы пропорциональна фурье-образу (спектральной плотности) поля на входе линзы. В этом проявляется фокусирующее действие линзы.

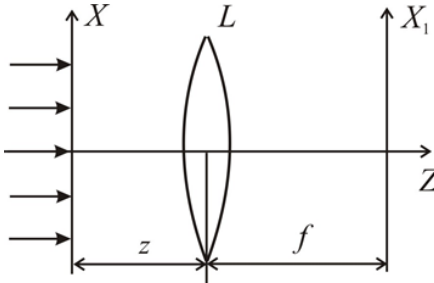


Рис. 10.12

Поэтому можно утверждать, что в основе фокусирующего действия линзы лежит процесс преобразования сигнала в его спектр.

Рассмотрим теперь общий случай, когда транспарант находится на расстоянии z от линзы в плоскости XU , т.е.

когда между транспарантом и линзой имеется слой свободного пространства толщиной z (рис. 10.12). Пусть $E(x, y)$ – амплитуда поля в плоскости непосредственно за транспарантом. Амплитуда

поля в плоскости непосредственно перед линзой найдется с помощью преобразования Френеля:

$$E(\xi, \eta) = \frac{e^{ikz}}{i\lambda z} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp \left[ik \frac{(\xi - x)^2 + (\eta - y)^2}{2z} \right] dx dy \quad (10.25)$$

или как свертка поля $E(x, y)$ с импульсной характеристикой свободного пространства:

$$E(\xi, \eta) = E(x, y) * h(x, y; \xi, \eta),$$

где ξ и η – координаты в плоскости линзы. Произведя фурье-преобразование от обеих частей этого равенства и используя теорему свертки, получим

$$\begin{aligned} F\{E(\xi, \eta)\} &= F\{E(x, y)\} \cdot F\{h(x, y, \xi, \eta)\} = \\ &= E(u, v) \cdot H(u, v), \end{aligned} \quad (10.26)$$

где $H(u, v)$ – частотная характеристика слоя свободного пространства на частотах $u = x_1 / (\lambda f)$ и $v = y_1 / (\lambda f)$.

Соотношение (10.26) справедливо не только, когда перед модулятором (в нашем случае перед линзой) расположен слой свободного пространства, но и любой фильтр с частотной характеристикой (коэффициентом передачи) $H(u, v)$ ⁹.

Поскольку распределение амплитуды поля в фокальной плоскости линзы равно фурье-образу амплитуды поля в плоскости непосредственно перед линзой $F\{E(\xi, \eta)\}$, умноженному на

величину $\frac{e^{ikf}}{i\lambda f} \exp \left(ik \frac{x_1^2 + y_1^2}{2f} \right)$, то получим

⁹ Об этом будет подробно говориться в главе 12.

$$E_1(x_1, y_1) = \frac{e^{ikf}}{i\lambda f} \exp\left(ik \frac{x_1^2 + y_1^2}{2f}\right) E(u, v) H(u, v).$$

Учитывая, что

$$E(u, v) = E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right),$$

$$\begin{aligned} H(u, v) &= \exp[-i\pi\lambda z(u^2 + v^2)] = \\ &= \exp\left\{-i\pi\lambda z \left[\left(\frac{x_1}{\lambda f}\right)^2 + \left(\frac{y_1}{\lambda f}\right)^2\right]\right\} = \exp\left(-ikz \frac{x_1^2 + y_1^2}{2f^2}\right). \end{aligned}$$

и опуская несущественный фазовый множитель $-i \exp(ikf)$, будем иметь

$$E_1(x_1, y_1) = \frac{1}{\lambda f} \exp\left[ik \frac{x_1^2 + y_1^2}{2f} \left(1 - \frac{z}{f}\right)\right] E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right). \quad (10.27)$$

Как видим, и в том случае, когда транспарант расположен на некотором расстоянии z от линзы, распределение амплитуды поля в фокальной плоскости линзы пропорционально фурье-образу распределения поля на плоскости непосредственно за транспарантом, но с коэффициентом

$$\frac{1}{\lambda f} \exp\left[ik \frac{x_1^2 + y_1^2}{2f} \left(1 - \frac{z}{f}\right)\right],$$

зависящим от толщины слоя z .

Модуль же комплексной амплитуды в

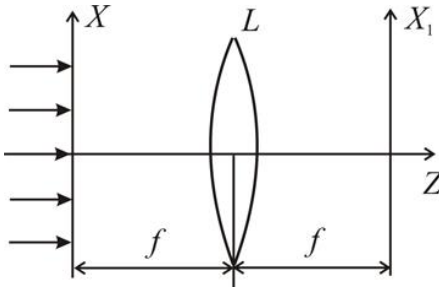


Рис. 10.13

фокальной плоскости просто равен модулю спектральной плотности поля во входной плоскости и не зависит от расстояния z .

Указанный квадратичный фазовый множитель, не влияя на пространственное распределение поля,

указывает на то, что в каждой точке фокальной плоскости поле представляет собой сферическую волну, сходящуюся в точку, расположенную на оси Z на расстоянии

$$l = \left[\frac{1}{f} \left(1 - \frac{z}{f} \right) \right]^{-1} = \frac{f^2}{f - z}$$

от фокальной плоскости линзы, и поэтому несет информацию о положении плоскости изображения. При $l < 0$ (что соответствует случаю $z > f$) эта точка лежит в плоскости, находящейся за фокальной плоскостью. При $l > 0$ ($z < f$) плоскость изображения мнимая, так как через точки фокальной плоскости проходят расходящиеся сферические волны. Если транспарант располагается в передней фокальной плоскости линзы (рис. 10.13), т.е. если толщина слоя свободного пространства между транспарантом и линзой $z = f$ ($l = \infty$), то фазовый множитель в (10.27) обратится в единицу, и мы получим

$$E_1(x_1, y_1) = \frac{1}{\lambda f} E \left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f} \right). \quad (10.28)$$

В этом случае фокальную плоскость пересекают пучки параллельных лучей, и распределение поля в фокальной плоскости линзы с точностью до постоянного множителя $1/(\lambda f)$

и отброшенного фазового множителя $\exp[i(kf - \pi / 2)]$ будет равно фурье-

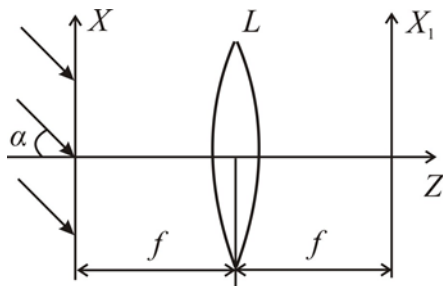


Рис. 10.14

преобразованию распределения амплитуды поля на предмете. Схема, изображенная на рис. 10.13, является основной схемой реализации преобразования Фурье.

Из проведенного рассмотрения видно, что для осуществления преобразования Фурье

необходимо не только чтобы свет прошел через линзу, но и чтобы он прошел после этого в свободном пространстве путь, равный фокусному расстоянию линзы. Даже в процессе распространения света в свободном пространстве происходит изменение его комплексной амплитуды.

Если в оптической системе, в которой транспарант располагается в переднем фокусе линзы (в схеме с $z = f$) плоская волна падает на транспарант произвольно (не перпендикулярно транспаранту, рис. 10.14), то поле, создаваемое этой волной в плоскости $z = 0$, определяется формулой

$$E_0(x, y) = E_0 \exp[i(k_x x + k_y y)] = E_0 \exp[i2\pi(x \cos \alpha / \lambda + y \cos \beta / \lambda)],$$

где α и β – направляющие углы вектора \mathbf{k} волны. Поле в фокальной плоскости в этом случае будет

$$E_1(x_1, y_1) = \frac{1}{i\lambda f} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(\xi, \eta) \exp[i2\pi(x \cos \alpha / \lambda + y \cos \beta / \lambda)] \times \exp[-i2\pi(u\xi + v\eta)] d\xi d\eta =$$

$$\begin{aligned}
&= \frac{1}{i\lambda f} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(\zeta, \eta) \exp\{-i2\pi[\zeta(u - \cos \alpha / \lambda) + \\
&+ \eta(v - \cos \beta / \lambda)]\} d\zeta d\eta = \\
&= \frac{1}{i\lambda f} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(\zeta, \eta) \exp[-i2\pi(u'\zeta + v'\eta)] d\zeta d\eta = \frac{1}{\lambda f} E(u', v'),
\end{aligned}$$

где

$$\begin{aligned}
u' &= u - \frac{\cos \alpha}{\lambda} = \frac{x_f}{\lambda f} - \frac{\cos \alpha}{\lambda} = \frac{x_f - f \cos \alpha}{\lambda f}, \\
v' &= v - \frac{\cos \beta}{\lambda} = \frac{y_f}{\lambda f} - \frac{\cos \beta}{\lambda} = \frac{y_f - f \cos \beta}{\lambda f}.
\end{aligned}$$

С учетом этого получаем

$$E_1(x_1, y_1) = \frac{1}{\lambda f} E\left(\frac{x_f - f \cos \alpha}{\lambda f}, \frac{y_f - f \cos \beta}{\lambda f}\right).$$

Это выражение показывает, что при произвольном падении плоской волны на транспарант спектр входного сигнала в фокальной плоскости линзы смещается вдоль оси X на величину $f \cos \alpha$ и вдоль оси Y на величину $f \cos \beta$.

Предположим теперь, что транспарант, расположенный в плоскости XY , освещается расходящейся сферической волной, источник S которой расположен на оси Z на расстоянии a перед транспарантом, а сам транспарант – непосредственно перед линзой (рис. 11.15). Тогда поле непосредственно перед линзой будет

$$E(x, y) = t(x, y) \exp\left(ik \frac{x^2 + y^2}{2a}\right),$$

а после прохождения через линзу

$$E'(x, y) = t(x, y) \exp\left(-ik \frac{x^2 + y^2}{2f}\right) \exp\left(ik \frac{x^2 + y^2}{2a}\right) =$$

$$= t(x, y) \exp\left(-ik \frac{x^2 + y^2}{2f'}\right),$$

где

$$\frac{1}{f'} = \frac{1}{f} - \frac{1}{a}. \quad (10.29)$$

Это соотношение показывает, что освещение транспаранта сферической волной радиуса a эквивалентно изменению фокусного

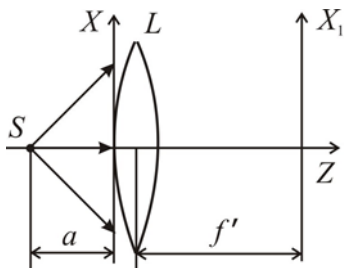


Рис. 10.15

расстояния f линзы на f' , определяемому формулой (10.29). Рассматриваемая система аналогична системе, изображенной на рис. 10.11. Плоскость X_1Y_1 при этом должна находиться на расстоянии f' от линзы. Поле в этой фокальной

плоскости найдем, как и при получении формулы (10.24). Тогда с точностью до множителей, не зависящих от координат x_1, y_1 , будем иметь

$$E_1(x_1, y_1) = \exp\left(ik \frac{x_1^2 + y_1^2}{2f'}\right) T\left(\frac{x_1}{\lambda f'}, \frac{x_1}{\lambda f'}\right), \quad (10.30)$$

где $T\left(\frac{x_1}{\lambda f'}, \frac{y_1}{\lambda f'}\right)$ – фурье-образ функции пропускания транспаранта на пространственных частотах

$$u = x_1 / (\lambda f'), \quad v = y_1 / (\lambda f').$$

Из формулы (10.29) следует, что если $1/f = 0$, то $f' = -a$. А поскольку фокусное расстояние f' должно быть положительным, то должно быть $a < 0$. Это означает, что поле непосредственно перед транспарантом будет таким же, как если бы он освещался сходящейся сферической волной $\exp\left(ik \frac{x^2 + y^2}{2a}\right)$ радиуса $a < 0$

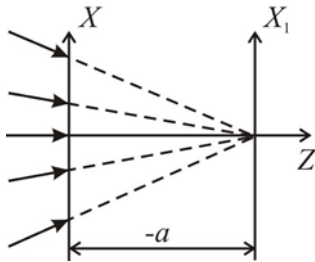


Рис. 10.16

(постоянный коэффициент перед экспонентой опущен). Следовательно, преобразование Фурье можно получить и в безлинзовой оптической системе, если транспарант освещать сходящейся сферической волной радиуса $a < 0$ (рис. 10.16). Плоскость X_1Y_1 должна при этом находиться также на расстоянии $-a$ за

транспарантом. Действительно, поле непосредственно перед транспарантом (в плоскости XY) есть поле, создаваемое сходящейся сферической волной радиуса a , а непосредственно за транспарантом оно будет

$$E(x, y) = \exp\left(ik \frac{x^2 + y^2}{2a}\right)t(x, y).$$

Поле в плоскости X_1Y_1 найдем, используя преобразование Френеля (10.22), положив в нем $z = -a$. Подставив сюда

выражение для $E(x, y)$ и опустив не зависящие от x_1, y_1 коэффициенты, будем иметь

$$\begin{aligned}
 E_1(x_1, y_1) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y) \exp\left(ik \frac{x^2 + y^2}{2a}\right) \times \\
 &\quad \times \exp\left(ik \frac{(x_1 - x)^2 + (y_1 - y)^2}{2(-a)}\right) dx dy = \\
 &= \exp\left(-ik \frac{x_1^2 + y_1^2}{2a}\right) \times \\
 &\quad \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y) \exp\left[i2\pi\left(\frac{x_1}{\lambda a} x + \frac{y_1}{\lambda a} y\right)\right] dx dy.
 \end{aligned}$$

Представив интеграл в этом выражении в виде

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y) \exp\left[-i2\pi\left(\frac{x_1}{-\lambda a} x + \frac{y_1}{-\lambda a} y\right)\right] dx dy,$$

видим, что он представляет собой преобразование Фурье функции $E(x, y)$ на пространственных частотах

$$u = x_1 / (-\lambda a), \quad v = y_1 / (-\lambda a).$$

С учетом этого получаем

$$E_1(x_1, y_1) = \exp\left(-ik \frac{x_1^2 + y_1^2}{2a}\right) T\left(\frac{x_1}{-\lambda a}, \frac{y_1}{-\lambda a}\right).$$

К этому результату можно прийти и заменив в формуле (10.32) величину f' на $-a$.

Результаты проведенного рассмотрения могут быть использованы для построения новых однолинзовых оптических систем, реализующих преобразование Фурье. Одна из таких систем показана на рис. 10.17. В этой системе сферическая волна, освещающая транспарант, создается собирающей линзой с фокусным расстоянием f , на которую падает плоская волна, распространяющаяся вдоль оси Z . Поле за линзой в этом случае

будет $E'(\xi, \eta) = E_0 \exp\left(-ik \frac{\xi^2 + \eta^2}{2f}\right)$, где ξ и η – координаты

в плоскости линзы, а перед транспарантом (с точностью до постоянных коэффициентов) его можно определить с помощью преобразования Френеля (10.25), произведя в нем замену $\xi \leftrightarrow x$, $\eta \leftrightarrow y$ и положив $z = a_1$. Подставив в полученное соотношение входной сигнал $E'(\xi, \eta)$, будем иметь

$$E(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-ik \frac{\xi^2 + \eta^2}{2f}\right) \times \\ \times \exp\left[ik \frac{(x - \xi)^2 + (y - \eta)^2}{2a_1}\right] d\xi d\eta = \exp\left(-ik \frac{x^2 + y^2}{2a_2}\right).$$

При получении этого результата было учтено, что $a_1 + a_2 = f$ (так как плоскость X_1Y_1 есть фокальная плоскость линзы) и использовано соотношение (9.21).

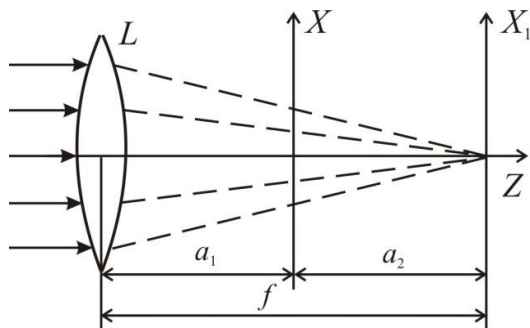


Рис. 10.17

Поле в плоскости X_1Y_1 определим по формуле (10.24), положив в ней $z = a_2$. Подставляя сюда выражение

$$E(x, y) = \exp\left(-ik \frac{x^2 + y^2}{2a_2}\right)t(x, y)$$

и опуская несущественные постоянные множители, получим

$$E_1(x_1, y_1) = \exp\left(ik \frac{x_1^2 + y_1^2}{2a_2}\right)T\left(\frac{x_1}{\lambda a_2}, \frac{x_1}{\lambda a_2}\right). \quad (10.31)$$

Достоинством этой системы является то, что она позволяет, изменяя расстояние a_2 , изменять масштаб выходного сигнала (сигнала в плоскости X_1Y_1). При уменьшении расстояния a_2 уменьшается и масштаб.

Еще одна оптическая система такого типа показана на рис. 10.18. В этой оптической системе собирающая линза с фокусным

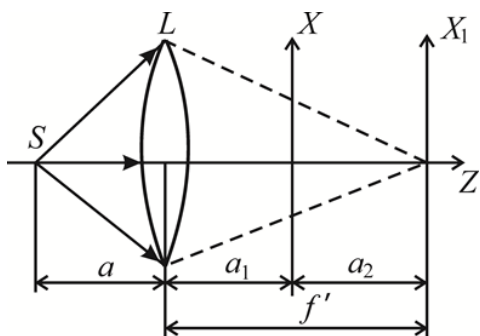


Рис. 10.18

расстоянием f
освещается
сферической волной,
исходящей из
точечного источника S ,
расположенного на оси
 Z на расстоянии a
перед линзой. После
прохождения через
линзу она
преобразуется в
сходящуюся

сферическую волну, радиус a' которой определяется формулой (10.19), причем $|a| > f$. Это эквивалентно пропусканию через линзу с фокусным расстоянием $f' = a'$ плоской волны, распространяющейся вдоль оси Z . Фокусное расстояние f' при этом можно определить из равенства (10.19), заменив в нем a' на f' :

$$\frac{1}{f'} = \frac{1}{f} - \frac{1}{a}.$$

Кроме того,

$$f' = a_1 + a_2. \quad (10.32)$$

Поле, создаваемое этой волной в плоскости непосредственно за линзой в пренебрежении сомножителями, не зависящими от координат плоскости линзы, будет $E(\xi, \eta) = \exp\left(-ik \frac{\xi^2 + \eta^2}{2f'}\right)$.

Поле перед транспарантом найдем с помощью преобразования Френеля (10.25) с заменой $\xi \leftrightarrow x$, $\eta \leftrightarrow y$. Подставив сюда выражение для $E(\xi, \eta)$, положив $z = a_1$, опустив

несущественный множитель перед интегралом и применив формулу (9.21), получим

$$E(x, y) = \exp \left[ik \left(1 - \frac{b}{a_1} \right) \frac{x^2 + y^2}{2a_1} \right],$$

где параметр b определяется равенством

$$\frac{1}{b} = \frac{1}{a_1} - \frac{1}{f'}. \quad (10.33)$$

Поле за транспарантом будет

$$E(x, y) = \exp \left[ik \left(1 - \frac{b}{a_1} \right) \frac{x^2 + y^2}{2a_1} \right] t(x, y).$$

Поле в фокальной плоскости линзы найдем с помощью преобразования Френеля (10.22). Будем иметь

$$\begin{aligned} E_1(x_1, y_1) = & \exp \left(ik \frac{x^2 + y^2}{2a_2} \right) \times \\ & \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y) \exp \left\{ ik \left[\left(1 - \frac{b}{a_1} \right) \frac{1}{a_1} + \frac{1}{a_2} \right] \frac{x^2 + y^2}{2} \right\} \times \\ & \times \exp \left[-ik 2\pi \left(\frac{x_1}{\lambda a_2} x + \frac{y_1}{\lambda a_2} y \right) \right] dx dy. \end{aligned} \quad (10.34)$$

Используя соотношения (10.32) и (10.33), получим, что квадратичный фазовый множитель в (10.34) равен единице. В результате придем к соотношению

$$E_1(x_1, y_1) = \exp \left(ik \frac{x_1^2 + y_1^2}{2a_2} \right) T \left(\frac{x_1}{\lambda a_2}, \frac{y_1}{\lambda a_2} \right),$$

совпадающему с соотношением (10.31).

В этой оптической системе также возможно изменение масштаба выходного сигнала при изменении величины a_2 .

Таким образом, и при освещении транспаранта сферической волной распределение поля в фокальной плоскости линзы пропорционально фурье-образу функции пропускания транспаранта.

10.3.4. Двойное преобразование Фурье

В теории преобразования Фурье преобразование из координатной плоскости в плоскость пространственных частот (прямое преобразование Фурье) осуществляется с помощью ядра $\exp[-i2\pi(ux + vy)]$, а преобразование из плоскости пространственных частот в координатную плоскость (обратное преобразование Фурье) – с помощью комплексно-сопряженного ядра $\exp[i2\pi(ux + vy)]$. Соотношение (10.23) показывает, что положительная линза вводит ядро $\exp[-i2\pi(ux + vy)]$. Следовательно, с помощью однолинзовой оптической системы можно выполнить только прямое преобразование Фурье. Для совершения обратного преобразования Фурье следует на двойном фокусном расстоянии от линзы расположить вторую такую же линзу, а в задней фокальной плоскости этой линзы, где получается обратное преобразование, изменить направления координатных осей на противоположные:

$$F\{F\{E(x, y)\}\} = E(-x, -y).$$

Это вытекает из того, что при прохождении через две линзы световая волна с пространственными частотами u и v преобразуется в волну с пространственными частотами $-u$ и $-v$ (x - и y -компоненты волнового вектора \mathbf{k} волны изменяют знак; ниже мы увидим, что это приводит к инверсии изображения – изображение оказывается перевернутым). Соответствующая схема представлена на рис. 10.19. Эта схема осуществляет два

последовательных преобразования Фурье оптического сигнала. Входной сигнал $E(x, y)$ в плоскости XY подвергается фурье-преобразованию линзой L_1 .

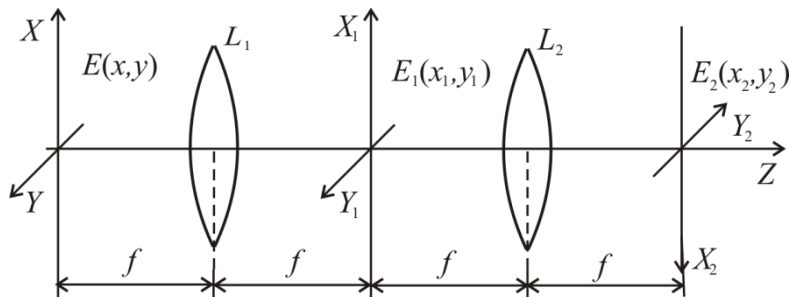


Рис. 10.19

На задней фокальной плоскости X_1Y_1 первой линзы, согласно (10.28), формируется распределение поля $E_1(x_1, y_1)$, пропорциональное спектральной плотности сигнала $E(u, v)$:

$$E_1(x_1, y_1) = (\lambda f)^{-1} E(u, v),$$

где учтено, что $x_1 / (\lambda f) = u$, $y_1 / (\lambda f) = v$. Линза L_2 производит такое же преобразование Фурье (с множителем $(\lambda f)^{-1}$), как и линза L_1 , но теперь преобразованию подвергается поле $E_1(x_1, y_1) = (\lambda f)^{-1} E(u, v)$. В результате распределение поля в задней фокальной плоскости X_2Y_2 линзы L_2 может быть представлено в виде

$$E_2(x_2, y_2) = \frac{1}{(\lambda f)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) \exp[-i2\pi(ux_2 + vy_2)] dx_1 dy_1$$

или, учитывая, что $x_1 = \lambda f u$, $y_1 = \lambda f v$,

$$E_2(x_2, y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) \exp \{i2\pi[u(-x_2) + v(-y_2)]\} dudv =$$

$$= E(-x_2, -y_2).$$

Следовательно, $E_2(x_2, y_2) = E(-x_2, -y_2)$. Таким образом, схема с двойным преобразованием Фурье восстанавливает в задней фокальной плоскости второй линзы входной сигнал. Восстановленный сигнал в указанной плоскости является перевернутым. Если изменить направление координатных осей X_2 и Y_2 на противоположное, как показано на рис. 10.19, то можно записать

$$E_2(x_2, y_2) = F^{-1}\{E_1(x_1, y_1)\} = F^{-1}\{F\{E(x, y)\}\}.$$

Следовательно, сигнал на выходе можно представить как последовательное осуществление прямого и обратного пространственных фурие-преобразований.

Заметим, что преобразование, выполняемое второй линзой, соответствует обратному фурие-преобразованию (восстанавливает исходный сигнал), хотя фактически является прямым фурие-преобразованием.

Следует отметить при этом, что распределение поля в плоскости X_2Y_2 , определяемое двойным преобразованием Фурье поля во входной плоскости XU , можно представить как результат двойной дифракции световой волны от плоскости объекта XU до плоскости зрачка линзы и от плоскости зрачка до выходной плоскости X_2Y_2 .

Основным свойством линз является, конечно, их способность формировать оптическое изображение. Если поместить перед линзой предмет и осветить его, то при определенных условиях в другой плоскости возникнет распределение интенсивности, которое будет почти точной копией предмета. Изображение может быть действительным, если в плоскости за линзой возникает действительное распределение, и мнимым, если свет за линзой кажется исходящим из новой плоскости, расположенной

перед линзой. Вопросы формирования оптического изображения с помощью линз подробно будут рассмотрены нами в следующих главах.

10.3.5. Частотная характеристика тонкой линзы

Как модулятор тонкая линза обладает свойством преобразовывать пространственный спектр входного сигнала. Покажем это. Пусть входной сигнал определяется функцией $E(x, y)$, а его спектральная плотность

$$E(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp[-i2\pi(ux + vy)] dx dy .$$

Выходной сигнал $E'(x, y)$ определится соотношением (10.17). Спектральная плотность выходного сигнала

$$E'(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) t(x, y) \exp[-i2\pi(ux + vy)] dx dy .$$

Подставив сюда разложение

$$E(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) \exp[i2\pi(ux + vy)] du dv ,$$

для спектральной плотности выходного сигнала будем иметь

$$E'(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y) \exp[-i2\pi(ux + vy)] \times \\ \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(\xi, \eta) \exp[i2\pi(\xi x + \eta y)] d\xi d\eta dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(\xi, \xi) \times$$

$$\times \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y) \exp\{-i2\pi[(u - \xi)x + (v - \eta)y]\} dx dy \right) d\xi d\eta.$$

Введя обозначение

$$H(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y) \exp[-i2\pi(ux + vy)] dx dy, \quad (10.35)$$

получим

$$E'(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(\xi, \eta) H(u - \xi; v - \eta) d\xi d\eta.$$

Определенную соотношением (10.35) функцию $H(u, v)$ называют передаточной функцией или частотной характеристикой модулятора, функция пропускания которого $t(x, y)$. Следовательно, спектральная плотность выходного сигнала равна свертке спектральной плотности входного сигнала и частотной характеристики линзы. Это и доказывает указанное выше утверждение. Полученное соотношение, называемое также интегралом Дюамеля, имеет место для любого модулятора.

Если функция пропускания линзы определяется формулой (10.14), то ее частотная характеристика

$$\begin{aligned} H(u, v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-ik \frac{x^2 + y^2}{2f}\right) \exp[-i2\pi(ux + vy)] dx dy = \\ &= \int_{-\infty}^{\infty} \exp\left(-ik \frac{x^2}{2f} - i2\pi ux\right) dx \int_{-\infty}^{\infty} \exp\left(-ik \frac{y^2}{2f} - i2\pi vy\right) dy = \\ &= C \exp\left(i4\pi^2 f \frac{u^2 + v^2}{2k}\right) = C \exp[i\pi\lambda f(u^2 + v^2)]. \end{aligned}$$

Следовательно,

$$H(u, v) = C \exp[i2\pi\lambda f(u^2 + v^2)], \quad (10.36)$$

где C – некоторая постоянная (не зависящая от пространственных частот u и v).

Таким образом, свойства линзы как модулятора можно описать одной из двух функций: функцией пропускания $t(x, y)$ или частотной характеристикой $H(u, v)$. Заметим, что частотная характеристика есть фурье-преобразование функции пропускания. В этом смысле она является сопряженной по отношению к функции пропускания. Точно так же импульсная характеристика и передаточная функция слоя свободного пространства связаны друг с другом также преобразованием Фурье и, следовательно, являются сопряженными. Более того, они оказываются сопряженными и к соответствующим характеристикам линзы. Действительно, передаточная функция слоя свободного пространства и функция пропускания линзы при $z = f$, а также импульсная характеристика слоя и частотная характеристика линзы связаны между собой преобразованием Фурье. Поэтому линза и слой свободного пространства являются сопряженными оптическими элементами.

Отметим также, что свойства линзы, как и слоя свободного пространства, определяются только одним параметром. Таким параметром для линзы является ее фокусное расстояние f , а для слоя – его толщина z .

Кроме того, сравнивая выражения для импульсной характеристики слоя свободного пространства толщиной z и функции пропускания собирающей линзы с фокусным расстоянием f , т.е. соотношения (9.19) и (10.14), а также (9.20) и (10.36), видим, что эти функции имеют близкую структуру и их можно объединить. Характеристические параметры z и f , определяющие свойства этих оптических элементов, входят в знаменатель аргумента экспоненты. Для удобства записи обычно вводят так называемые диоптрийные параметры – обратное

фокусное расстояние $\Phi = 1/f$ для линзы и обратное расстояние $d = 1/z$ для слоя свободного пространства. Эти величины аналогичны электрической проводимости электрической цепи, которая равна величине, обратной сопротивлению цепи. Объединение указанных выше функций осуществляется введением так называемой обобщенной функции Френеля

$$Z_0(x, y; f) = \exp\left[i \frac{kf}{2}(x^2 + y^2)\right],$$

с помощью которой можно выразить как импульсную характеристику слоя свободного пространства

$$h(x, y) = \frac{d \exp(ik/d)}{i\lambda} Z_0(x, y; f),$$

так и функцию пропускания линзы

$$t(x, y) = Z_0^*(x, y; f).$$

Таким образом, мы показали, что тонкая линза, как и слой свободного пространства, обладает свойством преобразования пространственного спектра входящего в нее сигнала.

Глава 11

ФОРМИРОВАНИЕ ОПТИЧЕСКОГО ИЗОБРАЖЕНИЯ ПРИ КОГЕРЕНТНОМ ОСВЕЩЕНИИ

11.1. Искажение изображения при распространении света в свободном пространстве

Пусть транспарант с изображением произвольного предмета освещается нормально падающим на него монохроматическим пучком света, например от системы лазер – расширитель пучка, создающей когерентное освещение плоской волной. Если между

транспарантом и расположенным за ним экраном имеется только свободное пространство, то сходство между предметом и его изображением на экране будет весьма приблизительным, причем с увеличением расстояния все труднее в картине на экране уловить сходство с оригиналом. При распространении света в свободном пространстве теряется сходство между наблюдаемыми оптическими картинками в плоскости наблюдения (экран) и предметной плоскости (транспарант). Выясним причину этого явления.

Изображение предмета в плоскости $X'Y'$ будет определяться распределением интенсивности, а значит, и амплитуды светового поля в этой плоскости. Это поле может быть получено как сигнал на выходе слоя пространства, непосредственно примыкающего к плоскости транспаранта XY . При не слишком больших расстояниях z между плоскостями XY и $X'Y'$ это поле определится преобразованием Френеля:

$$E'(x', y') = \frac{e^{ikz}}{i\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp\left[ik \frac{(x' - x)^2 + (y' - y)^2}{2z} \right] dx dy. \quad (11.1)$$

Этот интеграл свертки позволяет объяснить, почему распространение света в свободном пространстве приводит к искажению изображения. Видно, что поле E' в любой точке (x', y') плоскости наблюдения $X'Y'$ с точностью до несущественных множителей перед интегралом получается суммированием полей от всех точек (x, y) предметного поля, помноженных на фазовый множитель, квадратично зависящий от координат точек предметного поля. Наличие этого множителя и приводит к искажению изображения.

Процессы, связанные с искажением изображения в свободном пространстве, удобно рассматривать, перейдя в пространство фурье-координат или, иначе, в область пространственных частот

u и v . Произведя преобразование Фурье соотношения (11.1) и учтя теорему свертки, получим

$$E'(u, v) = \\ = E(u, v) \exp\left(i \frac{2\pi z}{\lambda}\right) \cdot \exp[-i\pi\lambda z(u^2 + v^2)], \quad (11.2)$$

где $E(u, v)$ и $E'(u, v)$ – фурье-образ распределения поля в плоскости предмета и плоскости наблюдения соответственно, произведение двух последних сомножителей представляет собой передаточную функцию $H(u, v)$ свободного пространства в приближении Френеля. Из (11.2) видно, что действие свободного пространства приводит к изменению фазы плоских волн, достигающих плоскости наблюдения, на величину, пропорциональную квадратам пространственных частот u и v , причем это изменение будет тем больше, чем больше расстояние z от плоскости предмета до плоскости наблюдения (линейный фазовый множитель $\exp\left(i \frac{2\pi z}{\lambda}\right)$ на распределение частот не влияет). Это обстоятельство является определяющим для искажения изображения предмета, так как фазовые соотношения между слагаемыми плоскими волнами в плоскости предмета и плоскости наблюдения, отстоящей на расстояние z , различны. Различие фазовых соотношений приводит к различию результатов интерференции составляющих плоских волн. Поэтому результирующее распределение поля в плоскости наблюдения будет кардинально отличаться от распределения поля в плоскости предмета. Отметим, что это различие, а значит, и величина искажения изображения будут зависеть от расстояния z от предмета до плоскости наблюдения.

Проиллюстрируем сказанное на следующем примере. Пусть предметом является одномерная амплитудная гармоническая решетка с функцией пропускания

$$t(x) = t_0 + t_1 \cos 2\pi ux,$$

где коэффициенты t_0 и t_1 удовлетворяют тем же условиям, что и в соотношении (7.1). При освещении решетки плоской волной амплитуды E_0 распределение светового поля непосредственно за плоскостью предмета (при $z = 0$) можно представить в виде

$$E(x) = E_0 t_0 + \frac{1}{2} E_0 t_1 \exp(i2\pi ux) + \frac{1}{2} E_0 t_1 \exp(-i2\pi ux).$$

Данному граничному условию соответствует в свободном пространстве за предметом световое поле из трех плоских волн

$$E(x, z) = E_0 t_0 \exp(ikz) + \frac{1}{2} E_0 t_1 \exp[i2\pi(ux + wz)] + \\ + \frac{1}{2} E_0 t_1 \exp[-i2\pi(ux - wz)].$$

Считая, что $u \ll 1/\lambda$, пространственную частоту w представим в виде $w = \sqrt{1/\lambda^2 - u^2} \approx 1/\lambda - \frac{1}{2}\lambda u^2$. Тогда для поля в плоскости $X'Y'$ получим

$$E(x', z) = E_0 t_0 \exp(i2\pi z/\lambda) + \frac{1}{2} E_0 t_1 \exp[i2\pi(ux' + z/\lambda - \\ - \frac{1}{2}\lambda z u^2)] + \frac{1}{2} E_0 t_1 \exp[-i2\pi(ux' - z/\lambda + \frac{1}{2}\lambda z u^2)] = \\ = E_0 \exp(i2\pi z/\lambda) \{t_0 + \frac{1}{2} t_1 E_0 \exp(-i\pi\lambda z u^2) \times \\ \times [\exp(i2\pi ux') + \exp(-i2\pi ux')]\}.$$

Применив к выражению в квадратных скобках формулу Эйлера, для поля в плоскости $X'Y'$ будем иметь

$$E'(x', z) = E_0 \exp(i2\pi z / \lambda) [t_0 + t_1 \exp(-i\pi\lambda z u^2) \cos 2\pi u x'] \quad (11.3)$$

Выражение (11.3) показывает, что гармоническая составляющая функции пропускания решетки на пространственной частоте u умножается на квадратичный фазовый множитель $\exp(-i\pi\lambda z u^2)$.

Распределение интенсивности в плоскости наблюдения

$$I(x', z) \sim |E'(x', z)|^2 = |E_0|^2 [t_0^2 + 2t_0 t_1 \cos 2\pi u x' \times \cos \pi\lambda z u^2 + t_1^2 \cos^2 2\pi u x'].$$

Полученное выражение отличается от распределения интенсивности в плоскости предмета

$$I(x', 0) \sim |E'(x', 0)|^2 = |E_0|^2 [t_0^2 + 2t_0 t_1 \cos 2\pi u x' + t_1^2 \cos^2 2\pi u x']$$

фазовым множителем $\cos \pi\lambda z u^2$. Этот множитель равен единице при z , удовлетворяющих условию $\pi\lambda z u^2 = 2\pi m$ ($m = 0, 1, 2, 3, \dots$), т.е. при

$$z = m(2d^2 / \lambda), \quad (11.4)$$

где $d = 1 / u$ – пространственный период решетки. В этом случае распределение интенсивности светового поля в плоскости наблюдения $I(x', z)$ будет в точности совпадать с распределением интенсивности светового поля в плоскости предмета $I(x', 0)$ – изображение получается полностью идентичным предмету. На расстояниях z , не удовлетворяющих условию (11.4), изображение будет в большей или меньшей степени отличаться от предмета, т.е. будет искаженным. Следовательно, если экран перемещать вдоль оси Z , изображение решетки на экране будет то появляться, то исчезать с периодичностью

$$\Delta z = 2d^2 / \lambda.$$

Это явление, впервые описанное Талботом, называют эффектом самовоспроизведения. Заметим, что Δz определяется квадратом пространственного периода решетки. Это является следствием нелинейности фазовой характеристики свободного пространства по пространственным частотам u и v .

Если транспарант имеет произвольную функцию пропускания $t(x)$, то ее можно разложить в ряд или в интеграл Фурье, т.е. представить транспарант в виде конечного или бесконечного числа гармонических решеток с различными периодами $d = 1 / u$. Плоские волны от каждой такой решетки, распространяясь за транспарантом, будут давать их изображения на расстояниях вдоль оси Z , определяемых формулой (11.4). Однако в общем случае за транспарантом не будет такой плоскости, где все решетки дают свои изображения в фазе – изображение предмета будет искаженным.

Таким образом, искажение изображения предмета происходит из-за тех фазовых искажений, которые появляются при распространении световых волн в свободном пространстве между транспарантом и экраном. Восстановить изображение – это значит скомпенсировать эти искажения. Такая компенсация в оптике осуществляется при помощи линз, зонных пластинок или голографическим методом. Линза с соответствующим фокусным расстоянием и соответствующим образом расположенная между предметом и экраном так воздействует на фазу световых волн, что фазовые искажения компенсируются с избытком и только действие свободного пространства за линзой приводит к восстановлению изображения. Это согласуется с тем, что, как было отмечено в предыдущем параграфе, фокусирующим, а значит, и изображающим свойством обладает линза вместе с примыкающим к ней слоем свободного пространства. Линза компенсирует таким образом действие свободного пространства между предметом и линзой и между линзой и изображением, что и отражено в формуле линзы, связывающей между собой

фокусное расстояние линзы с толщинами слоев свободного пространства, расположенных перед линзой и за линзой. Однако, как мы увидим ниже, из-за конечных размеров линзы и возникающей вследствие этого дифракции на оправе линзы изображение предмета, даваемое линзой, не будет полностью идентично самому предмету. Оно будет несколько размытым, нечетким.

11.2. Изображение линзой точечного объекта

Пусть на оптической оси Z в начале координатной плоскости XU на расстоянии a от линзы имеется точечный источник света S . Будем последовательно определять распределение комплексных амплитуд поля в различных плоскостях оптической системы по ходу распространения световой волны от этого источника (вдоль положительного направления оси Z). В плоскости XU непосредственно за источником S он создает распределение поля, описываемое δ -функцией:

$$E(x, y) = E_0 \delta(x, y). \quad (11.5)$$

В плоскости непосредственно перед линзой распределение поля можно найти с помощью преобразования Френеля (11.1), подставив в него выражение (11.5) и заменив в нем x' и y' на координаты в плоскости линзы ξ и η , а z на a :

$$E_1(\xi, \eta) = E_0 \frac{e^{ika}}{i\lambda a} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x, y) \exp \left[ik \frac{(\xi - x)^2 + (\eta - y)^2}{2a} \right] dx dy.$$

Учитывая фильтрующее свойство δ -функции, получим

$$E_1(\xi, \eta) = E_0 \frac{e^{ika}}{i\lambda a} \exp \left(ik \frac{\xi^2 + \eta^2}{2a} \right).$$

Заметим, что такое же поле в плоскости $\xi\eta$, т.е. в плоскости $z = a$, создавала бы расходящаяся сферическая волна с комплексной амплитудой

$$E(x, y) = E_0 \frac{e^{ikz}}{i\lambda z} \exp\left(ik \frac{x^2 + y^2}{2z}\right).$$

Это подтверждает возможность представления точечного источника света, создающего расходящуюся сферическую волну, δ -функцией.

Поле в плоскости непосредственно за линзой определится умножением поля непосредственно перед линзой $E_1(\xi, \eta)$ на функцию пропускания линзы $t(\xi, \eta) = \exp\left(-ik \frac{\xi^2 + \eta^2}{2f}\right)$.

Конечные размеры линзы учтем умножением $t(\xi, \eta)$ на зрачковую функцию линзы $P(\xi, \eta)$. Учитывая это, находим

$$E_1'(\xi, \eta) = \frac{E_0 e^{ika}}{i\lambda a} P(\xi, \eta) \exp\left[i \frac{k}{2} \left(\frac{1}{a} - \frac{1}{f}\right) (\xi^2 + \eta^2)\right].$$

Поле в плоскости $X'Y'$ на расстоянии a' от линзы найдется с помощью преобразования Френеля

$$E'(x', y') = \frac{e^{ika'}}{i\lambda a'} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E_1'(\xi, \eta) \exp\left[ik \frac{(x' - \xi)^2 + (y' - \eta)^2}{2a'}\right] d\xi d\eta.$$

Возведя в квадрат выражения в круглых скобках в показателе экспоненты, подставив значение $E_1'(\xi, \eta)$ и сгруппировав члены, получим

$$E'(x', y') = -E_0 \frac{e^{ik(a+a')}}{\lambda^2 aa'} e^{ik \frac{x'^2 + y'^2}{2a'}} \times$$

$$\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\xi, \eta) \exp \left[\left[i \frac{k}{2} \left(\frac{1}{a} - \frac{1}{f} + \frac{1}{a'} \right) (\xi^2 + \eta^2) \right] \right] \times \\ \times \exp \left(\left[-ik \left(\frac{x'}{a'} \xi + \frac{y'}{a'} \eta \right) \right] \right) d\xi d\eta.$$

Так как зрачковая функция $P(\xi, \eta)$ равна нулю вне апертуры линзы, то интегрирование проводится в бесконечных пределах. Будем считать, что a и a' удовлетворяют формуле линзы

$$\frac{1}{a} + \frac{1}{a'} = \frac{1}{f}.$$

В этом случае плоскость $X'Y'$ является плоскостью изображения. Тогда квадратичный фазовый множитель в подынтегральном выражении выпадает (обращается в единицу). Учитывая далее, что

$$k(x'/a') = k \operatorname{tg} \alpha \approx k \sin \alpha = k_x = 2\pi u,$$

$$k(y'/a') = k \operatorname{tg} \beta \approx k \sin \beta = k_y = 2\pi v,$$

для поля в плоскости изображения $X'Y'$ будем иметь:

$$E'(x', y') = \\ = -E_0 \frac{e^{ik(a+a')}}{\lambda^2 a a'} e^{ik \frac{x'^2 + y'^2}{2a'}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\xi, \eta) \exp \left[-i(k_x \xi + k_y \eta) \right] d\xi d\eta. \quad (11.6)$$

Предположим теперь, что $P(\xi, \eta) = 1$, т.е. радиус линзы достаточно велик, так что все лучи, исходящие из источника S , проходят через линзу. В этом случае интеграл

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left[-i(k_x \xi + k_y \eta) \right] d\xi d\eta$$

в (11.6) будет представлять собой фурье-образ единицы. Последний равен δ -функции, умноженной на $(2\pi)^2$:

$$\int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} \exp[-i(k_x \xi + k_y \eta)] d\xi d\eta = (2\pi)^2 \delta(k_x, k_y) =$$

$$= (2\pi)^2 \delta\left(\frac{x'}{a'}, \frac{y'}{a'}\right).$$

Поскольку изображение наблюдают или регистрируют в виде распределения интенсивности, а при определении интенсивности фазовый множитель в комплексной амплитуде выпадает, то фазовые множители в (11.6) можно положить равными единице. Тогда получим

$$E(x', y') = -E_0 \frac{4\pi^2}{\lambda^2 aa'} \delta\left(\frac{x'}{a'}, \frac{y'}{a'}\right).$$

Правая часть этого выражения обращается в нуль всюду, кроме точки с координатами $x' = y' = 0$. Отсюда видно, что падающая на линзу сферическая волна сходится в плоскости изображения в математическую точку с координатами $x' = y' = 0, z = a + a'$.

Таким образом, даваемое линзой неограниченных размеров изображение точечного объекта, расположенного на оптической оси (оси Z) на расстоянии a от линзы, представляет собой геометрическую точку, расположенную на расстоянии a' от линзы, причем расстояния a и a' связаны между собой формулой линзы. Следовательно, формула линзы описывает условие, при котором сферическое световое поле будет сходиться и воспроизводить изображение точечного объекта в плоскости изображения. Полученный результат полностью согласуется с геометрической теорией изображения, согласно которой точка в пространстве предметов изображается идеальной линзой в виде точки в пространстве изображений. На языке волн это означает, что идеальная линза превращает сферическую изофазную

поверхность предметной волны в сферическую поверхность изображающей волны с иным радиусом кривизны. Центры кривизны этих волн представляют собой точечный объект и его изображение, трактуемое с точки зрения геометрической оптики.

11.3. Изображение линзой протяженного предмета

Предположим теперь, что в плоскости XU расположен транспарант, функция пропускания которого $t(x, y)$. Транспарант освещается плоской монохроматической волной с комплексной амплитудой $E(z) = E_0 \exp(ikz)$. В плоскости непосредственно за транспарантом эта волна создает световое поле $E(x, y) = E_0 t(x, y)$. Оптическая система, содержащая модулирующий элемент (в нашем случае транспарант), слой свободного пространства, собирающую сферическую линзу и затем снова слой свободного пространства, называется элементарной оптической системой или оптическим каскадом. Для расчета распределения поля в плоскости изображения $X'U'$ оптический каскад удобно представить в виде двух последовательно соединенных линейных систем, для каждой из которых сигнал на выходе соответствует распределению поля в области дифракции Френеля. Функциональная схема процесса образования изображения показана на рис. 11.1.

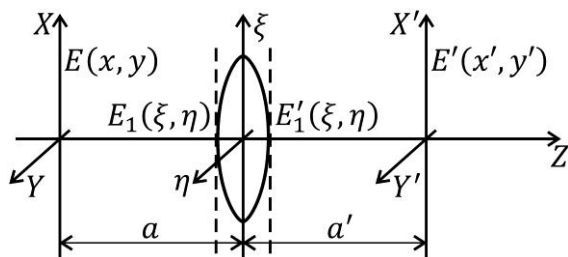


Рис. 11.1

В результате прохождения первого свободного пространства протяженностью a поле в плоскости непосредственно перед линзой определится преобразованием Френеля

$$E_1(\xi, \eta) = \frac{e^{ika}}{i\lambda a} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[ik \frac{(\xi - x)^2 + (\eta - y)^2}{2a} \right] dx dy.$$

(11.7)

Это можно переписать в виде свертки

$$E_1(\xi, \eta) = E(x, y) * h_1(x, y, \xi, \eta),$$

где

$$h_1(x, y; \xi, \eta) = \frac{e^{ika}}{i\lambda a} \exp \left[ik \frac{(\xi - x)^2 + (\eta - y)^2}{2a} \right]$$

– импульсный отклик первого слоя свободного пространства, ξ, η – координаты в плоскости линзы. Соотношение (11.7) описывает зависимость между входом и выходом аналога линейной системы и выражает тот факт, что распределение комплексного светового поля непосредственно перед линзой обусловлено распределением комплексного светового поля в плоскости предмета.

После прохождения линзы поле в плоскости непосредственно за линзой будет

$$E_1'(\xi, \eta) = E_1(\xi, \eta) t(\xi, \eta) P(\xi, \eta).$$

Оптический сигнал $E_1'(\xi, \eta)$ является входным сигналом второй системы – второго свободного пространства глубиной a' . Поле в плоскости изображения найдется как свертка функции $E_1'(\xi, \eta)$ с импульсным откликом

$$h_2(\xi, \eta, x', y') = \frac{e^{ika'}}{i\lambda a'} \exp \left[ik \frac{(x' - \xi)^2 + (y' - \eta)^2}{2a'} \right]$$

второго свободного пространства:

$$E'(x', y') = E_1'(\xi, \eta) * h_2(\xi, \eta, x', y').$$

Таким образом, поле в плоскости изображения

$$E'(x', y') = \{ [E(x, y) * h_1(x, y, \xi, \eta)] t(\xi, \eta) \times \\ \times P(\xi, \eta) \} * h_2(\xi, \eta, x', y').$$

Это выражение можно записать в интегральной форме:

$$E'(x', y') = - \frac{e^{ik(a+a')}}{\lambda^2 aa'} \times \\ \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) P(\xi, \eta) \exp \left\{ ik \left[\frac{(\xi - x)^2 + (\eta - y)^2}{2a} - \frac{\xi^2 + \eta^2}{2f} + \right. \right. \\ \left. \left. + \frac{(x' - \xi)^2 + (y' - \eta)^2}{2a'} \right] \right\} d\xi d\eta dx dy. \quad (11.8)$$

Мы видим, что распределение поля в плоскости изображения можно представить как результат двойной дифракции световой волны от плоскости объекта до плоскости линзы и от плоскости линзы до плоскости изображения (подробнее об этом в п. 11.7). Произведение множителя перед интегралом на двойной интеграл по ξ, η от экспоненциального множителя обозначим через $h(x' - x, y' - y)$. Тогда поле $E'(x', y')$ можно представить в виде интеграла свертки:

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) h(x' - x; y' - y) dx dy, \quad (11.9)$$

где

$$\begin{aligned}
 h(x' - x; y' - y) \equiv h(x, y, x', y') = & - \frac{\exp[ik(a + a')]}{\lambda^2 aa'} \times \\
 & \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\xi, \eta) \exp \left\{ ik \left[\frac{(\xi - x)^2 + (\eta - y)^2}{2a} - \frac{\xi^2 + \eta^2}{2f} + \right. \right. \\
 & \left. \left. + \frac{(x' - x)^2 + (y' - y)^2}{2a'} \right] \right\} d\xi d\eta. \tag{11.10}
 \end{aligned}$$

Функцию $h(x, y, x', y')$ называют импульсной характеристикой (или функцией пропускания) элементарной оптической системы. Она представляет собой поле точечного источника в точке (x, y) , рассматриваемое в точке (x', y') плоскости изображения $X'Y'$. Раскрыв скобки в экспоненциальном множителе, после некоторых преобразований для импульсной функции будем иметь

$$\begin{aligned}
 h(x, y, x', y') = & - \frac{e^{ik(a+a')}}{\lambda^2 aa'} \cdot \exp \left(ik \frac{x^2 + y^2}{2a} \right) \times \\
 & \times \exp \left(ik \frac{x'^2 + y'^2}{2a'} \right) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\xi, \eta) \exp \left[i \frac{k}{2} \left(\frac{1}{a} + \frac{1}{a'} - \right. \right. \\
 & \left. \left. - \frac{1}{f} \right) (\xi^2 + \eta^2) \right] \exp \left\{ -ik \left[\left(\frac{x}{a} + \frac{x'}{a'} \right) \xi + \left(\frac{y}{a} + \frac{y'}{a'} \right) \eta \right] \right\} d\xi d\eta. \tag{11.11}
 \end{aligned}$$

По указанным в предыдущем параграфе соображениям фазовые множители перед интегралом в $h(x, y, x', y')$, не зависящие от переменных интегрирования x, y , можно положить равными

единице. Если считать, что $(k/2a)(x^2 + y^2)_{\max} \ll \pi$, т.е. если $a \gg (x^2 + y^2)_{\max} / \lambda$, первый квадратичный фазовый множитель тоже можно положить равным единице. А поскольку a и a' удовлетворяют формуле линзы, то и квадратичный фазовый множитель в подынтегральном выражении будет равен единице. С учетом этого для импульсной функции идеальной элементарной оптической системы будем иметь

$$h(x, y, x', y') = -\frac{1}{\lambda^2 a a'} \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\xi, \eta) \exp \left\{ -ik \left[\left(\frac{x}{a} + \frac{x'}{a'} \right) \xi + \left(\frac{y}{a} + \frac{y'}{a'} \right) \eta \right] \right\} d\xi d\eta. \quad (11.12)$$

Полученное выражение показывает, что с точностью до постоянного множителя перед интегралом импульсная функция элементарной оптической системы представляет собой фурье-образ функции зрачка на пространственных частотах $\frac{1}{\lambda} \left(\frac{x}{a} + \frac{x'}{a'} \right)$

и $\frac{1}{\lambda} \left(\frac{y}{a} + \frac{y'}{a'} \right)$. Можно сказать, что импульсный отклик элементарной оптической системы есть дифрагировавшее на оправе линзы поле точечного источника (в приближении Фраунгофера).

Введем величину $\beta = x'/x = y'/y = a'/a$, называемую коэффициентом линейного увеличения линзы. Подставив в выражение (11.12) значение $a = a' / \beta$ и учтя, что $k = 2\pi / \lambda$, выражение для импульсной функции представим в виде

$$h(x, y, x', y') = -\frac{\beta}{(\lambda a')^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\xi, \eta) \times$$

$$\times \exp \left\{ -i \frac{2\pi}{\lambda a'} [(x' + \beta x)\zeta + (y' + \beta y)\eta] \right\} d\zeta d\eta.$$

Произведем замену переменных $\xi / \lambda a' = x_1$, $\eta / \lambda a' = y_1$. Тогда получим

$$h(x, y, x', y') = -\beta \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} P(\lambda a' x_1, \lambda a' y_1) \times \\ \times \exp \left\{ -i 2\pi [(x' + \beta x)x_1 + (y' + \beta y)y_1] \right\} dx_1 dy_1. \quad (11.13)$$

Положим теперь, что зрачковая функция $P(\lambda a' x_1, \lambda a' y_1) = 1$. Это допущение предполагает, что линза перехватывает все лучи, идущие со стороны предмета. Тогда интеграл (11.13) будет представлять собой δ -функцию и, следовательно,

$$h(x, y, x', y') \approx -\beta \delta(x' + \beta x, y' + \beta y)$$

или, учитывая, что $\delta(mx, ny) = |mn|^{-1} \delta(x, y)$,

$$h(x, y, x', y') = -\frac{1}{\beta} \delta \left(x + \frac{x'}{\beta}, y + \frac{y'}{\beta} \right).$$

Подставляя это в интеграл суперпозиции (11.9), для распределения поля в плоскости изображения получим

$$E'(x', y') = -\frac{1}{\beta} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \delta \left(x + \frac{x'}{\beta}, y + \frac{y'}{\beta} \right) dx dy.$$

Учитывая фильтрующее свойство δ -функции, окончательно находим

$$E'(x', y') = -\frac{1}{\beta} E \left(-\frac{x'}{\beta}, -\frac{y'}{\beta} \right). \quad (11.14)$$

Мы видим, что каждой точке (x, y) предмета соответствует определенная точка (x', y') изображения, как и должно быть по геометрической теории оптических изображений. Распределение поля в плоскости изображения получается заменой в распределении поля в плоскости предмета координат x и y на координаты $-x'/\beta$ и $-y'/\beta$ и умножением поля на коэффициент $-1/\beta$. Световое поле в плоскости изображения изменяется обратно пропорционально коэффициенту линейного увеличения линзы. При этом изменяется масштаб осей в β раз, а их направление – на обратное. Таким образом, изображение, получающееся в пренебрежении конечными размерами и аберрациями линзы, представляет собой точную копию предмета, увеличенную в β раз и перевернутую в плоскости изображения, т.е. точно такое, какое получается в геометрической оптике.

В приближении геометрической оптики, которое имеет место при $\lambda \rightarrow 0$, функция зрачка $P(\lambda a'x_1, \lambda a'y_1)$ принимается равной единице при всех значениях переменных x_1 и y_1 . Действительно, если радиус оправы линзы равен R , то функция зрачка $P(\lambda a'x_1, \lambda a'y_1) = 1$ для всех $\lambda a'x_1$ и $\lambda a'y_1$, лежащих внутри круга

$$(\lambda a'x_1)^2 + (\lambda a'y_1)^2 \leq R^2.$$

Отсюда получаем $x_1^2 + y_1^2 \leq (R/\lambda a')^2$. Следовательно, при $\lambda \rightarrow 0$ имеем $x_1 \rightarrow \infty$, $y_1 \rightarrow \infty$. Очевидно, что переход к пределу $\lambda \rightarrow 0$ равносильен переходу к линзе неограниченных размеров ($R \rightarrow \infty$). Линза, неограниченных размеров и не имеющая аберраций, называется идеальной. Идеальная линза дает идеальное изображение – геометрически подобное предмету.

Соотношение (11.14) показывает, что рассматриваемая оптическая система является линейной по отношению к комплексной амплитуде. Но так и должно быть при когерентном освещении. В идеальной оптической системе (в смысле геометрической оптики) формирования изображения, как мы

видели, импульсная функция является δ -функцией. Соответствующая передаточная функция

$$H(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta\left(x + \frac{x'}{\beta}, y + \frac{y'}{\beta}\right) \exp[-i2\pi(ux + vy)] dx dy = \\ = \exp\left[i \frac{2\pi}{\beta}(ux' + vy')\right].$$

Так как в итоге регистрируется интенсивность изображения, то интерес представляет величина $|H(u, v)|^2$. Эта величина равна единице при всех u и v . Следовательно, идеальная оптическая система формирования изображения имеет передаточную характеристику с неограниченной полосой пространственных частот.

Следует отметить также, что при выполнении формулы линзы квадратичный фазовый множитель в подынтегральном выражении (11.11), т.е. тот, который приводит к искажению изображения при распространении света в свободном пространстве, исчезает. Тем самым линза, как уже отмечалось, компенсирует действие свободного пространства между предметом и линзой и между линзой и изображением.

Полученные выше выражения и выражения, которые будут получены ниже в этой главе, можно считать справедливыми для любой оптической системы, формирующей изображение. Все конструктивные усложнения реальных систем служат для того, чтобы приблизить их свойства к свойствам идеализированной тонкой линзы. Дифракционные эффекты, возникающие в сложной системе можно интерпретировать как эффекты дифракции на входном зрачке системы.

Рассмотрим случай, когда $a' = f$, т.е. плоскость $X'Y'$ является задней фокальной плоскостью линзы. Оптическую систему будем считать идеальной, т.е. что $P = 1$. Импульсная функция

элементарной оптической системы (11.11) в этом случае принимает вид

$$\begin{aligned}
 h(x, y, x', y') = & -\frac{e^{ik(a+f)}}{\lambda^2 af} \cdot \exp\left(ik \frac{x^2 + y^2}{2a}\right) \times \\
 & \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(ik \frac{\xi^2 + \eta^2}{2a}\right) \exp\left(ik \frac{x'^2 + y'^2}{2f}\right) \times \\
 & \times \exp\left\{-ik \left[\left(\frac{x}{a} + \frac{x'}{f}\right)\xi + \left(\frac{y}{a} + \frac{y'}{f}\right)\eta\right]\right\} d\xi d\eta. \quad (11.15)
 \end{aligned}$$

Интеграл в этом выражении можно вычислить, используя формулу (9.21). Полагая

$$\alpha = \frac{k}{2ia}, \quad \beta_1 = ik \left(\frac{x}{a} + \frac{x'}{f}\right), \quad \beta_2 = ik \left(\frac{y}{a} + \frac{y'}{f}\right),$$

получим

$$\begin{aligned}
 & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(ik \frac{\xi^2 + \eta^2}{2a}\right) \exp\left\{-ik \left[\left(\frac{x}{a} + \frac{x'}{f}\right)\xi + \left(\frac{y}{a} + \frac{y'}{f}\right)\eta\right]\right\} d\xi d\eta = \\
 & = \frac{i2\pi a}{k} \exp\left(-ik \frac{x^2 + y^2}{2a}\right) \cdot \exp\left[-ik \frac{a(x'^2 + y'^2)}{2f^2}\right] \times \\
 & \times \exp\left(-ik \frac{xx' + yy'}{f}\right).
 \end{aligned}$$

Тогда получим

$$h(x, y, x', y') = \frac{1}{\lambda f} \exp\left[ik \left(1 - \frac{a}{f}\right) \frac{x'^2 + y'^2}{2f}\right] \times$$

$$\times \exp \left[-i \left(\frac{kx'}{f} x + \frac{ky'}{f} y \right) \right].$$

Фазовый множитель

$$-i \exp [ik(a + f)] = \exp \{i[k(a + f) - \pi/2]\}$$

опущен, так как он не влияет на распределение поля в плоскости $X'Y'$. Подставляя выражение $h(x, y, x', y')$ в исходное соотношение (11.9) и учитывая, что $k = 2\pi/\lambda$, получим

$$E'(x', y') = \frac{1}{\lambda f} \exp \left[ik \left(1 - \frac{a}{f} \right) \frac{x'^2 + y'^2}{2f} \right] \times \\ \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[-i2\pi \left(\frac{x'}{\lambda f} x + \frac{y'}{\lambda f} y \right) \right] dx dy$$

или

$$E'(x', y') = \frac{1}{\lambda f} \exp \left[ik \left(1 - \frac{a}{f} \right) \frac{x'^2 + y'^2}{2f} \right] E \left(\frac{x'}{\lambda f}, \frac{y'}{\lambda f} \right).$$

Мы получили выражение, почти совпадающее с выражением (10.27), если в нем заменить величины с индексом 1 на величины со штрихом, а z на a . Оно показывает, что в задней фокальной плоскости линзы с точностью до множителей перед интегралом формируется фурье-спектр распределения поля в плоскости предмета. Заметим, что если предметная плоскость совпадает с передней фокальной плоскостью линзы, т.е. если $a = f$, то фазовый множитель перед интегралом обратится в единицу и мы получаем точное преобразование Фурье на пространственных частотах

$$u = x' / (\lambda f), \quad v = y' / (\lambda f).$$

Таким образом, идеальная элементарная оптическая система позволяет получать одновременно и пространственный спектр, и изображение сигнала. В плоскости изображения формируется изображение входного сигнала, а в фокальной плоскости линзы – его спектр.

Рассмотрим теперь случай, когда предмет находится в передней фокальной плоскости линзы, т.е. когда $a = f$, а наблюдение распределения поля производится на произвольном расстоянии b от него. Проведем расчеты, подобные предыдущим, в предположении $P = 1$, для этого случая получим

$$h(x, y, x', y') = \frac{1}{i\lambda f} \exp \left[ik \left(1 - \frac{b}{f} \right) \frac{x^2 + y^2}{2f} \right] \exp \left[-i \left(\frac{kx'}{f} x + \frac{ky'}{f} y \right) \right].$$

Несущественный фазовый множитель $\exp[ik(a' + f)]$ здесь опущен. Распределение поля в плоскости наблюдения

$$E'(x', y') = \frac{1}{i\lambda f} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[ik \left(1 - \frac{a'}{f} \right) \frac{x^2 + y^2}{2f} \right] \times \\ \times \exp \left[-i \left(\frac{kx'}{f} x + \frac{ky'}{f} y \right) \right] dx dy.$$

В этом случае, в отличие от предыдущего, не существует однозначного соответствия между наблюдаемыми и исходными точками объекта или его спектра. Преобразуем это выражение, подставив в него обратное преобразование Фурье функции $E(x, y)$, получим

$$E'(x', y') = \frac{1}{i\lambda f} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) \exp \left[ik \left(1 - \frac{a'}{f} \right) \frac{x^2 + y^2}{2f} + \right.$$

$$\begin{aligned}
& + i \left(2\pi u - \frac{kx'}{f} \right) x + i \left(2\pi v - \frac{ky'}{f} \right) y \Big] dx dy dudv = \\
& = \frac{1}{\lambda f} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left[ik \left(1 - \frac{a'}{f} \right) \frac{1}{2f} (x^2 + y^2) + \right. \right. \\
& \left. \left. + i2\pi \left(u - \frac{x'}{\lambda f} \right) x + i2\pi \left(v - \frac{y'}{\lambda f} \right) y \right] dx dy \right) dudv,
\end{aligned}$$

где учтено, что $k = 2\pi/\lambda$.

Внутренний интеграл, который мы обозначим через I , можно вычислить. Используя формулу (9.21) при значениях параметров

$$\alpha = \frac{\pi(1 - a'f^{-1})}{i\lambda f}, \quad \beta_1 = i2\pi \left(u - \frac{x'}{\lambda f} \right), \quad \beta_2 = i2\pi \left(v - \frac{y'}{\lambda f} \right),$$

будем иметь

$$\begin{aligned}
I & = \frac{i\lambda f}{1 - a'f^{-1}} \times \\
& \times \exp \left[-i\pi\lambda \frac{f^2}{f - a'} (u^2 + v^2) + i2\pi(ux' + vy') \frac{f}{f - a'} - ik \frac{x'^2 + y'^2}{2(f - a')} \right].
\end{aligned}$$

Учитывая это, находим

$$\begin{aligned}
E'(x', y') & = \frac{f}{f - a'} \exp \left[-ik \frac{x'^2 + y'^2}{2(f - a')} \right] \times \\
& \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(u, v) \exp \left[-i\pi\lambda \frac{f^2}{f - a'} (u^2 + v^2) \right] \times
\end{aligned}$$

$$\times \exp \left[i2\pi \left(u \frac{f}{f-a'} x' + v \frac{f}{f-a'} y' \right) \right] dudv.$$

Сравнивая полученное выражение с выражением (5.12), которое в приближении $u^2 + v^2 \ll 1/\lambda^2$, когда

$$2\pi z \sqrt{1/\lambda^2 - u^2 - v^2} = kz - \pi\lambda z(u^2 + v^2),$$

можно записать в виде

$$\begin{aligned} E'(x', y', z) &= \\ &= e^{ikz} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(u, v) \exp[-i\pi\lambda z(u^2 + v^2) \exp[i2\pi(ux' + vy')]] dudv. \end{aligned}$$

Мы видим, что они аналогичны и с точностью до фазовых множителей перед интегралами эти выражения совпадают при условии, что

$$z = f^2 / (f - a'). \quad (11.16)$$

Используя это совпадение, можно записать

$$\begin{aligned} E'(x', y') &= \frac{f}{f-a'} \exp \left[-ik \frac{x'^2 + y'^2}{2(f-a')} \right] \exp \left(-ik \frac{f^2}{f-a'} \right) \times \\ &\times E \left(\frac{f}{f-a'} x', \frac{f}{f-a'} y', \frac{f^2}{f-a'} \right). \end{aligned}$$

Теперь также наблюдается однозначное соответствие поля изображения с преобразованным полем предмета. Преобразование состоит в том, что изменяется масштаб поля предмета по осям X и Y в $f/(f-a')$ раз и оно проходит путь в свободном пространстве, определяемый величиной (11.16). Эта величина может быть больше или меньше нуля в зависимости от

того, будет ли расстояние b меньше или больше фокусного расстояния линзы f .

Можно показать, что эти два рассмотренных случая ($a = f$ и $a' = f$) включают в себя все остальные случаи, в том числе и случай, когда расстояния a и b удовлетворяют формуле линзы.

11.4. Образование изображения при когерентном освещении как двойное преобразование Фурье

Покажем, что комплексная амплитуда светового поля в плоскости изображения связана с комплексной амплитудой в плоскости предмета через два преобразования Фурье (прямого и обратного). Для этого рассмотрим снова процесс образования изображения элементарной оптической системой. Как и прежде, будем считать линзу неограниченной, поэтому интегрирование будем проводить в бесконечных пределах.

Световое поле в плоскости непосредственно перед линзой определится сверткой функции предмета $E(x, y)$ с импульсным откликом свободного пространства перед линзой:

$$\begin{aligned}
 E_1(\xi, \eta) &= \frac{e^{ika}}{i\lambda a} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[ik \frac{(\xi - x)^2 + (\eta - y)^2}{2a} \right] dx dy = \\
 &= \frac{\exp \left(ik \frac{\xi^2 + \eta^2}{2a} \right)}{i\lambda a} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left(ik \frac{x^2 + y^2}{2a} \right) \times \\
 &\quad \times \exp \left[-i \frac{k}{a} (x\xi + y\eta) \right] dx dy. \quad (11.17)
 \end{aligned}$$

В конечном выражении фазовый множитель $\exp(ika)$, одинаковый для всех точек (ξ, η) и учитывающий только набег фазы при распространении света от плоскости предмета до

плоскости линзы, опущен. Интеграл (11.17) представляет собой преобразование Фурье от функции $E(x, y)\exp\left(ik\frac{x^2 + y^2}{2a}\right)$ на пространственных частотах $k_x = k(\xi/a)$, $k_y = k(\eta/a)$.

Введем обозначение

$$F\left\{E(x, y)\exp\left(ik\frac{x^2 + y^2}{2a}\right)\right\} = F\left(\frac{k}{a}\xi, \frac{k}{a}\eta\right).$$

С учетом этого для поля непосредственно за линзой можно записать выражение

$$\begin{aligned} E_1'(\xi, \eta) &= E_1(\xi, \eta) \exp\left(-ik\frac{\xi^2 + \eta^2}{2f}\right) = \\ &= F\left(\frac{k}{a}\xi, \frac{k}{a}\eta\right) \exp\left(-ik\frac{\xi^2 + \eta^2}{2f}\right) = \\ &= \frac{1}{i\lambda a} F\left(\frac{k}{a}\xi, \frac{k}{a}\eta\right) \exp\left(-ik\frac{\xi^2 + \eta^2}{2a'}\right), \end{aligned}$$

где учтено также, что в соответствии с формулой линзы

$$1/a - 1/f = 1/a'.$$

Поле в плоскости изображения определится сверткой функции $E_1'(\xi, \eta)$ с импульсным откликом второго свободного пространства:

$$\begin{aligned} E'(x', y') &= \frac{1}{i\lambda a'} \iint_{-\infty-\infty}^{\infty\infty} E'(\xi, \eta) \exp\left[ik\frac{(x' - \xi)^2 + (y' - \eta)^2}{2a'}\right] d\xi d\eta = \\ &= -\frac{1}{\lambda^2 aa'} \iint_{-\infty-\infty}^{\infty\infty} F\left(\frac{k}{a}\xi, \frac{k}{a}\eta\right) \exp\left(-ik\frac{\xi^2 + \eta^2}{2a'}\right) \times \end{aligned}$$

$$\begin{aligned} & \times \exp\left(ik \frac{x'^2 + y'^2}{2a'}\right) d\zeta d\eta = \\ & = -\frac{\exp\left(ik \frac{x'^2 + y'^2}{2a'}\right)}{\lambda^2 aa'} \times \\ & \quad \times \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} F\left(\frac{k}{a}\zeta, \frac{k}{a}\eta\right) \exp\left[-i \frac{k}{a'}(x'\zeta + y'\eta)\right] d\zeta d\eta. \end{aligned}$$

В этом выражении фазовый множитель $\exp(ika')$ опущен как не влияющий на распределение амплитуды поля в плоскости $X Y'$

Перейдем к новым переменным $p = -(k/a)\zeta$, $q = -(k/a)\eta$. Тогда

$$\begin{aligned} E'(x', y') &= -\frac{\exp\left(ik \frac{x'^2 + y'^2}{2a'}\right)}{\lambda^2 aa'} \frac{a^2}{k^2} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} F(-p, -q) \times \\ & \times \exp\left[i\left(\frac{a}{a'}x'p + \frac{a}{a'}y'q\right)\right] dpdq. \end{aligned}$$

Интеграл в этом выражении является обратным фурье-преобразованием функции $F(-p, -q)$. Учитывая свойство фурье-преобразований $F^{-1}\{F\{f(x, y)\}\} = 4\pi^2 f(x, y)$ и заменив в функции $f(x, y) = E(x, y) \exp\left(ik \frac{x^2 + y^2}{2a}\right)$ переменные x, y соответственно на $-(a/a')x', -(a/a')y'$, получим

$$E'(x', y') = -\frac{a}{a'} E\left(-\frac{a}{a'}x', -\frac{a}{a'}y'\right) \times$$

$$\begin{aligned}
& \times \exp \left\{ i \frac{k}{2a} \left[\left(-\frac{a}{a'} x' \right)^2 + \left(-\frac{a}{a'} y' \right)^2 \right] \right\} \exp \left(ik \frac{x'^2 + y'^2}{2a'} \right) = \\
& = -\frac{a}{a'} \exp \left(ik \frac{a}{a'} \frac{x'^2 + y'^2}{2f} \right) E \left(-\frac{a}{a'} x', -\frac{a}{a'} y' \right) = \\
& = -\exp \left(ik \frac{1}{\beta} \frac{x'^2 + y'^2}{2f} \right) \cdot \frac{1}{\beta} E \left(-\frac{x'}{\beta}, -\frac{y'}{\beta} \right),
\end{aligned}$$

где $\beta = a' / a$ – коэффициент увеличения линзы. При получении этого выражения мы снова учли формулу линзы и что $k = 2\pi / \lambda$.

Квадратичный фазовый множитель $\exp \left(ik \frac{1}{\beta} \frac{x'^2 + y'^2}{2f} \right)$ определяет фазовые искажения изображения. При

$$\frac{1}{\beta} \frac{k}{2f} (x'^2 + y'^2)_{\max} \ll \pi$$

указанный фазовый множитель можно положить равным единице, и тогда

$$E'(x', y') = -\frac{1}{\beta} E \left(-\frac{x'}{\beta}, -\frac{y'}{\beta} \right).$$

Мы снова пришли к выражению (11.14).

Тем самым мы показали, что распределение комплексной амплитуды в изображении представляет собой двойное преобразование Фурье:

$$E'(x', y') = F^{-1} \{ F \{ E(x, y) \} \cdot t(x, y) \}.$$

Это можно проиллюстрировать и на эквивалентной оптической схеме, приведенной на рис. 10.2. Амплитуда волны в задней фокальной плоскости первой линзы равна фурье-образу

амплитуды в плоскости предмета, а амплитуда волны в плоскости изображения в свою очередь является фурье-образом амплитуды волны света в задней фокальной плоскости первой линзы. Процесс образования изображения можно рассматривать и как последовательность двух процессов дифракции Фраунгофера, описываемой преобразованием Фурье.

11.5. Влияние на изображение конечных размеров линзы

Будем считать теперь, что линза имеет конечный радиус R . Рассмотрим опять сначала точечный объект, расположенный на расстоянии a от линзы. Если a и a' связаны между собой формулой линзы, то в соответствии с выражением (11.6) определение распределения поля в плоскости изображения сводится к вычислению интеграла

$$I = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} P(\xi, \eta) \exp[-i(k_x \xi + k_y \eta)] d\xi d\eta,$$

т.е. к фурье-преобразованию зрачковой функции линзы. Зрачковая функция имеет такой же вид, как и апертурная функция круглого отверстия, поэтому можно воспользоваться результатом, полученным при рассмотрении дифракции Фраунгофера на круглом отверстии. В п. 6.7 было показано, что

$$I(\alpha) = I(0) \left(\frac{2J_1(\alpha)}{\alpha} \right)^2. \quad (11.18)$$

Здесь $J_1(\alpha)$ – функция Бесселя первого рода первого порядка,

$$\alpha = Rk_{\perp} = R\sqrt{k_x^2 + k_y^2},$$

или, учитывая, что

$$k_x = k(x'/a'), \quad k_y = k(y'/a'), \quad r' = \sqrt{x'^2 + y'^2}, \quad \alpha = \frac{2\pi R}{\lambda a'} r'.$$

Функция $\left(\frac{2J_1(\alpha)}{\alpha}\right)^2$ показана на рис. 6.8. Она имеет максимум при $\alpha = 0$, т.е. при $r' = 0$. Первый минимум этой функции достигается при $\alpha = 3,83$, т.е. на расстоянии $r'_1 = 0,61\lambda a' / R$ от начала координат плоскости $X'Y'$, расположенного на оптической оси линзы. Отсюда следует, что в отличие от линзы с бесконечными размерами, которая фокусирует плоскую волну в математическую точку (дельта-функция в выражении (11.5)), линза с конечным радиусом R преобразует падающую на нее часть плоской волны в пятно конечной ширины. Изображение точечного источника, даваемое линзой конечных размеров, не является точечным, как это было в случае неограниченной линзы, а размывается в целую дифракционную картину в виде светового кружка (кружка Эри) радиусом r'_1 , окруженного чередующимися темными и светлыми кольцами. Центр дифракционной картины расположен на оси линзы в точке идеального изображения $x' = y' = 0$.

Выражение (11.18) можно получить и рассматривая образование распределения поля в плоскости изображения как результат двойной дифракции световой волны от плоскости объекта до плоскости зрачка линзы и от плоскости зрачка до плоскости изображения. Так как оптическая система предполагается идеальной, дифракцию можно считать фраунгоферовской. Более точно распределение поля на зрачке соответствует фурье-образу распределения поля на входной плоскости. Следовательно, распределение поля в плоскости изображения найдется как

$$E'(x', y') = F^{-1}\{F\{E(x, y)\}P(\xi, \eta)\},$$

где F – обозначение фурье-преобразования с ядром $\exp \left[-i \frac{k}{2f} (x\xi + y\eta) \right]$, $P(\xi, \eta)$ – функция зрачка линзы. Отсюда для точечного источника, расположенного в начале координат ($x = 0, y = 0$), будем иметь

$$E'(x', y') = F^{-1} \{ F \{ \delta(x, y) \} P(\xi, \eta) \} = C \frac{2J_1(\alpha)}{\alpha}.$$

Как уже отмечалось в п.6.7, изображение точки в виде пятна конечных размеров является следствием ограничения фронта падающей на оптическую систему световой волны, обусловленного конечными поперечными размерами диафрагмы и самой линзы. На краях диафрагмы или на оправе (на апертуре) линзы возникает более или менее заметная дифракция. Световая волна после дифракции на линзе будет создавать в фокальной плоскости линзы или в плоскости, где возникает изображение источника света, дифракционную картину. В этом отношении линза конечных размеров ведет себя подобно круглому отверстию в непрозрачном экране при дифракции Фраунгофера. Таким образом, дифракция света ставит определенный предел получению стигматического изображения. Никакая, даже самая совершенная линза не дает идеального изображения, изображением точки будет являться не точка, а целая дифракционная картина Эри. Распределение интенсивности света в окрестности точки, соответствующей геометрическому изображению, определяется как результат интерференции когерентного света, исходящего из каждой точки волновой поверхности волны, прошедшей сквозь линзу (или выходной зрачок идеальной оптической системы) и являющейся, согласно принципу Гюйгенса – Френеля, источником когерентного излучения. Величину r_1' можно считать мерой степени неопределенности, с которой точка в плоскости пространственных частот (в фокальной плоскости линзы) соответствует пространственной частоте аксиальной плоской

волны, падающей на линзу конечного радиуса R . Эта неопределенность является следствием того, что линза конечных размеров собирает лишь часть пространственной информации, которую несет световая волна.

В более общем случае, при наличии aberrаций (недостатков) линз и других причин, даваемое линзой изображение точечного источника (или, как говорят, отклик системы на точечное (импульсное) воздействие) будет представлять собой световое пятно с изменяющейся от точки к точке интенсивностью, которое называется функцией рассеяния точки (ФРТ) оптической системы¹⁰.

Чтобы найти связь между распределением амплитуд $E'(x', y')$ и $E(x, y)$ в случае объекта конечных размеров, обратимся снова к соотношению (11.13). Перейдем в этом соотношении к новым переменным: $\beta x = -s$, $\beta y = -t$. Тогда получим

$$h(x, y, x', y') = \\ = -\beta \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\lambda a' x_1, \lambda a' y_1) \exp \left\{ -i2\pi [(x' - s)x_1 + (y' - t)y_1] \right\} dx_1 dy_1.$$

В новых переменных функция отклика системы $h(x, y, x', y')$ зависит только от разности координат. Если ввести в рассмотрение функцию

$$\bar{h}(x' - s, y' - t) = -\frac{1}{\beta} h(x, y, x', y'),$$

то интеграл суперпозиции (11.9) с учетом замены переменных можно привести к виду

$$E'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[-\frac{1}{\beta} E \left(-\frac{s}{\beta}, -\frac{t}{\beta} \right) \right] \bar{h}(x' - s, y' - t) ds dt,$$

¹⁰ Об этом будет подробно говориться в следующей главе.

где

$$\bar{h}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\lambda a' x_1, \lambda a' y_1) \exp[-i2\pi(xx_1 + yy_1)] dx_1 dy_1. \quad (11.19)$$

Полученное выражение для $E'(x', y')$ представляет собой свертку импульсной характеристики системы с идеальным изображением. Оно и определяет связь между распределением комплексных амплитуд светового поля в плоскости предмета и плоскости изображения с учетом ограниченных размеров линзы. Как видим, учет конечных размеров апертуры линзы приводит к замене однозначного соответствия между точками предмета и изображения преобразованием свертки.

Таким образом, изображение, даваемое линзой, не является точной копией предмета. Поскольку это изображение определяется сверткой, а при свертке каждая точка входной функции, описываемой δ -функцией, расширяется на выходе в пятно (кружок Эри), то оно является несколько размытым, неконтрастным. Физической причиной этой размытости и неконтрастности является дифракция света на апертуре линзы.

Кроме дифракционных эффектов, на качество изображения существенное влияние оказывают и различные аберрации (недостатки) линз, полное устранение которых оказывается практически невозможным.

Для рассмотрения вопроса о влиянии конечных размеров апертуры линзы на распределение поля в задней фокальной плоскости линзы положим в формуле (11.11) $a' = f$. В этом случае импульсная функция примет вид

$$h(x, y, x', y') = -\frac{e^{ik(a+f)}}{\lambda^2 a f} \cdot \exp\left(ik \frac{x^2 + y^2}{2a}\right) \times \\ \times \exp\left(ik \frac{x'^2 + y'^2}{2f}\right) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\xi, \eta) \exp\left(ik \frac{\xi^2 + \eta^2}{2a}\right) \times$$

$$\times \exp \left\{ -ik \left[\left(\frac{x}{a} + \frac{x'}{f} \right) \xi + \left(\frac{y}{a} + \frac{y'}{f} \right) \eta \right] \right\} d\xi d\eta.$$

Подставим это выражение в формулу (11.9). После некоторых несложных преобразований эту формулу приведем к виду

$$E'(x', y') = \frac{1}{\lambda^2 a f} \exp \left[ik \left(1 - \frac{a}{f} \right) \frac{x'^2 + y'^2}{2f} \right] \times \\ \times \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp \left[-ik \left(\frac{x'}{f} x + \frac{y'}{f} y \right) \right] A(x, y; x', y') dx dy,$$

где

$$A(x, y; x', y') = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} P(\xi, \eta) \times \\ \times \exp \left\{ i \frac{k}{2a} \left[\left(\xi - \left(x + \frac{a}{f} x' \right) \right)^2 + \left(\eta - \left(y + \frac{a}{f} y' \right) \right)^2 \right] \right\} d\xi d\eta$$

и несущественный множитель $-\exp[ik(a+f)]$ опущен. Показатель экспоненты, стоящей под знаком интеграла в выражении $A(x, y; x', y')$, имеет вид квадратичной формы. Кроме того, апертурная функция $P(\xi, \eta)$ для типичных для оптики значений λ является медленно меняющейся, поэтому для вычисления второго интеграла можно воспользоваться методом стационарной фазы.

Согласно этому методу, если $F(\xi)$ – медленно меняющаяся функция, то

$$\int_{-\infty}^{\infty} F(\xi) \exp[if(\xi)] d\xi = \sqrt{\frac{2\pi}{\mu f''(\xi_0)}} F(\xi_0) \exp \left\{ i \left[f(\xi_0) \pm \frac{\pi}{4} \right] \right\},$$

где ξ_0 – корень уравнения $f'(\xi) = 0$. В двумерном случае для медленно меняющейся функции $F(\xi, \eta)$ будем иметь

$$\int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} F(\xi, \eta) \exp[i f(\xi, \eta)] d\xi d\eta = \frac{2\pi}{\sqrt{\mu f''_{\xi}(\xi_0, \eta_0) f''_{\eta}(\xi_0, \eta_0)}} \times \\ \times F(\xi_0, \eta_0) \exp\left\{i \left[f(\xi_0, \eta_0) \mu \frac{\pi}{2} \right]\right\},$$

где ξ_0, η_0 – корни системы уравнений $f'_{\xi}(\xi, \eta) = 0, f'_{\eta}(\xi, \eta) = 0$. Знаки \pm под корнем в знаменателях правых частей поставлены для того, чтобы обеспечить положительность подкоренных выражений. Отметим также, что под медленно меняющейся функцией понимается функция, медленно меняющаяся в интервале шириной

$$\Delta\xi = \sqrt{\frac{2\pi}{\pm f''(\xi_0)}}$$

в одномерном случае и в интервалах изменения переменных ξ, η шириной

$$\Delta\xi = \sqrt{\frac{2\pi}{\pm f''_{\xi}(\xi_0, \eta_0)}}, \Delta\eta = \sqrt{\frac{2\pi}{\pm f''_{\eta}(\xi_0, \eta_0)}}$$

соответственно в двумерном случае.

Применяя вторую формулу вычисления интеграла по методу стационарной фазы к выражению $A(x, y; x', y')$ и учтя, что в нашем случае

$$\xi_0 = x + \frac{a}{f} x', \eta_0 = y + \frac{a}{f} y', f''_{\xi}(\xi_0, \eta_0) = f''_{\eta}(\xi_0, \eta_0) = \frac{k}{a},$$

получим

$$A(x, y; x', y') = \lambda a P \left(x + \frac{a}{f} x', y + \frac{a}{f} y' \right).$$

Фазовый множитель $\exp\left(\pm i \frac{\pi}{2}\right)$ опущен как несущественный.

Ширина интервалов, в которых апертурная функция должна медленно изменяться $\Delta \xi = \Delta \eta = \sqrt{\lambda a}$. Ввиду малости λ эти величины меньше интервалов изменения переменных ξ и η , в пределах которых апертурная функция остается постоянной, равной единице, поэтому условие применимости метода стационарной фазы к вычислению выражения $A(x, y, x', y')$ выполняется. С учетом этого находим распределение комплексной амплитуды поля в задней фокальной плоскости линзы:

$$\begin{aligned} E'(x', y') &= \frac{1}{\lambda f} \exp \left[ik \left(1 - \frac{a}{f} \right) \frac{x'^2 + y'^2}{2f} \right] \times \\ &\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) P \left(x + \frac{a}{f} x'; y + \frac{a}{f} y' \right) \times \\ &\times \exp \left[-ik \left(\frac{x'}{f} x, \frac{y'}{f} y \right) \right] dx dy. \end{aligned} \quad (11.20)$$

Поскольку координаты x' и y' связаны с пространственными частотами u и v соотношениями $x' = \lambda f u$ и $y' = \lambda f v$, то формулу (11.20) можно записать в следующем виде:

$$E'(u, v) = \frac{1}{\lambda f} \exp \left[i\pi \lambda f \left(1 - \frac{a}{f} \right) (u^2 + v^2) \right] \times$$

$$\times \int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) P(x + \lambda au; y + \lambda av) \exp[-i2\pi(ux + vy)] dx dy .$$

Из полученного соотношения видно, что при учете влияния ограниченного размера зрачка линзы поле в задней фокальной плоскости линзы и поле в плоскости предмета не связаны преобразованием Фурье. Лишь при $a = 0$, когда предмет расположен вплотную к линзе и его физические размеры меньше диаметра зрачка линзы, влиянием зрачка можно пренебречь. В этом случае с точностью до квадратичного фазового множителя $\exp[i\pi\lambda f(u^2 + v^2)]$ указанные поля будут связаны преобразованием Фурье.

11.6. Влияние конечных размеров линзы на спектр пространственных частот

Рассмотрим теперь вопрос о влиянии конечных размеров линзы, перейдя в область пространственных частот. Разложим функцию распределения комплексных амплитуд объекта $E(x, y)$ в интеграл Фурье, т.е. представим входной сигнал в виде совокупности плоских волн (гармоник) всевозможных пространственных частот u и v и распространяющихся под различными углами к направлению распространения освещающей объект плоской волны. Точно так же представим в виде суперпозиции пространственных гармоник всех возможных пространственных частот u и v и распределение комплексных амплитуд поля $E'(x', y')$ в плоскости изображения. Взяв преобразование Фурье от (11.9) и используя теорему о свертке, для фурье-спектра распределения комплексных амплитуд светового поля в плоскости изображения $E'(u, v)$ получим

$$E'(u, v) = H(u, v) E(u, v), \quad (11.21)$$

где $E(u, v)$ – фурье-спектр функции $E(x, y)$,

$$H(u, v) = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} \bar{h}(x, y) e^{-i2\pi(ux+vy)} dx dy$$

– передаточная функция оптической системы. Отсюда следует, что изображение, построенное линейной пространственно-инвариантной (изопланарной) системой, получается суммированием составляющих (по частоте) гармоник объекта, для каждой из которых действие изображающей системы учитывается своим коэффициентом $H(u, v)$ по формуле (11.21). Передаточная функция определяет, каким образом каждая частотная составляющая спектра объекта передается оптической системой с учетом дифракции и aberrаций. Учитывая, что, согласно (11.19), $\bar{h}(x, y)$ есть преобразование Фурье функции зрачка, получим, что передаточная функция есть двойное преобразование Фурье зрачковой функции системы:

$$H(u, v) = F\{F\{\bar{h}(x, y)\}\}.$$

Откуда

$$\begin{aligned} H(u, v) &= \\ &= \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} \left(\int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} P(\lambda a' \zeta, \lambda a' \eta) e^{-i2\pi(x\zeta+y\eta)} d\zeta d\eta \right) e^{-i2\pi(ux+vy)} dx dy = \\ &= \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} P(\lambda a' \zeta, \lambda a' \eta) \left(\int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} e^{-i2\pi[(u+\zeta)x+(v+\eta)y]} dx dy \right) d\zeta d\eta = \\ &= \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} P(\lambda a' \zeta, \lambda a' \eta) \delta(u+\zeta, v+\eta) d\zeta d\eta = P(-\lambda a' u, -\lambda a' v). \end{aligned}$$

Перейдя в плоскости зрачка к отраженной системе координат, т.е. произведя замену $x \rightarrow -x$, $y \rightarrow -y$, будем иметь

$$H(u, v) = P(\lambda a' u, \lambda a' v).$$

Эта формула справедлива для дифракционно-ограниченных оптических систем со зрачком любой формы. Величина a' в этом соотношении в общем случае представляет собой расстояние от плоскости зрачка до плоскости изображения. Введем новое обозначение этого расстояния: l . В этих обозначениях

$$H(u, v) = P(\lambda l u, \lambda l v). \quad (11.22)$$

Из выражения (11.22) следует, что передаточная функция $H(u, v)$ при когерентном освещении равна значению функции зрачка при $x = \lambda l u$, $y = \lambda l v$, поэтому для получения передаточной функции $H(u, v)$ нужно в зрачковой функции $P(x, y)$ просто заменить переменные x и y соответственно на $\lambda l u$ и $\lambda l v$. Например, для прямоугольной апертуры размером $a \times b$ зрачковая функция имеет вид

$$P(x, y) = \text{rect}\left(\frac{x}{a}\right) \text{rect}\left(\frac{y}{b}\right) = 1$$

при $|x| \leq a/2$, $|y| \leq b/2$;

$$P(x, y) = 0$$

вне указанного прямоугольника.

Эта функция показана на рис. 11.2, а. Произведя в этом выражении указанную выше замену, получим соответствующую передаточную функцию

$$H(u, v) = \text{rect}\left(\frac{\lambda l u}{a}\right) \text{rect}\left(\frac{\lambda l v}{b}\right) = 1$$

при $|\lambda l u| \leq a/2$, $|\lambda l v| \leq b/2$;

$$H(u, v) = 0$$

в остальных случаях.

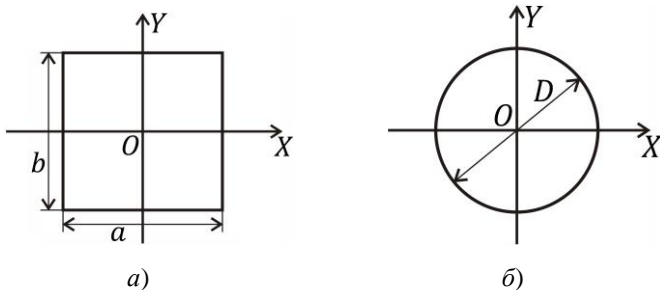


Рис.11.2

Аналогично можно получить передаточную функцию и для круговой апертуры радиуса R . Зрачковая функция круговой апертуры определяется выражением

$$P(x, y) = \text{circ}(\sqrt{x^2 + y^2} / R) = 1$$

при x, y , заключенных внутри круга $x^2 + y^2 \leq R^2$ и $P(x, y) = 0$ вне указанного круга (рис. 11.2, б), а ее передаточная функция

$$H(u, v) = \text{circ}\left(\frac{\sqrt{u^2 + v^2}}{R / \lambda l}\right) = 1$$

при выполнении условия $u^2 + v^2 \leq \left(\frac{R}{\lambda l}\right)^2$;

$$H(u, v) = 0$$

в остальных случаях.

Зрачковая функция круговой апертуры показана на рис. 11.2, б. Введя обозначение $\rho = \sqrt{u^2 + v^2}$, выражение для передаточной функции зрачка круговой формы можно записать в виде

$$H(\rho) = \text{circ}\left(\frac{\lambda l \rho}{R}\right).$$

Графики указанных передаточных функций показаны на рис. 11.3, *a* и 11.3, *б* соответственно

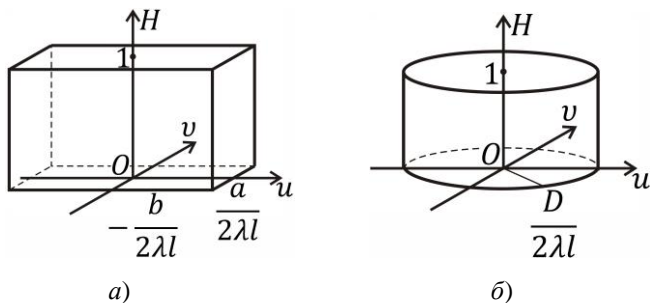


Рис. 11.3

Соотношение (11.22) показывает, что часть гармоник входного сигнала пропускается оптической системой без искажений ($P(\lambda l u, \lambda l v) = 1$), другая же часть, начиная с некоторых пространственных частот u и v , не пропускается совсем ($P(\lambda l u, \lambda l v) = 0$). В случае круговой апертуры будут пропускаться лишь те гармоники, пространственные частоты которых удовлетворяют условию

$$u^2 + v^2 \leq (R/\lambda l)^2, \quad (11.23)$$

т.е. внутри круга радиуса $R/\lambda l$. Гармоники, частоты которых расположены за пределами этого круга, отфильтровываются оптической системой. Предельные частоты гармоник, пропущенных системой, $u_0 = v_0 = R/\lambda l$ определяются, таким образом, угловым радиусом зрачка R/l и длиной волны излучения λ . Аналогично для зрачка прямоугольной формы пропускаются гармоники с пространственными частотами

$$u \leq a/(2\lambda l), v \leq b/(2\lambda l) \quad (11.24)$$

и отфильтровываются оптической системой гармоники с пространственными частотами, превышающими предельные значения

$$u_0 = a / (2\lambda l), \quad v_0 = b / (2\lambda l). \quad (11.25)$$

Мы видим, таким образом, что влияние оптической системы на пространственный спектр изображения равносильно действию линейного фильтра пространственных частот.

Изображение предмета было бы идеальным, если бы линза (вообще изображающая оптическая система) собирала все световые волны, несущие информацию о предмете. Соотношение (11.23) и аналогичные соотношения (11.24) для прямоугольного зрачка определяют условия, при которых линза преобразует в изображение всю световую информацию о предмете, которую несут гармоники с пространственными частотами u и v . Пространственные частоты компонент, не удовлетворяющие указанным выше условиям, не проходят через линзу и не участвуют в формировании изображения предмета. Несущая такими гармониками информация о предмете теряется, что и приводит к потере качества изображения. Наибольшей фильтрации подвергаются гармоники высоких пространственных частот, которые несут информацию о мелких деталях объекта. Поэтому мелкие детали предмета будут ослаблены в изображении, вследствие чего изображение предмета получается размытым, неконтрастным.

Проведенное рассмотрение показывает, что переход от предмета к изображению эквивалентен действию линейного фильтра пространственных частот. Этот факт является следствием пространственной ограниченности системы (при $R \rightarrow \infty$ или при $a \rightarrow \infty$, $b \rightarrow \infty$ граничные пространственные частоты u_0 и v_0 также стремятся к бесконечности), и более конкретно — следствием дифракции света на входном зрачке системы.

11.7. Формирование оптического изображения по Аббе – Портеру

Существуют две модели формирования оптического изображения, каждая из которых наиболее удачно применяется в совершенно определенных условиях. Согласно первой модели (модели Аббе – Портера), образование изображения при когерентном освещении сводится к образованию дифракционной картины в фокальной плоскости линзы и последующему суммированию в плоскости изображения интерференционных полос, создаваемых лучами, проходящими через пары дифракционных максимумов $\pm m$ -го порядка, но с непрерывным диапазоном разнесения полос и их ориентацией. Или к двойному преобразованию Фурье (так как образованию фраунгоферовской дифракционной картины соответствует преобразование Фурье от функции амплитудного распределения по объекту, а суммированию интерференционных полос – суммирование фурье-компонент). Согласно второй модели (модели Рэлея), применяемой для описания формирования изображения, как при когерентном, так и при некогерентном освещении, изображение протяженного объекта можно представить как комбинацию элементарных изображений, которые оптическая система создает отдельно для света от каждой точки объекта. Если освещение когерентно, то имеет место интерференция и тогда изображение математически можно представить в виде комбинации элементарных изображений с комплексными амплитудами. В этом случае при представлении объекта в виде последовательности равно разнесенных точечных источников для описания суммирования регулярно расположенных элементарных изображений, формирующих изображение объекта, можно использовать ряды Фурье. При этом члены рядов соответствуют порядкам дифракции модели Аббе – Портера. Если же освещение не когерентно, то в плоскости изображения будут суммироваться не амплитуды, а интенсивности элементарных изображений, полученные от каждой точки объекта. Рассмотрим здесь первую модель подробно. Вторая модель будет описана в п. 12.4.

Предположим сначала, что объектом является дифракционная решетка PP , освещаемая параллельным пучком монохроматических лучей. Изображение решетки создается в плоскости $P'P'$ (рис. 11.4). Вследствие дифракции на решетке свет, прошедший через решетку, будет состоять из дискретного ряда плоских дифрагировавших волн, распространяющихся от решетки по различным направлениям. В фокальной плоскости F линзы L возникает фраунгоферова дифракционная картина в виде набора ярких точек – дифракционных максимумов различных порядков. В точке пересечения этой плоскости с оптической осью линзы находится максимум нулевого порядка M_0 ($m = 0$), по обе стороны от которого располагаются максимумы первого M_1 и M_{-1} ($m = \pm 1$), второго M_2 и M_{-2} ($m = \pm 2$), третьего M_3 и M_{-3} ($m = \pm 3$) и т.д. порядков согласно формуле $d \sin \varphi_m = m\lambda$, где d – период решетки. Интенсивность этих максимумов убывает по мере того, как увеличивается их порядок.

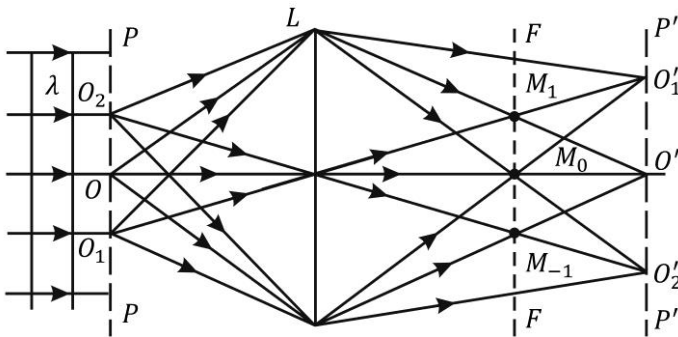


Рис. 11.4

Совокупность точек $M_0, M_1, M_{-1}, M_2, M_{-2}$ и т.д. Э. Аббе назвал первичным или дифракционным изображением. Эти точки – дифракционные максимумы – являются источниками вторичных когерентных волн Гюйгенса, которые, распространяясь далее за

фокальной плоскостью FF , встречаются в сопряженной с P плоскости изображения PP' . Интерферируя между собой, эти вторичные волны в плоскости PP' создают дифракционную картину, называемую вторичным изображением Аббе, которая и представляет собой более или менее точное изображение решетки. При этом линза выполняет задачу сведения дифрагировавших на решетке световых пучков в плоскости изображения. Процесс распространения света от фурье-плоскости до плоскости изображения Аббе назвал второй дифракцией. Таким образом, формирование оптического изображения решетки происходит в две стадии: на первой (предварительной) стадии происходит дифракция света на решетке с образованием фраунгоферовской дифракционной картины в фокальной плоскости линзы, а на второй (окончательной) – рекомбинация дифрагированного света (отклоненных пучков) линзой в реальное изображение. Отметим, что если на рис.11.4 выделить два луча, идущих, например, из точки O_1 в точку O_1' , один проходящий через центр линзы, другой – через ее нижний край, то мы увидим, что здесь нет противоречия между этим волновым описанием и методом лучей геометрической оптики.

Взятая сама по себе любая пара максимумов M_m и M_{-m} создает интерференционную картину на плоскости изображения. Это напоминает интерференционный опыт Юнга, в котором пара узких отверстий в непрозрачном экране действует таким же образом. Интерференционную картину в опыте Юнга часто называют дифракционной картиной, имея в виду физический процесс (дифракцию), при котором свет проходит через каждое отверстие, чтобы попасть к месту интерференции. В этом смысле формирование изображения можно рассматривать как двойной процесс дифракции (идея, принадлежащая Цернике).

Чтобы уяснить, каким способом строится изображение объекта типа решетки, показанной на рис. 11.4, рассмотрим вклады в изображение, вносимые парами дифрагировавших лучей, включающих различные порядки дифракции, т.е. вклады, которые

вносят пары максимумов $M_{\pm m}$, $m = 1, 2, 3, \dots$. На рис. 11.5 показана схема образования пары дифрагировавших лучей m -го порядка, испускаемых решеткой. В плоскости изображения PP' они интерферируют, создавая распределение интенсивности по гармоническому закону с периодом d'_m , равным ширине интерференционной полосы h .

Рассматривая дифракционные максимумы $M_{\pm m}$ как когерентные источники, для определения ширины полосы h воспользуемся формулой (3.25) $h = \lambda l / a$, где a – расстояние между источниками M_m и M_{-m} , l – расстояние от источников M_{+m} и M_{-m} до плоскости PP' , в которой наблюдается интерференционная картина. С помощью рис. 11.5 находим $a = l \operatorname{tg} \varphi'$, где φ' – угол схождения интерферирующих лучей, т.е. угол, под которым из точки O'_1 виден отрезок длиной a , соединяющий источники M_m и M_{-m} , а так как φ' – мал, то можно считать, что $a \approx l \sin \varphi'$. Следовательно,

$$d'_m = h \approx \lambda / \sin \varphi'.$$

С другой стороны, условием для формирования максимумов m -го порядка от дифракционной решетки является выражение $d \sin \varphi = m \lambda$. С учетом этого получаем связь между периодом d'_m гармонического распределения интенсивности, создаваемым парой максимумов m -го порядка, и периодом решетки d :

$$d'_m \approx (d / m) \cdot (\sin \varphi / \sin \varphi').$$

Множитель во вторых скобках является постоянным по величине и связан с увеличением линзы, поэтому имеем: $d'_m \sim d / m$.

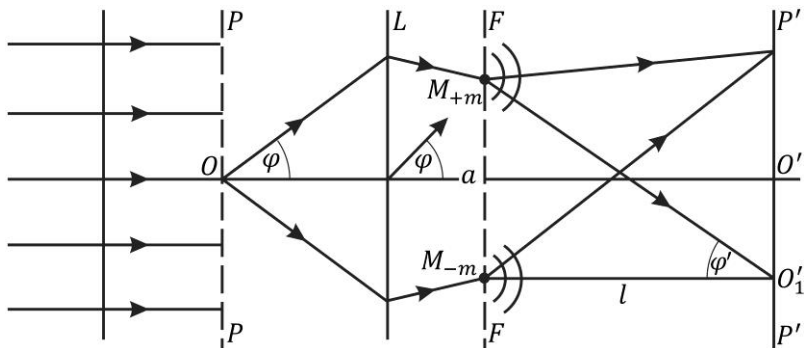


Рис.11.5

Волны, проходящие через максимумы первого порядка, интерферируют в плоскости изображения PP' , создавая в этой плоскости гармонические осцилляции интенсивности, которые соответствуют основному периоду решетки ($d'_m \sim d$). Этот период дает минимальную информацию об изображаемом объекте (решетке) без тонких деталей его оптической структуры. Каждая пара последующих максимумов более высокого порядка добавляет последовательно к общему распределению интенсивности гармоники (колебания интенсивности) более короткого периода ($d'_m \sim \sim d/m$), которые формируют изображение.

Аббе продемонстрировал это, перекрывая с помощью различных экранов и диафрагм часть главных дифракционных максимумов $M_{\pm m}$ с помощью диафрагмы, помещенной в фокальную плоскость линзы. При таком перекрытии изображение решетки в плоскости PP' получается искаженным, так как оно соответствует решетке, у которой при дифракции эти максимумы не возникают. Когда диафрагма пропускает только один пучок от M_0 , достигающий плоскости PP' , свет будет состоять только из одной волны, которой не с чем интерферировать. Изображение получается бесструктурным и представляет собой более или менее равномерно освещенное поле, какое наблюдалось бы, если

бы на месте решетки PP находилась прозрачная плоскопараллельная пластинка. Когда диафрагма в плоскости FF открывает кроме главного максимума M_0 еще максимумы первого порядка M_1 и M_{-1} , исходящие из них волны интерферируют в плоскости PP' . Возникает изображение решетки с плавным переходом от прозрачных к непрозрачным участкам. Точно такая же картина наблюдалась бы в случае, когда на месте щелевой решетки PP находилась бы решетка того же периода d , но с гармонической функцией пропускания. Действительно, при дифракции на такой решетке возникают главные максимумы только порядков $m = 0, \pm 1$, т.е. максимумы M_0, M_1 и M_{-1} , а это и имеет место в рассматриваемом случае. При увеличении числа открываемых максимумов изображение решетки будет все больше и больше становиться подобным самой решетке. Однако точное подобие достигнуто не будет, так как из-за конечных размеров линзы не все лучи, проходящие через максимумы, попадают в линзу и, кроме того, само число дифракционных максимумов, создаваемых решеткой, ограничено: $m_{\max} = \text{целое число} \leq d / \lambda$.

Поскольку сходство изображения решетки с самой решеткой (когда в изображении решетки оказывается разрешенным основной период d решетки) появляется при $m = \pm 1$, то наименьший период (размер) объекта, который может быть разрешен при заданной апертуре линзы $2\sigma_A$, можно определить из условия $d \sin \sigma_A = 1 \cdot \lambda$, откуда находим, что $d = \lambda / \sin \sigma_A$.

Закроем все максимумы, кроме центрального максимума M_0 и двух боковых максимумов M_2 и M_{-2} . Тогда опять получим случай гармонической решетки, с той лишь разницей, что максимумы $M_{\pm 2}$ будут более удалены друг от друга, чем максимумы $M_{\pm 1}$. Если оставить открытыми только M_0 и два максимума $M_{\pm 3}$, то период решетки, изображаемой в плоскости PP' , станет еще меньше (в три раза) и т.д. Все происходит так,

как если бы решетка-объект PP представляла собой суперпозицию бесконечного числа гармонических решеток с разными периодами $d, d/2, d/3$ и т.д. Математически это можно выразить, представив функцию пропускания щелевой решетки $t(x)$ в виде ряда Фурье (п. 7.3.2):

$$t(x) = \sum_{m=-\infty}^{\infty} \frac{A_m}{2} \cos \frac{2\pi m}{d} x,$$

где

$$A_m/2 = \frac{a}{d} \sin \frac{\pi ma}{d} \Big/ \frac{\pi ma}{d},$$

a – ширина щели. Каждый член этого ряда описывает одну гармоническую решетку. Каждая из таких решеток дает в плоскости FF две освещенные точки справа и слева от M_0 , соответствующие главным максимумам ± 1 -го порядка, полученным от этой решетки. Пары главных дифракционных максимумов порядка $\pm m$ обусловлены соответствующими парами гармоник (фурье-компонентами, т.е. членами ряда $t_m(x) = \frac{A_m}{2} \cos \frac{2\pi m}{d} x$) с амплитудами $M_{\pm m}/2$, пространственными частотами m/d и пространственными периодами d/m . Член $t_0 = A_0/2$ соответствует равномерному распределению света в плоскости решетки, и можно считать, что он принадлежит гармонической решетке с бесконечно большим периодом.

Все детали изображения решетки строятся способом, вполне аналогичным фурье-синтезу (сложению фурье-компонент). Но дифракционные максимумы сами заключают в себе разложение Фурье (фурье-синтез) решеточного объекта (точнее, его функции пропускания), а дифракционная плоскость описывается как фурье-плоскость. Поэтому процесс формирования изображения в рассматриваемом примере можно интерпретировать как двойное преобразование Фурье. Первое преобразование – это разложение

светового излучения в фурье-спектр в фокальной плоскости линзы, т.е. образование фраунгоферовской дифракционной картины, а второе – суммирование этих фурье-компонент (фурье-синтез данного фурье-анализа) в плоскости изображения, т.е. образование собственно изображения. Такую интерпретацию легко понять, если воспользоваться принципом обратимости. Все порядки дифракции, которые создают изображение путем суммирования гармоник, возвращают к решеточному объекту, где они рекомбинируют (на второй стадии формирования изображения), образуя первоначальное распределение интенсивности на решетке.

Таким образом, согласно теории Аббе – Портера, изображение получается двумя последовательными этапами: дифракцией на предмете и образованием фраунгоферовой дифракционной картины в фокальной плоскости линзы и последующим преобразованием дифрагировавших пучков линзой в реальное оптическое изображение в сопряженной плоскости. Первому этапу соответствует фурье-анализ, а второму – фурье-синтез.

Представление о двойном преобразовании Фурье полностью согласуется с интерпретацией двойной дифракции Цернике. По существу это два различных способа выражения одного и того же явления.

Рассмотренный выше подход можно применить и к любым непериодическим структурам, так как дискретные порядки дифракции не являются его необходимыми условиями. В основе проведенного рассмотрения лежит тот факт, что изображение в плоскости PP' решеточного объекта, расположенного в плоскости PP , формируется только световыми пучками, проходящими через плоскость FF , т.е. пучками, образующими дифракционные спектры. Непериодический объект можно рассматривать как одну щель решетки, но не с прямоугольной, а с произвольной функцией пропускания $t(x, y)$. Как и в случае щели, функция пропускания такого объекта может быть представлена не рядом, а интегралом Фурье. Дифракционная картина в фокальной плоскости линзы (дифракционной плоскости) в этом случае будет

представлять собой картину непрерывного рассеяния волн с угловым изменением амплитуды и фазы, зависящим от функции пропускания объекта. При этом крупные детали объекта, которым соответствуют низкие пространственные частоты, дают дифрагировавшие лучи, которые мало отклоняются от направления прямого пучка. Мелкие детали, которым соответствуют высокие частоты, дают дифрагировавшие лучи, сильно отклоненные от прямого пучка. Эта картина является двумерным преобразованием Фурье от функции распределения комплексной амплитуды поля по объекту $E(x, y)$. Восстановление этой картины в плоскости изображения сводится к суммированию интерференционных полос, создаваемых парами дифрагировавших лучей (под углами $\pm \varphi'$ на рис 11.5), но с произвольным расстоянием между полосами и их ориентацией.

Формирование изображения и в случае непериодических объектов можно описать как процесс двойного преобразования Фурье. При освещении транспаранта параллельным пучком монохроматического света проходящая через транспарант плоская монохроматическая волна дифрагирует на неоднородностях прозрачности изображенного на транспаранте предмета (модулируется по амплитуде и фазе в плоскости предмета) так, что в плоскости XU непосредственно за транспарантом возникает распределение комплексной амплитуды $E(x, y) = E_0 t(x, y)$, где E_0 – амплитуда падающей на транспарант плоской волны, $t(x, y)$ – функция пропускания транспаранта. Поле световой волны в плоскости XU можно разложить в двумерный интеграл Фурье, т.е. представить в виде совокупности бесконечного числа плоских дифрагировавших волн с пространственными частотами u и v и соответствующими амплитудами и фазами. Комплексная амплитуда дифрагировавшей волны с волновым вектором $\mathbf{k}(2\pi u, 2\pi v)$ пропорциональна соответствующей фурье-компоненте (пространственной гармонике) $E(u, v)$ функции $E(x, y)$ предмета:

$$E(u, v) = \int_{-\infty-\infty}^{\infty\infty} E(x, y) e^{-i2\pi(ux+vy)} dx dy .$$

(11.26)

Каждый пучок параллельных лучей, получившийся в результате дифракции на прозрачной пластине, собирается линзой в соответствующей точке фокальной плоскости. Возникает дифракционная картина, в которой, как было показано в п. 10.5, распределение амплитуды поля $E_f(x_f, y_f)$ представляет собой двумерный фурье-образ $E(u, v)$ функции $E(x, y)$ предмета.

Проходящие через расположенные в фокальной плоскости линзы точки максимумов дифрагировавшие волны в плоскости PP' изображения интерферируют между собой. В линзе с исправленными aberrациями оптические пути между всеми парами сопряженных точек одинаковы, поэтому при интерференции волны в точках (x', y') плоскости PP' их сложение происходит с теми же относительными фазами, которые они имели в соответствующих точках (x, y) плоскости PP предмета. Распределение амплитуды светового поля $E'(x', y')$ в плоскости изображения PP' , будет описываться (с точностью до масштаба) таким же интегралом Фурье, что и в плоскости PP предмета:

$$E'(x', y') = \int_{-\infty-\infty}^{\infty\infty} E(u, v) e^{i2\pi(ux'+vy')} du dv .$$

(11.27)

Отметим, что первое преобразование Фурье (прямое) описывает первую дифракцию – дифракцию на апертуре линзы, а второе (обратное) – вторую дифракцию, которая осуществляется в результате распространения света в свободном пространстве между задней фокальной плоскостью линзы и плоскостью изображения.

Хотя пределы интегрирования по u и v , т.е. по направлениям дифрагировавших волн (так как каждому значению u и v соответствует вполне определенное направление распространения дифрагировавшей волны), в интеграле (11.27) указаны равными $\pm\infty$, на самом деле интегрирование производится в пределах, определенных апертурой линзы.

В плоскости изображения PP' в результате интерференции дифрагировавших волн возникает распределение светового поля, воспроизводящее (в измененном масштабе) распределение поля $E(x, y)$ в плоскости PP предмета. Однако это воспроизведение не будет точным, поэтому восстановленное изображение не будет полностью идентично изображению предмета на транспаранте. Хотя распределение амплитуды светового поля $E'(x', y')$ в плоскости изображения описывается (с точностью до масштаба) тем же интегралом Фурье (11.27), что и в плоскости предмета, но в этот интеграл входят не все фурье-компоненты $E(u, v)$ функции предмета $E(x, y)$, а только те, которым соответствуют прошедшие через линзу дифрагировавшие волны. Угол дифракционной расходимости φ дифрагировавшего пучка определяется соотношением $\sin\varphi \sim \lambda / a$, где a – линейный размер объекта. Поэтому если объект очень мал ($a \approx \lambda$), то дифрагировавшие на нем волны будут идти под большим углом к оптической оси и могут не попасть в линзу. По этой причине пространственные гармоники $E(u, v)$, соответствующие мелким деталям предмета, будут отсутствовать в формируемом линзой изображении, если соответствующие им дифрагировавшие волны имеют угол дифракции $\varphi > \sigma_A$, где $2\sigma_A$ – апертура линзы. Это и приводит к тому, что распределение амплитуды светового поля $E'(x', y')$ в плоскости изображения не дает точного воспроизведения функции $E(x, y)$ предмета. Линза из-за конечных размеров апертуры не способна отобразить объект, размеры которого соизмеримы с длиной волны света. Вследствие этого и происходит потеря точности восстановления изображения.

Аналогично и в случае решеточного объекта. Изображение решетки получилось бы полностью подобным самой решетке со всеми ее деталями, если бы распределение света в плоскости изображения PP' представлялось бы рядом Фурье с теми же коэффициентами, что и в плоскости предмета. Однако это оказывается невозможным. Из ряда Фурье выпадают компоненты, соответствующие волнам, которые не попадают в линзу из-за конечных размеров ее апертуры. Изображение соответствует суперпозиции только тех гармонических решеток, главные дифракционные максимумы которых вошли в оптическую систему.

Проиллюстрируем представление о двойной дифракции на примере когерентно освещенного одномерного периодического решеточного объекта с гармонической функцией пропускания

$$t(x) = \frac{1}{2} \left(1 + \cos \frac{2\pi}{d} x \right),$$

где d – период решетки. Распределение амплитуды поля по объекту

$$E(x) = E_0 t(x) = \frac{1}{2} E_0 \left(1 + \cos \frac{2\pi}{d} x \right),$$

где E_0 – амплитуда плоской волны, падающей на объект. При первой дифракции в фокальной плоскости линзы образуется дифракционная картина (спектр)

$$\begin{aligned} E(u) &= \int_{-\infty}^{\infty} E(x) e^{-i2\pi ux} dx = \\ &= \frac{1}{2} E_0 \left[\int_{-\infty}^{\infty} \exp(-i2\pi ux) dx + \frac{1}{2} \int_{-\infty}^{\infty} \exp[-i2\pi(u - 1/d)x] dx + \right. \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{2} \int_{-\infty}^{\infty} \exp[-i2\pi(u + 1/d)] dx \Big] = \frac{1}{2} E_0 \left[\delta(u) + \frac{1}{2} \delta(u - 1/d) + \right. \\
& \left. + \frac{1}{2} \delta(u + 1/d) \right]. \tag{11.28}
\end{aligned}$$

При получении этого выражения поле $E(x)$ было представлено в виде

$$E(x) = \frac{1}{2} E_0 \left[1 + \frac{1}{2} \exp\left(i \frac{2\pi}{d} x\right) + \frac{1}{2} \exp\left(-i \frac{2\pi}{d} x\right) \right].$$

В данном случае дифракционная картина состоит из трех разделенных плоских волн, которые условно изображаются тремя пиками на спектре – спектральными линиями. Фурье-координаты этих пиков $u = 0, \pm \frac{1}{d}$. Декартовы координаты этих пиков в фокальной плоскости линзы определяются значениями (11.28). Если эти спектральные линии попадают в апертуру линзы, то, согласно представлению о двойной дифракции, поле в плоскости изображения (с точностью до масштаба) будет

$$\begin{aligned}
E'(x') &= \int_{-\infty}^{\infty} E(u) e^{i2\pi u x'} du = \\
&= \frac{1}{2} E_0 \left[\int_{-\infty}^{\infty} \delta(u) e^{i2\pi u x'} du + \frac{1}{2} \int_{-\infty}^{\infty} \delta(u - 1/d) e^{i2\pi u x'} du + \right. \\
&+ \left. \frac{1}{2} \int_{-\infty}^{\infty} \delta(u + 1/d) e^{i2\pi u x'} du \right] = \\
&= \frac{1}{2} E_0 \left[1 + \frac{1}{2} \exp\left(i \frac{2\pi}{d} x\right) + \frac{1}{2} \exp\left(-i \frac{2\pi}{d} x\right) \right].
\end{aligned}$$

Это выражение показывает, что амплитуда изображения имеет ту же периодичность, что и амплитуда предмета. С точностью до масштаба изображение объекта оказывается полностью идентичным самому объекту.

Найдем максимальное значение пространственной частоты, пропускаемой линзой. Представим себе, что функция пропускания $t(x, y)$ объекта разложена на фурье-компоненты, описывающие пропускание гармонических решеток. Рассматривая распространение световых волн в плоскости XZ , возьмем одну из фурье-компонент (одну гармоническую решетку), составляющих пропускание $t(x, y)$, с пространственной частотой u . Если плоская волна падает на решетку неограниченных размеров, то дифрагировавшие волны будут плоскими, а углы дифракции будут определяться формулой $\sin \varphi = \pm \lambda / d = \pm \lambda u$. Если пренебречь дифракцией на краях транспаранта (решетки), то можно считать, что любой малый пучок лучей, падающий на произвольно малый участок решетки, дифрагирует под углом φ . Рассмотрим для простоты пучок лучей с вершиной, расположенной на оси Z . Максимальный угол, под которым эти лучи могут дифрагировать, попадая при этом в линзу радиусом R , определяется отношением $\operatorname{tg} \varphi = R / a$, где a – расстояние от транспаранта до линзы. Будем считать, что угол дифракции φ мал, так что $R / a = \operatorname{tg} \varphi \approx \sin \varphi$. Тогда для центральных лучей получим, что максимальная пространственная частота u_0 входной решетки, которую линза может преобразовать в изображение, определяется соотношением

$$\sin \varphi = \lambda u_0 \leq R / d,$$

т.е. не превосходит величины

$$u_0 = R / \lambda a.$$

(11.29)

Таким образом, линза конечного радиуса R преобразует в изображение всю световую информацию о предмете, которую несут компоненты с пространственными частотами вплоть до частоты u_0 . Если пространственная частота какой-либо компоненты пропускания превышает значение u_0 , то соответствующая ей информация теряется.

Мы видим, что оптическое изображение можно рассматривать как передачу информации. Отображаемый предмет осуществляет пространственную модуляцию световой волны, вызывая появление дифрагировавших волн. Эти отклоненные на разные углы волны несут информацию о структуре предмета. При этом информация о структуре предмета в получаемом изображении будет полной, если все дифрагировавшие волны проходят через линзу и участвуют в создании дифракционной картины на экране (участвуют в формировании изображения) и неполной – если какие-либо волны не попадут в линзу. Чем больше дифрагировавших волн различных порядков проходит через линзу, тем совершеннее получается изображение, тем, следовательно, больше информации о структуре предмета оно несет.

Теория формирования изображения Аббе – Портера послужила основой создания метода фильтрации пространственных гармоник, который применяется при оптической обработке. Эта теория сыграла важную роль в разработке теоретических основ фазового контраста и его разновидностей. Метод фазового контраста применяется в микроскопии и в настоящее время. Теория Аббе – Портера сыграла роль и при создании нового метода получения высококачественного изображения, называемого голографией. Все эти вопросы будут рассмотрены нами в 12 и 13 главах.

получается изображение, тем, следовательно, больше информации о структуре предмета оно несет.

Теория формирования изображения Аббе – Портера послужила основой создания метода фильтрации пространственных гармоник, который применяется при оптической обработке. Эта теория сыграла важную роль в разработке теоретических основ

фазового контраста и его разновидностей. Метод фазового контраста применяется в микроскопии и в настоящее время. Теория Аббе – Портера сыграла роль и при создании нового метода получения высококачественного изображения, называемого голографией. Все эти вопросы будут рассмотрены нами в 12 и 13 главах.

Глава 12

ФОРМИРОВАНИЕ ОПТИЧЕСКОГО ИЗОБРАЖЕНИЯ ПРИ НЕКОГЕРЕНТНОМ И ЧАСТИЧНО КОГЕРЕНТНОМ ОСВЕЩЕНИИ

12.1. Разрешающая способность оптических систем

Как уже отмечалось, изображение светящейся точки, даваемое линзой конечного размера, вследствие дифракции света на оправе линзы представляет собой не точку, а целую дифракционную картину в виде светового кружка (кружка Эри), окруженного чередующимися темными и светлыми концентрическими кольцами.

Представим себе теперь две близко расположенные друг к другу точки S_1 и S_2 какого-либо протяженного предмета. Если эти точки являются некогерентными источниками, то испускаемые ими световые волны не интерферируют и в плоскости изображения происходит сложение не амплитуд волн, а их интенсивностей, поэтому изображение точек представляет собой простое наложение дифракционных картин от отдельных точек. Будем считать, что изображаемый предмет находится на большом удалении от линзы. В этом случае лучи, идущие в линзу от каждой точки S_1 и S_2 , можно считать параллельными, а дифракцию на линзе – фраунгоферовской. Если расстояние между центрами их изображений мало по сравнению с размерами кружков Эри, то результирующая картина практически не будет отличаться от картины, представляющей изображение одной точки; изображения точек наложатся друг на друга и дадут одно неравномерно освещенное световое пятно. Следовательно, две очень близкие точки будут изображаться даже идеальной линзой не как две отдельные точки, а как одна точка. В таком случае говорят, что линза не разрешает рассматриваемые точки. Как бы ни было велико по размерам изображение предмета, на нем не

будут видны соответствующие детали; изображение предмета будет не резким, а размытым.

Если увеличивать расстояние между изображаемыми точками, то расстояние между центрами их изображений также будет увеличиваться при неизменном угловом размере соответствующих им кружков Эри. Начиная с некоторого расстояния между точками, на графике суммарного распределения интенсивности посередине появится провал, т.е. разделяющая дифракционные пятна более темная полоса. Интенсивность света в провале будет отличаться от максимального значения на относительную величину

$$\Delta I = (I_{\max} - I_{\min}) / I_{\max}.$$

Если величина этого провала достигнет 20% от максимальной интенсивности, то такая картина будет восприниматься как раздельное изображение двух точек. В этом случае говорят, что линза разрешает точки S_1 и S_2 . Вместо глубины провала часто пользуются контрастом изображения, под которым понимают отношение разности максимального и минимального значений интенсивности к их сумме: $K = (I_{\max} - I_{\min}) / (I_{\max} + I_{\min})$. При этом двадцатипроцентному провалу, как легко убедиться, соответствует одиннадцатипроцентный контраст.

Все сказанное в полной мере относится и к системе линз, заключенных в оправу, т.е. к объективу. Возможность объектива

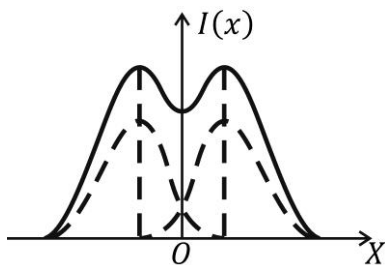


Рис. 12.1

строить раздельно изображение двух близко расположенных точек объекта называют его разрешающей способностью. Количественный критерий разрешающей способности объектива был предложен Рэлеем. Согласно Рэлею, изображения двух точечных некогерентных источников

считаются разрешенными (находятся на пределе разрешения), если центральный максимум интенсивности дифракционной картины изображения одного из них совпадет с первым дифракционным минимумом другого (рис. 12.1). В этом случае провал на кривой суммарного распределения интенсивности даже превышает указанное выше значение 20%. Действительно, радиальное распределение интенсивности в изображении точки описывается функцией

$$I(z) = I(0)[2J_1(z)/z]^2,$$

где $z = 2\pi R\varphi / \lambda$. Первый нуль этой функции имеет место при $z_1 = 3,93$. При выполнении критерия Рэля кривые, описывающие центральные максимумы двух дифракционных картин, пересекаются в точке $z = z_1 / 2 = 1,965$. Значение функции $I(z)$ в этой точке равно $0,37I(0)$, поэтому интенсивность в провале составляет $2 \cdot 37\% = 74\%$ от максимальной интенсивности. Следовательно, величина провала будет составлять 26% от максимальной интенсивности.

Обозначим через $\delta\psi$ (рис. 12.2) наименьшее угловое расстояние между двумя точками (стрелками на рис. 12.2 по-

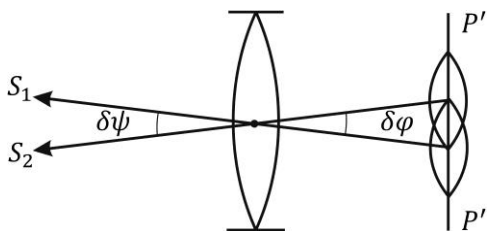


Рис. 12.2

казаны направления к этим точкам), при котором они еще разрешаются объективом. Эту величину называют угловым разрешением. Так как aberrации, вносимые объективом,

не учитываются, а учитывается только дифракция света на оправе объектива, то указанную величину $\delta\psi$ называют также дифракционным пределом разрешения. Величина R , обратная $\delta\psi$, называется разрешающей силой объектива: $R = 1/\delta\psi$.

Найдем эту величину. Из рис. 12.2 видно, что критерий Рэлея будет выполняться, если угловое расстояние $\delta\psi$ между точками S_1 и S_2 окажется равным угловому размеру кружка Эри $\delta\varphi = 1,22\lambda / D$, где D – диаметр оправы линзы или диаметр апертурной диафрагмы объектива, т.е. если $\delta\psi = \delta\varphi = 1,22\lambda / D$. Отсюда находим разрешающую силу объектива:

$$R = D / (1,22\lambda). \quad (12.1)$$

Как видим, разрешающая сила объектива тем больше, чем больше диаметр его апертурной диафрагмы и чем меньше длина волны применяемого для освещения предмета света.

Формула (12.1) применима к объективам фото- и киноаппарата, зрительной трубы, телескопа. Например, для лучшего российского телескопа с диаметром объектива $D \approx 6$ м расчет по вышеприведенной формуле дает для наименьшего углового расстояния между звездами

$$\delta\psi = 1,22 \cdot 0,5 \cdot 10^{-6} / 6 \approx 1 \cdot 10^{-7} \text{ рад.}$$

(Расчет проведен для средней длины волны $\lambda = 0,5 \cdot 10^{-6}$ мкм.) Предел разрешения оптической системы зависит от многих факторов: длины волны света, размеров и геометрии линз, а также от типа системы формирования изображения.

Глаз при рассматривании удаленных предметов действует принципиально так же, как объектив телескопа, поэтому формула (12.1) применима и к глазу; роль D здесь играет диаметр зрачка глаза. Поскольку диаметр зрачка глаза $D \approx 3$ мм, то, полагая $\lambda = 0,5 \cdot 10^{-6}$ мкм, получим $\delta\psi = 2 \cdot 10^{-4}$ рад $\approx 0,7'$. Следовательно, минимальное угловое расстояние между двумя удаленными на большое расстояние от глаза точками, при котором глаз воспринимает их раздельно как две точки, примерно равно одной угловой минуте. Полученный результат хорошо согласуется с физиологической оценкой разрешающей способности глаза (расстояние между соседними светочувствительными элементами

сетчатки глаза соответствует этому угловому расстоянию). Этот пример демонстрирует удивительную способность живого организма приспосабливаться к окружающим условиям и в процессе эволюции достигать максимума того, что принципиально допускается законами природы.

Увеличение окуляра оптического прибора должно подбираться так, чтобы дифракционные кружки, разрешенные объективом, были разрешены и глазом, т.е. были видны под углом около $1'$. При фотографировании объекта линейное расстояние между центрами дифракционных картин двух близко расположенных точек должно превосходить размеры зерен эмульсии, что соответствует расстояниям порядка $0,01$ мм.

Если точечный источник света находится на конечном расстоянии a от линзы, то дифракционная картина с кружком Эри возникает не в фокальной плоскости, а на расстоянии a' за линзой в плоскости изображения. Радиус кружка Эри в этом случае будет определяться формулой

$$r_1' = 1,22\lambda a' / D.$$

Эта величина и является размером изображения точечного источника. При выполнении критерия Рэлея радиус r_1' равен расстоянию $\Delta r'$ между центрами кружков Эри. Расстоянию $\Delta r'$ в плоскости изображения соответствует расстояние $\Delta x = (a / a') \Delta r'$ в плоскости объекта (это соотношение вытекает из равенства $\delta\psi = \delta\varphi$). С учетом этого получаем оценку минимально разрешимого расстояния между точками объекта

$$\Delta x = 1,22\lambda a / D.$$

Произведем оценку Δx для глаза. Взяв в качестве величины a расстояние наилучшего зрения ($a = 25$ см), диаметр зрачка глаза $D = 3$ мм и $\lambda = 0,5 \cdot 10^{-3}$ мм, получим $\Delta x = 1,22 \cdot 0,5 \cdot 10^{-3} \cdot 250 / 3 \approx 0,05$ мм. Это и есть минимальное расстояние между

двумя точками объекта, при котором человеческий глаз воспринимает изображение этих точек отдельно как две точки.

Следует отметить, что на практике разрешающую способность фото- и кинообъективов определяют не по двум близко расположенным светящимся точкам (как в случае телеобъективов), а рассматривая визуально или фотографируя изображения испытательных тест-объектов, называемых мирами и представляющих собой систему параллельных друг другу светлых полос одинаковой толщины, разделенных темными промежутками той же толщины. При этом, однако, оказывается, что наименьшее разрешаемое расстояние между серединами изображений двух светлых полос примерно такое же, какое получается для двух светящихся точек.

Разрешающая способность микроскопа ограничена теми же физическими причинами, что и разрешающая способность рассмотренных выше оптических приборов – дифракцией света. Поскольку в микроскопе обычно реализуется случай когерентного освещения, то для определения разрешающей способности микроскопа можно воспользоваться рассуждениями, приведенными в предыдущем параграфе.

Для получения изображения необходимо, чтобы в объектив микроскопа попали лучи, проходящие через пары максимумов по меньшей мере нулевого и первого порядков (при этом воспроизводятся в изображении наиболее крупные детали объекта), в противном случае все поле зрения будет засвечено более или менее равномерно. Если объектом является решетка с периодом d , то указанное условие будет выполнено, т.е. основной период решетки будет разрешен, если угол дифракционной расходимости φ_1 , определяющий направление на первый дифракционный максимум, создаваемый решеткой, будет не больше половины угла раскрытия (апертуры) $2\sigma_A$ объектива (рис. 12.3), т.е. если $\sin \sigma_A \geq \sin \varphi_1 = 1 \cdot \lambda / d$. При таком условии обеспечивается возможность взаимодействия трех пучков $m = 0, \pm 1$. Следовательно, минимальный период решетки, который

может быть разрешен, будет равен $d = \lambda / \sin \sigma_A$. $d = \lambda / \sin \sigma_A$. В соответствии с этим наименьший размер доступный наблюдению объекта $h \approx d = \lambda / \sin \sigma_A$. Величину $R = 1 / h$ и принимают в качестве разрешающей силы микроскопа.

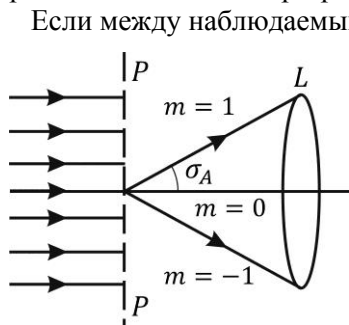


Рис. 12.3

Если между наблюдаемым объектом и объективом находится среда с показателем преломления n (иммерсия), то в приведенных выше соотношениях длину волны света в вакууме λ следует заменить длиной волны света в среде λ / n . Тогда $\sin \sigma_A \geq \lambda / n$. В этом случае минимальное расстояние, разрешаемое объективом микроскопа, будет $h = \lambda / (n \sin \sigma_A)$.

Величина $n \sin \sigma_A$ называется числовой апертурой микроскопа. Она примерно равна 1,5. Следовательно, минимальное разрешаемое расстояние $h \approx 0,7\lambda$. В видимой области $h \approx 0,4$ мкм.

Предел разрешения может быть сделан меньше указанного за счет применения наклонного освещения. При нормальном падении освещающего пучка за решеткой интерферируют три пучка – пучки, проходящие через максимумы порядков $m = 0, \pm 1$. При интерференции пучков нулевого и одного из первых порядков в изображении выявляется структура с основным периодом d . Такая же картина интерференции получается и при наложении пучка нулевого порядка с другим пучком первого порядка. Второй пучок первого порядка просто усиливает действие первого пучка того же порядка. Кроме того, пучки первого порядка интерферируют между собой, в результате чего на интерференционную картину накладываются новые, более слабые интерференционные полосы, расстояние между которыми вдвое меньше, т.е. равно $d / 2$. В этом проявляется действие первых гармоник в разло-

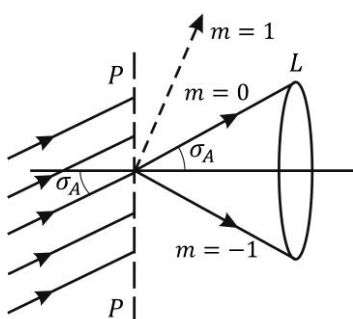


Рис. 12.4

жении функции пропускания решетки в ряд Фурье. Не меняя основного периода в изображении решетки, они несколько усиливают его контраст. Следовательно, комбинация максимума нулевого порядка с одним только максимумом первого порядка вполне достаточна для формирования изображения решетки, которое будет передавать основной

характер структуры решетки – ее периодичность с основным периодом d . Распределение интенсивности в плоскости изображения при этом будет синусоидальным, т.е. таким, каким оно было бы в случае решетки с гармонической функцией пропускания.

Максимально возможное разрешение получается, если освещающий пучок направить под углом σ_A к решетке, как показано на рис. 12.4. В этом случае не отклоненный ($m = 0$) и отклоненный ($m = 1$ или $m = -1$) пучки будут приходить на противоположные края апертуры объектива, т.е. в поле зрения оказывается нулевой и один из первых максимумов ($m = \pm 1$). Эти пучки могут интерферировать в плоскости изображения и создавать изображение решетки. В указанном направлении освещающего пучка угол раскрытия становится в два раза меньше: $\sin \sigma_A = 0,5\lambda/(nd)$. С учетом этого получаем наименьший размер объекта h , который еще можно наблюдать в микроскоп: $h = 0,5\lambda/(n \sin \sigma_A)$. Эта величина вдвое меньше полученного выше значения h .

Обратимся теперь снова к гармонической решетке. Такой решеточный объект будет полностью разрешим объективом, если все три его дифракционных максимума попадают в апертуру объектива, т.е. если координаты крайних дифракционных

максимумов (пиков) $x' = \pm \lambda f / d$ (формула (7.4)) не превосходят радиуса апертуры объектива: $(\lambda / d) f \leq D / 2$ или $d \geq (2\lambda / D) f$. Минимальный разрешимый период решетки

$$d = 2\lambda f / D. \quad (12.2)$$

Это соотношение называют критерием разрешения Аббе. Критерий Аббе можно сравнить с критерием Рэлея, который в случае когерентного освещения можно записать в виде

$$\delta x = 2\lambda f / D, \quad (12.3)$$

где δx – предел разрешения в плоскости предмета. Из сопоставления выражений (12.2) и (12.3) можно заключить, что критерии Аббе и Рэлея в действительности эквивалентны. Однако в большинстве случаев критерий Аббе позволяет лучше понять условия эксперимента, влияющие на окончательный предел разрешения.

Предел разрешения спектрального прибора с гармонической дифракционной решеткой можно вычислить по одной из следующих формул:

$$\delta\lambda = \lambda / N, \quad \delta x = \lambda / \theta, \quad \delta\varphi = \lambda / L, \quad (12.4)$$

где $\delta\lambda$ – минимальная разность длин волн двух спектральных линий, при которой эти линии воспринимаются как отдельные спектральные линии, λ – рабочая длина волны, N – полное число штрихов решетки, θ – угловая ширина решетки, $\delta\varphi$ – угловой предел разрешения, L – проекция ширины решетки на плоскость, перпендикулярную направлению наблюдения.

Разрешающая способность является важнейшей характеристикой и регистрирующей среды, которая также определяется как расстояние между двумя близко расположенными точками изображения, которые еще можно различить. Экспериментально эта величина находится с помощью

миры. Для обычных фотоматериалов разрешающая способность равна примерно 50 – 100 линий на 1мм.

12.2. Оптические передаточные функции дифракционно-ограниченных систем

Понятия оптической передаточной функции (ОПФ) $A(u, v)$ и дифракционной ограниченности оптической системы формирования изображения были введены в п. 9.7. Напомним, что оптическая передаточная функция определяется формулой (9.42) и представляет собой фурье-преобразование импульсной характеристики оптической системы по интенсивности $F_0(x, y)$:

$$A(u, v) = \int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} F_0(x, y) e^{-i2\pi(ux+vy)} dx dy.$$

Дифракционно-ограниченными системами называют такие системы, ограничение свойств которых обусловлено только конечными размерами реальных линз, т.е. дифракцией света на их апертуре. Оптическую передаточную функцию обычно определяют в нормированном виде:

$$A(u, v) = \frac{\int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} F_0(x, y) e^{-i2\pi(ux+vy)} dx dy}{\int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} F_0(x, y) dx dy}. \quad (12.5)$$

Так как $F_0(x, y) = \left| \bar{h}(x, y) \right|^2$, то интеграл в числителе (12.5) есть фурье-преобразование от $\left| \bar{h}(x, y) \right|^2$, а в знаменателе – то же фурье-преобразование, но при $u = v = 0$. Следовательно,

$$A(u, v) = \frac{\text{F}\left\{\left|\bar{h}(x, y)\right|^2\right\}}{\text{F}\left\{\left|\bar{h}(x, y)\right|^2\right\}_{u=v=0}} = \frac{\int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} |h(x, y)|^2 e^{-i2\pi(ux+vy)} dx dy}{\int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} |h(x, y)|^2 dx dy}.$$

Нормировка ОПФ на ее значение при нулевых частотах вводится для того, чтобы учесть всегда присутствующий фон, влияющий на контраст изображения. Учитывая, что $h(x, y)$ является обратным преобразованием Фурье от когерентной оптической функции (КОФ), т.е. от функции

$$h(x, y) = \int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} H(u', v') e^{i2\pi(u'x+v'y)} du' dv',$$

и используя так называемую теорему автокорреляции, из которой следует, что

$$\text{F}\left\{\left|\bar{h}(x, y)\right|^2\right\} = \int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} H(u', v') H^*(u' + u, v' + v) du' dv',$$

а значит,

$$\begin{aligned} \text{F}\left\{\left|\bar{h}(x, y)\right|^2\right\}_{u=v=0} &= \int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} H(u', v') H^*(u', v') du' dv' = \\ &= \int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} |H(u', v')|^2 du' dv', \end{aligned}$$

находим

$$A(u, v) = \frac{\int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} H(u', v') H^*(u' + u, v' + v) du' dv'}{\int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} |H(u', v')|^2 du' dv'}.$$

Произведя замену переменных $\xi = u' + u / 2$, $\eta = v' + v / 2$, перепишем выражение для $A(u, v)$ в симметричной форме

$$A(u, v) = \frac{\int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} H(\xi - u/2, \eta - v/2) H^*(\xi + u/2, \eta + v/2) d\xi d\eta}{\int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} |H(\xi, \eta)|^2 d\xi d\eta}.$$

Полученное выражение определяет основную связь между свойствами когерентных и некогерентных систем. Подставляя сюда вместо $H(u, v)$ выражение (11.22), $H(u, v) = P(\lambda u, \lambda v)$, получим

$$A(u, v) = \frac{\int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} P(\xi - \lambda lu/2, \eta - \lambda lv/2) P(\xi + \lambda lu/2, \eta + \lambda lv/2) d\xi d\eta}{\int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} P(\xi, \eta) d\xi d\eta}, \quad (12.6)$$

где учтено, что $P^2 = P$ (так как функция P^2 принимает те же два значения 1 или 0, что и функция P) и что функция P действительная, а значит, $P^* = P$; ξ, η – координаты в плоскости зрачка системы. Таким образом, ОПФ дифракционно-ограниченной оптической системы определяется дифракцией в зрачке и зависит только от формы, размеров и положения зрачка относительно плоскости изображения. Выражение (12.6) можно интерпретировать геометрически. Числитель этой дроби представляет собой площадь $\Delta S(u, v)$ области перекрытия двух смещенных относительно друг друга контуров зрачка, центр одного из которых лежит в точке $(\lambda lu/2, \lambda lv/2)$, а центр другого – в симметричной относительно начала координат точке

$(-\lambda l u / 2, -\lambda l v / 2)$; знаменатель дроби есть полная площадь S_0 зрачка. Таким образом, $A(u, v) = \Delta S(u, v) / S_0$.

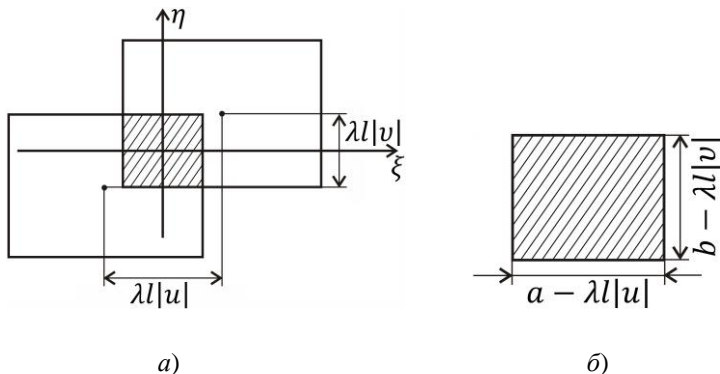


Рис. 12.5

Такая геометрическая интерпретация показывает, что для дифракционно-ограниченных оптических систем оптическая передаточная функция является действительной, неотрицательной функцией пространственных частот u и v , обращающейся в нуль, когда пространственные частоты становятся выше определенных граничных значений u_0, v_0 . На данной геометрической интерпретации и основан расчет ОПФ дифракционно-ограниченных оптических систем.

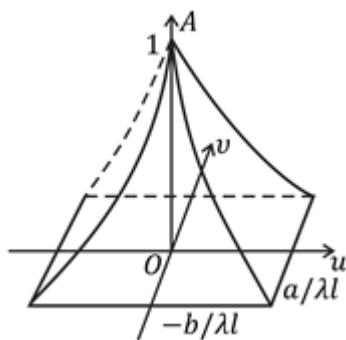


Рис. 12.6

Произведем расчет ОПФ дифракционно-ограниченных систем с прямоугольным и круглым входными зрачками (рис. 11.2). Рис. 12.5 иллюстрирует расчет для случая прямоугольного зрачка. Область перекрытия на рис. 12.5, a заштрихована. Размеры области перекрытия показаны

на рис. 12.5, б. Как видно на этом рисунке, площадь области перекрытия

$$\Delta S(u, v) = (a - \lambda l |u|)(b - \lambda l |v|)$$

при $|u| \leq a/\lambda l$, $|v| \leq b/\lambda l$ и

$$\Delta S(u, v) = 0$$

в остальных случаях, где a и b – размеры прямоугольного зрачка.

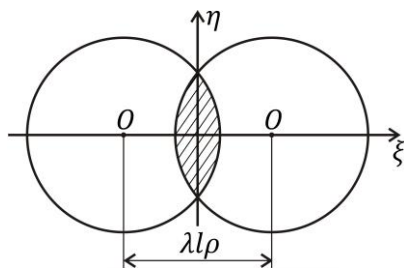
Разделив на площадь зрачка ab , получим

$$A(u, v) = \Lambda\left(\frac{\lambda l |u|}{a}\right) \Lambda\left(\frac{\lambda l |v|}{b}\right), \quad (12.7)$$

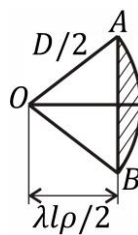
где $\Lambda(x)$ – введенная в п. 4.2 треугольная функция,

$$\Lambda(x) = \begin{cases} 1 - |x|, & |x| < 1; \\ 0, & |x| \geq 1 \end{cases}$$

(в рассматриваемом случае $x = \lambda l |u|/a$ и $x = \lambda l |v|/b$).



а)



б)

Рис. 12.7

График ОПФ $A(u, v)$ для прямоугольного зрачка показан на рис. 12.6. Предельные частоты (частоты, выше которых $A(u, v)$

обращается в нуль) $|u_0| = a / \lambda l$, $|v| = b / \lambda l$. Заметим, что эти частоты в два раза превышают соответствующие значения для когерентной передаточной функции (КПФ) $H(u, v)$.

Если в формуле (12.7) положить $a = b$, получим ОПФ для системы с квадратным зрачком:

$$A(u, v) = \Lambda\left(\frac{\lambda l |u|}{a}\right) \Lambda\left(\frac{\lambda l |v|}{a}\right)$$

Расчет ОПФ системы с круглым зрачком иллюстрируется рис. 12.7. В случае круглого выходного зрачка ОПФ будет, очевидно, обладать круговой симметрией. Поэтому достаточно определить ОПФ $A(u, v)$ вдоль произвольной радиальной оси в частотной плоскости. Пространственную частоту в этом направлении обозначим через ρ . Очевидно, что

$$\rho = \sqrt{u^2 + v^2}.$$

На рис. 12.7, б видно, что площадь области перекрытия равна удвоенной площади заштрихованного кругового сегмента. Эта площадь равна разности площадей кругового сектора и площади

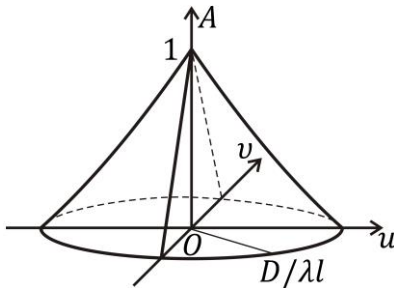


Рис. 12.8

треугольника OAB :

$$\begin{aligned} \frac{1}{2} \Delta S(u, v) &= \frac{1}{2\pi} \cdot \frac{\pi D^2}{4} - \frac{1}{2} \cdot \frac{\lambda l \rho}{2} \sqrt{\left(\frac{D}{2}\right)^2 - \left(\frac{\lambda l \rho}{2}\right)^2} = \\ &= \frac{D^2}{8} \arccos\left(\frac{\lambda l \rho}{D}\right), \end{aligned}$$

где D – диаметр зрачка. А так как $S_0 = \pi D^2 / 4$, то для ОПФ оптической системы с круглым зрачком будем иметь:

$$A(\rho) = \frac{2}{\pi} \left[\arccos\left(\frac{\lambda l \rho}{D}\right) - \frac{\lambda l \rho}{D} \sqrt{1 - \left(\frac{\lambda l \rho}{D}\right)^2} \right] \quad (12.8)$$

при $\rho \leq D / \lambda l$ и $A(\rho) = 0$ в остальных случаях. График этой функции показан на рис. 12.8. Граничная частота ОПФ в этом случае $\rho_0 = D / \lambda l$. Как видим, и для круглого зрачка граничная частота для ОПФ в два раза больше той же частоты для когерентной ной передаточной функции $H(u, v)$.

Графики, изображенные на рис. 12.6 и 12.8, показывают, что оптическая передаточная функция является убывающей функцией аргументов u и v ; она достигает своего максимального значения, равного единице при $u = v = 0$ и обращается в нуль при $u = u_0$ и $v = v_0$.

12.3. Функции рассеяния точки, линии и края дифракционно-ограниченных систем

Оценка качества оптической системы, формирующей изображение, представляет собой довольно сложную проблему. Общего критерия такой оценки не существует. Поэтому на практике используют различные упрощенные критерии, которые, хотя и не являются универсальными, вполне пригодны для сравнительной оценки качеств оптической системы. Одним из таких критериев является, например, экспериментальное определение разрешающей способности оптической системы. В настоящее время наибольшее распространение получил способ оценки, основанный на анализе изображения простейших объектов: светящейся точки, линии, границы полуплоскости (края), решетки со ступенчатой или синусоидальной функцией

пропускания и т.п. Функции, описывающие распределение интенсивности на изображении точки, линии и края (функции рассеяния), являются теми элементарными характеристиками объектива, на основе которых могут быть установлены любые возможные способы оценки качества оптического изображения, например передаточные функции системы. Эти функции обеспечивают более информативную оценку оптической системы, формирующей изображение, чем простое измерение ее предела разрешения. Светящаяся точка, линия и край полуплоскости – это те элементы, каждый из которых в своей совокупности образует изображаемый объект. Действие оптической системы на эти элементы можно суммировать (что допустимо ввиду линейности оптической системы) и тем самым описать изображение объекта.

Следует, однако, отметить, что и эти функции не дают полного решения проблемы оценки качества системы, особенно если в окончательном формировании изображения участвует глаз. Глаз является плохой системой формирования изображения, но он связан со сложной обработкой данных в сетчатке и мозге. Это существенно усложняет определение отклика системы.

Функции рассеяния были рассмотрены нами в п. 9.8. Здесь мы проведем вычисление этих функций для некоторых дифракционно-ограниченных систем с симметричными зрачками.

Найдем функцию рассеяния точки для дифракционно-ограниченной системы с прямоугольным и круглым зрачками. Для расчета используем формулу (9.45)

$$F_0(x, y) = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} A(u, v) e^{i2\pi(ux+vy)} dudv. \quad (12.9)$$

Оптическая передаточная функция $A(u, v)$ для системы с прямоугольным зрачком, как видно из соотношения (12.7), является функцией с разделяющимися переменными, поэтому и ее ФРТ можно представить в виде произведения функций одной переменной: $F_0(x, y) = F_{01}(x)F_{02}(y)$. Подставляя в (12.9) выражение (12.7) для ОПФ системы с прямоугольным зрачком и замечая, что переменные с такой ОПФ разделяются, ФРТ

рассматриваемой системы находим одномерным обратным преобразованием Фурье каждого сомножителя:

$$F_0(x, y) = \int_{-\infty}^{\infty} \Lambda\left(\frac{\lambda lu}{a}\right) e^{i2\pi ux} du \int_{-\infty}^{\infty} \Lambda\left(\frac{\lambda lv}{b}\right) e^{i2\pi vy} dv.$$

Учитывая, что

$$\int_{-\infty}^{\infty} \Lambda(cu) \exp(i2\pi ux) du = \frac{\text{sinc}^2(x/c)}{c},$$

получим для ФРТ оптической системы с прямоугольным зрачком размером $a \times b$

$$F_0(x, y) = \frac{a}{\lambda l} \text{sinc}^2\left(\frac{ax}{\lambda l}\right) \frac{b}{\lambda l} \text{sinc}^2\left(\frac{by}{\lambda l}\right). \quad (12.10)$$

Сравнивая это выражение с выражением (6.24), видим, что с точностью до постоянного множителя распределение интенсивности в изображении точки

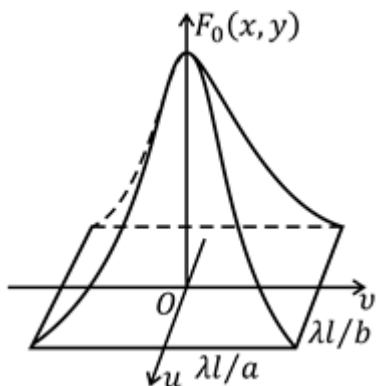


Рис. 12.9

в изображении точки оптической системой с прямоугольным зрачком описывается той же функцией, что и распределение интенсивности в дифракционной картине при дифракции Фраунгофера от прямоугольного отверстия. Отличие только в аргументах функций: в рассматриваемом случае аргументами являются переменные $ax/\lambda l$ и $by/\lambda l$, а

в функции $I(x', y')$ – соответственно $ax'/\lambda f$ и $by'/\lambda f$. Функция $F_0(x, y)$ имеет максимум, равный $ab/(\lambda l)^2$, при

$ax/\lambda l = by/\lambda l = 0$, т.е. при $x = y = 0$, и первый минимум при $ax/\lambda l = by/\lambda l = 1$, т.е. при $x = \lambda l/a$ и $y = \lambda l/b$. Затем эта функция снова возрастает и убывает, осциллируя стремится к нулю. Фигура, изображающая функцию $F_0(x, y)$ в пределах центрального максимума, показана на рис. 12.9. В основании этой фигуры лежит прямоугольник со сторонами $2\lambda l/a$ и $2\lambda l/b$. Заметим, что большей стороне прямоугольного зрачка соответствует меньшая сторона изображения и наоборот.

Положив в выражении (12.10) $a = b$, получим ФРТ для оптической системы с квадратным зрачком:

$$F_0(x, y) = \left(\frac{a}{\lambda l}\right)^2 \operatorname{sinc}^2\left(\frac{ax}{\lambda l}\right) \operatorname{sinc}^2\left(\frac{ay}{\lambda l}\right).$$

Для системы с круглым зрачком ФРТ обладает круговой симметрией (симметрией вращения). Для такой системы обратное преобразование Фурье заменяется обратным преобразованием Ганкеля:

$$f(r) = 2\pi \int_0^{\infty} F(\rho) J_0(2\pi r\rho) \rho d\rho. \quad (12.11)$$

Поэтому для определения ФРТ с круглым зрачком вместо формулы (12.9) следует использовать формулу обратного преобразования Ганкеля. Произведя в формуле (12.11) замену $f(r)$ на $F_0(r)$, а $F(\rho)$ на $A(\rho)$, получим формулу для определения ФРТ системы с круглым зрачком:

$$F_0(r) = 2\pi \int_0^{\infty} A(\rho) J_0(2\pi r\rho) \rho d\rho.$$

Подставляя сюда выражение (12.8) для ОПФ системы с круглым зрачком и опуская достаточно сложные выкладки, запишем конечный результат:

$$F_0(r) = \frac{2J_1^2(\alpha)}{\alpha^2}, \quad (12.12)$$

где $J_1(\alpha)$ – функция Бесселя первого рода первого порядка,

$$\alpha = \pi D r / (\lambda l).$$

Сравнение выражения (12.12) с выражением (6.29) показывает, что с точностью до постоянного множителя распределение интенсивности в изображении точки, даваемое оптической системой с круглым зрачком, такое же, как и распределение интенсивности в дифракционной картине при дифракции Фраунгофера от круглого отверстия. В плоскости изображения это распределение представляет собой диск Эри, окруженный чередующимися концентрическими темными и светлыми кольцами. Функция $F_0(r)$ имеет максимум при $\alpha = \pi D r / (\lambda l) = 0$, т.е. при $r = 0$. Первый минимум этой функции достигается при $\pi D r / (\lambda l) = 3,83$, откуда для радиуса r_1 первого темного кольца, окружающего центральный максимум (радиуса диска Эри), получаем $r_1 = 1,22 \lambda l / D$. График функции $F_0(r)$ показан на рис. 12.10.

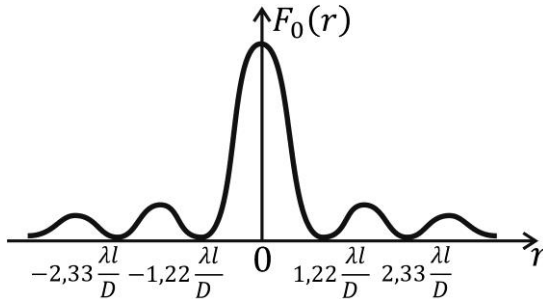


Рис. 12.10

Расчет функции рассеяния линии дифракционно-ограниченных систем можно выполнить с помощью формулы (9.49)

$$F_1(x) = \int_{-\infty}^{\infty} F_0(r) dy.$$

На основании того, что функция рассеяния линии систем, имеющих ФРТ с разделяющимися переменными, с точностью до постоянного множителя равна одному из сомножителей ФРТ ($F_1(x) = C_1 F_{01}(x)$) и соотношения (12.10), находим ФРЛ дифракционно-ограниченной системы с прямоугольным зрачком:

$$F_1(x) = C_1 \frac{a}{\lambda l} \operatorname{sinc}^2\left(\frac{ax}{\lambda l}\right).$$

Из условия нормировки

$$\int_{-\infty}^{\infty} F_1(x) dx = C_1 \int_{-\infty}^{\infty} \frac{a}{\lambda l} \operatorname{sinc}^2\left(\frac{ax}{\lambda l}\right) dx = C_1 \cdot 1 = 1$$

находим, что $C_1 = 1$. Следовательно, ФРЛ системы с прямоугольным зрачком определится как

$$F_1(x) = \frac{a}{\lambda l} \operatorname{sinc}^2\left(\frac{ax}{\lambda l}\right). \quad (12.13)$$

Если сравнить это выражение с выражением (6.15), легко убедиться, что распределение интенсивности в изображении светящейся линии (щели), даваемое оптической системой с прямоугольным зрачком, имеет тот же вид, что и распределение интенсивности при дифракции Фраунгофера от щели шириной a , если вместо переменной $ax' / (\lambda f)$ в функции (6.15) взять переменную $ax / (\lambda l)$. График функции рассеяния линии $F_1(x)$ для системы с прямоугольным зрачком представлен на рис. 12.11. Эта функция имеет максимум при $x = 0$ и первый минимум при $x = \pm \lambda l / a$.

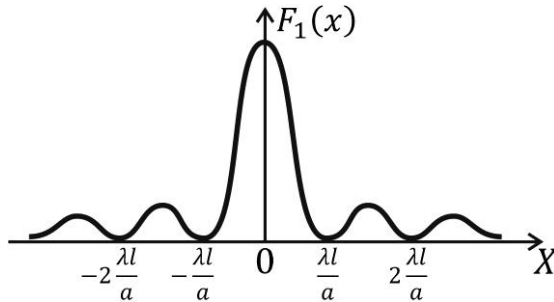


Рис. 12.11

Расчет ФРЛ для дифракционно-ограниченной оптической системы с круглым зрачком может быть произведен с помощью преобразования Абеля от функции рассеяния точки этой системы:

$$F_1(x) = \int_x^{\infty} \frac{rF_0(r)}{\sqrt{r^2 - x^2}} dr.$$

Положив в этой формуле

$$F_0(r) = 2 \left(\frac{J_1(r_i)}{r_i} \right)^2,$$

где $r_i = \pi D r / (\lambda l)$, придем к интегралу

$$F_1(x_i) = 4 \int_{x_i}^{\infty} \frac{J_1^2(r_i)}{r_i \sqrt{r_i^2 - x_i^2}} dr_i, \quad (12.14)$$

где $x_i = \pi D x / (\lambda l)$. Интеграл (12.14), называемый интегралом Струве, представляется знакпеременным рядом

$$F_1(x_i) = 1 - \frac{4}{3 \cdot 5} x_i^2 + \frac{4^2}{3 \cdot 5^2 \cdot 7} x_i^4 - \frac{4^3}{3 \cdot 5^2 \cdot 7^2 \cdot 9} x_i^6 + \dots$$

Функция рассеяния края дифракционно-ограниченных оптических систем производится с помощью выражения (9.50)

$$F_2(x) = \int_{-\infty}^x F_1(\xi) d\xi. \quad (12.15)$$

Однако расчет ФРК более сложен, чем расчет других функций рассеяния, поэтому ограничимся здесь более простым случаем системы с прямоугольным зрачком. Подставляя в формулу (12.15) выражение (12.13), получим

$$F_2(x) = \int_{-\infty}^x \frac{a}{\lambda l} \operatorname{sinc}^2\left(\frac{a\xi}{\lambda l}\right) d\xi.$$

Представим этот интеграл в виде суммы интегралов:

$$F_2(x) = \int_{-\infty}^0 \frac{a}{\lambda l} \operatorname{sinc}^2\left(\frac{a\xi}{\lambda l}\right) d\xi + \int_0^x \frac{a}{\lambda l} \operatorname{sinc}^2\left(\frac{a\xi}{\lambda l}\right) d\xi.$$

Первый интеграл, как нетрудно убедиться, равен 1/2. Следовательно,

$$F_2(x) = \frac{1}{2} + \int_0^x \frac{a}{\lambda l} \operatorname{sinc}^2\left(\frac{a\xi}{\lambda l}\right) d\xi.$$

Представим функцию $\operatorname{sinc}^2(x_1)$, где $x_1 = a\xi / (\lambda l)$, в виде ряда

$$\operatorname{sinc}^2(x_1) = \frac{1}{2} - \frac{x_1^2}{4!} + \frac{x_1^4}{6!} - \dots$$

и произведя почленное интегрирование, получим

$$F_2(x) = \frac{1}{2} + \frac{ax}{\lambda l} - \frac{1}{3 \cdot 4!} \left(\frac{ax}{\lambda l}\right)^3 + \frac{1}{5 \cdot 6!} \left(\frac{ax}{\lambda l}\right)^5 - \dots$$

График ФРК системы с прямоугольным зрачком, рассчитанной с помощью этого ряда, показан на рис. 9.4.

Расчет функций рассеяния не всегда возможен. Часто их находят экспериментальным путем. Следует, однако, отметить, что эти эксперименты (особенно эксперименты по определению ФРТ) достаточно сложные.

12.4. Формирование изображения при некогерентном освещении

Рассмотрим теперь другой предельный случай¹¹, когда изображаемый оптической системой объект просвечивается (в случае транспаранта с изображением предмета) или освещается (в случае непрозрачного объекта) некогерентным светом. В этом случае все точки объекта излучают некогерентный свет, а оптическая система, как было показано в п. 9.6, оказывается линейной по интенсивности. Распределение интенсивности в плоскости изображения $I'(x', y')$ в этом случае можно определить с помощью интеграла суперпозиции (9.38)

$$I'(x', y') = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} I(x, y) F_0(x' - x, y' - y) dx dy. \quad (12.16)$$

где $I(x, y)$ – распределение интенсивности света в плоскости предмета, $F_0(x' - x, y' - y) = |h(x' - x, y' - y)|^2$ – импульсный отклик системы по интенсивности или функция рассеяния точки (ФРТ). Функцию $F_0(x' - x; y' - y)$ называют также функцией размытия системы по интенсивности. Она представляет собой интенсивность поля в точке (x', y') изображения, которое возникает от изображения точечного объекта, описываемого интенсивностью в виде δ -функции в точке (x, y) . Координаты x и

¹¹ Первый предельный случай был рассмотрен в п. 11.7

x и y представляют собой координаты идеального изображения (приведенные координаты объекта) и определяются произведениями координат x_0 , y_0 точек объекта на коэффициент линейного увеличения β оптической системы: $x = \beta x_0$, $y = \beta y_0$. В этом случае численные значения координат точек предмета и их изображения будут одинаковыми.

Получим соотношение (12.16) другим способом. Пусть $I(x, y)$ – интенсивность в произвольной точке (x, y) плоскости предмета, а $F_0(x', y', x, y)$ – функция пропускания системы. Тогда интенсивность света, достигающего точки (x', y') в плоскости изображения и выходящего из элемента $dx dy$ с центром в точке (x, y) предмета в соответствии с принципом Гюйгенса – Френеля будет

$$dI'(x', y') = I(x, y)F_0(x, y)dx dy.$$

Поскольку свет, исходящий из различных точек предмета, не когерентен, интенсивности, создаваемые различными элементами предмета, складываются, так что полная интенсивность в точке (x', y') определится интегралом (12.16).

К интегралу суперпозиции (12.16) можно прийти и эвристически на основе следующих соображений. Вследствие многочисленных причин (дифракция на апертуре, aberrации линз и др.) изображение точки (x, y) предмета не сводится в сопряженную ей точку (x', y') , как в случае идеального изображения, а распространяется на некоторую область, окружающую эту точку. Эту область называют элементарным изображением. Согласно модели Рэлея формирования оптического изображения, которая применяется при некогерентном освещении, изображение всего протяженного объекта при этом будет определяться комбинацией таких элементарных изображений. В простейшем случае, когда все прочие эффекты, кроме эффекта дифракции света на апертуре, отсутствуют, элементарное изображение представляет собой

дифракционную картину в виде светового диска Эри, окруженного чередующимися темными и светлыми кольцами. В общем же случае по ряду причин (например, вследствие аберраций) элементарное изображение представляет собой более сложную картину распределения интенсивности. Распределение интенсивности в элементарном изображении можно представить функцией $F_0(x', y', x, y)$, являющейся функцией рассеяния точки. Будем считать, что оптическая система является пространственно-инвариантной, т.е. такой, что при изменении положения изображаемой точки ее изображение тоже изменяет только свое положение, но не изменяет своего вида. В этом случае функция, описывающая изображение точки, т.е. функция $F_0(x', y'; x, y)$, будет пространственно-инвариантной: $F_0(x', y', x, y) = F_0(x' - x, y' - y)$. Будем предполагать, что оптическая система, создающая изображение, является не только пространственно-инвариантной, но и линейной по интенсивности (т.е. при передаче интенсивности от предмета к изображению к системе применим принцип суперпозиции). Тогда, поскольку излучение от всех точек объекта не когерентно, распределение интенсивности $I'(x', y')$ на изображении можно представить в виде суммы интенсивностей, даваемых каждым из изображений, полученных от различных точек объекта. При этом вклады в интенсивность изображения от всех точек (x, y) предмета суммируются с «весами» (коэффициентами), равными соответствующим значениям интенсивности $I(x, y)$ в плоскости предмета. С учетом этих предположений мы и приходим к интегралу (12.16).

Соотношение (12.16) описывает процесс образования изображения, если элементы объекта излучают некогерентный свет. Интеграл суперпозиции (12.16), в отличие от интеграла суперпозиции (11.9), выражает собой суммирование изображений различных точек объекта по интенсивности, а не по комплексным амплитудам. Он выражает тот факт, что линейная по интенсивности оптическая система полностью характеризуется ее откликом на входные импульсы (отсюда и еще одно название

функции $F_0(x', y', x, y)$ – аппаратная функция). Действие оптической системы может быть полностью описано, если определены изображения точечных источников, расположенных по всему полю предмета.

Если бы оптическая система была идеальной, то поступающий на вход сигнал в виде δ -функции создавал бы на выходе сигнал тоже в виде δ -функции, а значит, точечный источник света, представляемый двумерной δ -функцией, изображался бы в виде точки. В этом случае распределение интенсивности в изображении $I'(x', y')$ качественно не отличалось бы от распределения интенсивности в плоскости предмета $I(x, y)$. Это распределение получалось бы из $I(x, y)$ простой заменой $x \rightarrow x'$, $y \rightarrow y'$. Действительно, для любой линейной идеальной (не искажающей входной сигнал) передающей системы выходной сигнал можно представить в виде линейной комбинации взвешенной (т.е. входящей с определенными коэффициентами – «весами») смещенной системы δ -функций, причем в качестве «весов» берутся значения, описывающие входной сигнал функции в соответствующих точках. Так что,

$$I'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) \delta(x' - x; y' - y) dx dy. \quad (12.17)$$

Откуда по фильтрующему свойству δ -функции получаем, что $I'(x', y') = I(x', y')$. Изображение в этом случае получается идеальным и увеличенным в плоскости изображения.

Но никакая оптическая система идеальной не является, и поэтому поступающий на вход бесконечно узкий сигнал в виде δ -функции после прохождения через оптическую систему преобразуется в сигнал конечной ширины, описываемый функцией $F_0(x', y', x, y)$: $\delta(x' - x, y' - y) \rightarrow F_0(x' - x, y' - y)$; точечный источник, находящийся в точке (x, y) , изображается не в виде точки (x', y') , а в виде отдельной картины, распределение

интенсивности в которой дается функцией $F_0(x' - x, y' - y)$. Заменяя в (12.17) $\delta(x' - x, y' - y)$ на $F_0(x' - x, y' - y)$, приходим к соотношению (12.16). В этом случае распределение интенсивности $I'(x', y')$ в плоскости изображения уже не будет просто повторять в увеличенном масштабе распределение интенсивности $I(x, y)$ в плоскости предмета. Оно будет размытым, и это размытие будет определяться функцией $F_0(x' - x, y' - y)$. (Действие размытия на изображение выражается в расширении изображения и в том, что переходы и границы между соседними частями получаются в изображении более плавными, более сглаженными). Но когда распределение одной функции происходит в соответствии с законом, определяемым другой функцией, говорят о свертке этих функций, чем и является соотношение (12.16).

Взяв преобразование Фурье от свертки (12.16), приходим к соотношению

$$I'(u, v) = I(u, v) A(u, v), \quad (12.18)$$

где $I'(u, v)$ и $I(u, v)$ – фурье-образы распределения интенсивности на изображении и на предмете соответственно, $A(u, v)$ – передаточная функция оптической системы. Это соотношение показывает, что распределение интенсивности на изображении получается с помощью последовательности, состоящей из преобразований Фурье распределения интенсивности на объекте и функции рассеяния, умножения полученных фурье-образов и следующего преобразования Фурье (на последнем этапе получения распределения $I'(x', y')$ из его фурье-образа $I'(u, v)$): $I'(-x', -y') = F\{I'(u, v)\}$, так как $F\{F\{f(x, y)\}\} = f(-x, -y)$. Такое представление образования изображения во многом дополняет модель Аббе – Портера, согласно которой формирование оптического изображения представляет собой процесс двух последовательных преобразований Фурье (прямого и обратного).

Соотношение (12.18) показывает также, что если объект представлен в виде суммы гармоник различных пространственных частот, то изображение, построенное линейной оптической системой, получается суммированием соответствующих гармоник, для каждой из которых действие изображающей системы учитывается своим весовым коэффициентом $A(u, v)$.

В одномерном случае, как было показано в п. 9.7 (формула (9.51)), распределение интенсивности на изображении

$$I'(x') = \int_{-\infty}^{\infty} I(x) F_1(x' - x) dx \quad (12.19)$$

или в соответствии со свойством коммутативности операции свертки

$$I'(x') = \int_{-\infty}^{\infty} F_1(x) I(x' - x) dx . \quad (12.20)$$

Произведя преобразование Фурье соотношения (12.20), получим в одномерном случае

$$I'(u) = A(u)I(u), \quad (12.21)$$

где функция пространственной частоты

$$A(u) = \int_{-\infty}^{\infty} F_1(x) e^{-i2\pi ux} dx \quad (12.22)$$

представляет собой одномерную оптическую передаточную функцию. Одномерный случай чаще всего встречается при измерениях ОПФ, когда объектом является решетка из параллельных штрихов, узкая щель и т.п.

Рассмотрим теперь несколько примеров применения формул (12.20) и (12.21).

12.4.1. Изображение простейших объектов

1. Равномерное освещение. Распределение интенсивности в плоскости предмета $I(x) = I_0 = \text{const}$. Интенсивность в плоскости изображения, согласно формуле (12.20), будет

$$I'(x') = I_0 \int_{-\infty}^{\infty} F_1(x) dx = I_0 = \text{const}.$$

Таким образом, при равномерном распределении интенсивности света в плоскости предмета получается равномерное распределение интенсивности и в плоскости изображения.

2. Тожественная передача. В случае тождественной передачи функция рассеяния $F_1(x' - x) = \delta(x' - x)$. Тогда на основании фильтрующего свойства δ -функции получим

$$I'(x') = \int_{-\infty}^{\infty} I(x) \delta(x' - x) dx = I(x').$$

Распределения $I'(x')$ и $I(x')$ идентичны, так как координата x' пробегает те же значения, что и x .

3. Край полуплоскости. Интенсивность в плоскости предмета задается распределением:

$$I(x) = \begin{cases} 0, & x > 0, \\ I_0, & x \leq 0. \end{cases}$$

Распределение интенсивности в плоскости изображения

$$I'(x') = I_0 \int_{-\infty}^0 F_1(x' - x) dx = I_0 \int_{x'}^{\infty} F_1(x) dx.$$

При этом напротив края интенсивность $I'(0) = I_0/2$, если функция $F_1(x)$ нечетная, т.е. если $F_1(-x) = -F_1(x)$. Распределение поля в плоскости изображения имеет вид, показанный на рис. 9.5.

4. Щель шириной a . Распределение интенсивности на объекте

$$I(x) = \begin{cases} I_0, & |x| \leq a/2, \\ 0, & |x| > a/2. \end{cases}$$

Распределение интенсивности на изображении

$$I'(x') = I_0 \int_{-a/2}^{a/2} F_1(x' - x) dx = I_0 \int_{x' - a/2}^{x' + a/2} F_1(x) dx.$$

При $x' = 0$ (напротив середины щели) получается максимальное значение интенсивности

$$I'(0) = I_0 \int_{-a/2}^{a/2} F_1(x) dx.$$

При $x' = a/2$ (напротив верхнего края щели) имеем

$$I'(a/2) = I_0 \int_0^a F_1(x) dx.$$

Аналогично напротив нижнего края ($x' = -a/2$)

$$I'(-a/2) = I_0 \int_{-a}^0 F_1(x) dx.$$

12.4.2. Изображение гармонической дифракционной решетки

Решение этой задачи имеет важное значение в фурье-оптике, так как оно позволяет выяснить, как оптическая система передает в плоскости изображения спектр пространственных частот, и найти величины, характеризующие эту передачу. Распределение интенсивности на объекте, т.е. функцию $I(x, y)$, можно разложить в двумерный интеграл Фурье, т.е. представить ее в виде совокупности бесконечного множества гармонических составляющих, характеризующихся определенной пространственной частотой, амплитудой и фазой, поэтому представляется важным выяснить, как оптическая система воздействует на отдельные компоненты пространственных частот объекта. С этой целью рассмотрим изображение объекта, обладающего распределением интенсивности, состоящим из одной гармонической компоненты определенной пространственной частоты.

Пусть объект в виде дифракционной решетки имеет косинусоидальную функцию пропускания по интенсивности

$$T(x) = a + b \cos 2\pi u x,$$

где $u = 1 / d$, d – период решетки, a и b – постоянные. Тогда распределение интенсивности в плоскости непосредственно за решеткой (входной плоскости)

$$I(x) = I_0 T(x) = \bar{I} + I_1 \cos 2\pi u x, \quad (12.23)$$

где $\bar{I} = aI_0$ – среднее значение интенсивности (интенсивность равномерного фона), $I_1 = bI_0$ – амплитуда колебания интенсивности во входной плоскости. Заменяя в (12.23) x на $x' - x$, подставляя это в интеграл суперпозиции (12.20) и раскрыв $\cos 2\pi u(x' - x)$, будем иметь

$$I'(x') = \int_{-\infty}^{\infty} F_1(x) [\bar{I} + I_1 \cos 2\pi u(x' - x)] dx = \bar{I} +$$

$$+ I_1 \left(\cos 2\pi ux' \int_{-\infty}^{\infty} F_1(x) \cos 2\pi uxdx + \right. \\ \left. + \sin 2\pi ux' \int_{-\infty}^{\infty} F_1(x) \sin 2\pi uxdx \right).$$

Введя обозначение

$$c(u) = \int_{-\infty}^{\infty} F_1(x) \cos 2\pi uxdx ; \quad s(u) = \int_{-\infty}^{\infty} F_1(x) \sin 2\pi uxdx ,$$

это соотношение можно переписать в виде

$$I'(x') = \bar{I} + I_1 [c(u) \cos 2\pi ux' + s(u) \sin 2\pi ux'] = \\ = \bar{I} + T(u) I_1 \cos[2\pi ux' - \varphi(u)], \quad (12.24)$$

где

$$T(u) = \sqrt{c^2(u) + s^2(u)}, \quad \varphi(u) = \arctg \frac{s(u)}{c(u)}.$$

Соотношение (12.24) показывает, что косинусоидальная решетка изображается в виде косинусоидальной же решетки, т.е. гармоническая компонента пространственного распределения передается линейной по интенсивности пространственно-инвариантной оптической системой также в виде гармонической компоненты. Если произвольный объект можно представить в виде суммы гармонических составляющих, то и изображение будет передаваться в виде суммы гармонических составляющих. При этом среднее значение интенсивности \bar{I} и пространственная частота u гармоники остаются неизменными; меняются только ее амплитуда I_1 (она изменяется в $T(u)$ раз) и фаза $\varphi(u)$.

Покажем, что функция $T(u)$ представляет собой модуль комплексной одномерной ОПФ, а $\varphi(u)$ – ее аргумент:

$$A(u) = T(u)\exp[-i\varphi(u)].$$

Действительно, используя формулу Эйлера, одномерную ОПФ можно представить в виде

$$A(u) = \int_{-\infty}^{\infty} F_1(x) \exp(-i2\pi ux) dx = \\ \int_{-\infty}^{\infty} F_1(x) \cos 2\pi ux dx - i \int_{-\infty}^{\infty} F_1(x) \sin 2\pi ux dx = c(u) - is(u).$$

Модуль этого комплексного числа, очевидно, равен $T(u)$, а аргумент – $\varphi(u)$. Функцию $T(u)$ называют одномерной частотно-контрастной характеристикой (ЧКХ), а $\varphi(u)$ – одномерной частотно-фазовой характеристикой (ЧФХ) оптической системы, или фазовой передаточной функцией.

Для выяснения физического смысла указанных выше характеристик введем также еще две передаточные характеристики оптической системы: коэффициент модуляции m , определив его отношением амплитуды I_1 гармонической составляющей к среднему значению \bar{I} :

$$m = I_1 / \bar{I}$$

и контраст K , определив его как

$$K = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \quad (12.25)$$

т.е. тем же выражением, что и видность V интерференционных полос на экране. Покажем, что контраст K в плоскости предмета определяется коэффициентом модуляции m . Действительно,

подставляя в (12.25) значения I_{\max} и I_{\min} , полученные из (12.23), будем иметь

$$K = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{(\bar{I} + I_1) - (\bar{I} - I_1)}{(\bar{I} + I_1) + (\bar{I} - I_1)} = \frac{I_1}{\bar{I}} = m.$$

Контраст изображения решетки

$$K' = \frac{I'_{\max} - I'_{\min}}{I'_{\max} + I'_{\min}} = \frac{T(u)I_1}{\bar{I}} = kT(u).$$

Коэффициент модуляции

$$m' = T(u) \frac{I_1}{\bar{I}} = mT(u).$$

Откуда

$$T(u) = \frac{K'}{K} = \frac{m'}{m}.$$

Таким образом, частотно-контрастная характеристика, т.е. функция $T(u)$ пространственной частоты u , характеризует отношение контраста изображения к контрасту самой решетки или отношение коэффициента модуляции изображения к коэффициенту модуляции объекта. ЧКХ обычно меньше единицы (кроме нулевой частоты, при которой она равна 1) и уменьшается с ростом частоты u . Примерная форма ЧКХ показана на рис. 12.12, а. Функция $T(u)$ определяет меру ослабления контраста изображения предмета для различных пространственных частот. По этой причине эту функцию и называют частотно-контрастной характеристикой оптической системы; ее называют также функцией передачи модуляции (ФПМ).

Частотно-контрастная характеристика является показателем эффективности системы формирования изображения на данной

пространственной частоте; она позволяет наилучшим образом оценить передачу мелких деталей объекта. Если известен коэффициент модуляции исходной решетки, то, умножая его на значение ЧКХ, соответствующее заданной пространственной частоте решетки, можно получить величину коэффициента модуляции для изображения решетки на этой пространственной частоте. Любой одномерный объект можно разложить в интеграл (или ряд) Фурье, т.е. представить его в виде совокупности конечного или бесконечного числа гармонических распределений (гармонических решеток). Это дает возможность рассчитать передачу такого объекта.

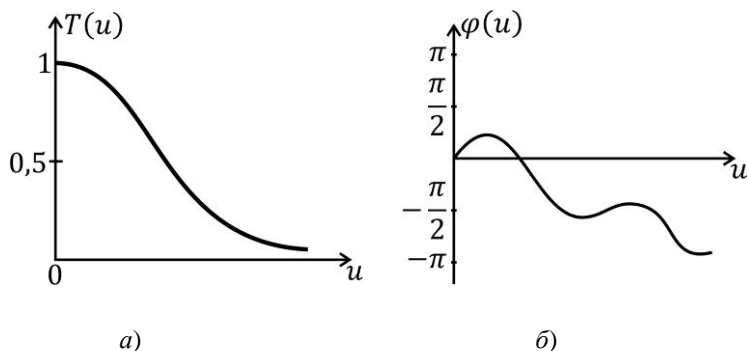


Рис. 12.12

Частотно-зависимые изменения фазы, вносимые оптической системой и характеризующиеся частотно-фазовой характеристикой $\varphi(u)$, проявляются на изображении в виде горизонтальных смещений гармонических компонент, составляющих изображение. Изображение решетки в общем случае не будет совпадать с ее положением, рассчитанным по законам оптики идеальной системы. Фазовая передаточная функция $\varphi(u)$ характеризует смещение фазы в изображении гармонической решетки при изменении пространственной частоты (примерный ее вид показан на рис. 12.12, б). Это смещение указывает на несоответствие координаты точки

изображения относительно соответствующей точки предмета. Линейное смещение косинусоиды Δx можно найти из того условия, что это смещение должно составлять, очевидно, такую же часть от периода, т.е. от $1/u$, какую фазовый угол сдвига $\varphi(u)$ составляет от 2π . Так что $\Delta x = \varphi(u) / (2\pi u)$. Для частоты $u = 0$ угол сдвига $\varphi(0) = 0$, а значит, и $\Delta x = 0$. В случае, когда функция рассеяния линии $F_1(x)$ является четной ($F_1(-x) = F_1(x)$), что всегда имеет место, например, для точки, расположенной на оптической оси, ввиду нечетности функции $\sin 2\pi ux$, а значит, и нечетности подынтегральной функции $F_1(x) \sin 2\pi ux$, интеграл $s(u) = 0$. В этом случае $\varphi(u) = 0$, т.е. смещение фазы будет отсутствовать, а одномерная ОПФ $A(u)$ будет величиной действительной и равной

$$c(u) = \int_{-\infty}^{\infty} F_1(x) \cos 2\pi u x dx .$$

При асимметричной функции рассеяния линии $F_1(x)$ (для оптических систем это может быть на краю поля зрения, при дефектах центровки и др.) ЧФХ для $u \neq 0$ может принимать различные значения, как показано на рис. 12.12, б

12.5. Сравнение формирования изображения при когерентном и некогерентном освещении

Как уже отмечалось, универсального критерия оценки качества оптического изображения, на основании которого можно было бы сравнить изображения при двух типах освещения, не существует. Поэтому сравнение изображений может касаться лишь каких-то отдельных его характеристик. Однако и при таком сравнении

можно выявить некоторые различия между двумя типами изображения.

Хотя при некогерентном освещении речь идет о распределении интенсивности на изображении, а при когерентном – о распределении комплексной амплитуды, в обоих случаях в конечном итоге наблюдается интенсивность. По этой величине – интенсивности изображения – и можно сравнить два типа изображения.

Одной из характеристик интенсивности изображения, по которой можно произвести сравнение в обоих случаях, является частотный спектр. В случае некогерентного освещения система является линейной по интенсивности; распределение интенсивности на изображении I' представляется сверткой $I' = |h|^2 * I$ или, учитывая, что $I \sim |E|^2$,

$$I' = |h|^2 * |E|^2, \quad (12.26)$$

где I и E – соответственно распределение интенсивности и комплексной амплитуды на предмете. При когерентном освещении система, будучи линейной по комплексной амплитуде, не является таковой в отношении интенсивности. В этом случае

$$I' \sim |E'|^2 = |h * E|^2. \quad (12.27)$$

Чтобы найти частотный спектр интенсивности изображения, следует произвести фурье-преобразование над обеими частями равенств (12.26) и (12.27) и использовать теорему свертки. Обозначив символом \otimes операцию автокорреляции

$$X(u, v) \otimes X(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X(u', v') X(u' + u; v' + v) du' dv',$$

частотный спектр интенсивности изображения для двух случаев можно записать следующим образом:

$$I'(u, v) \sim [H(u, v) \otimes H(u, v)] \cdot [E(u, v) \otimes E(u, v)]$$

– при некогерентном освещении,

$$I'(u, v) \sim H(u, v) E(u, v) \otimes H(u, v) E(u, v)$$

– при когерентном освещении.

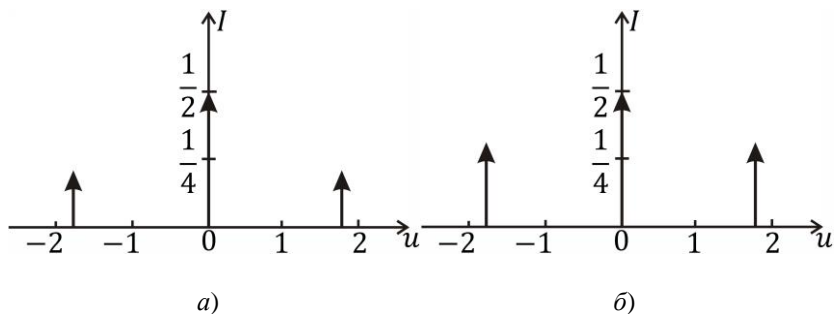


Рис. 12.13

Эти соотношения показывают, что частотные спектры интенсивности двух типов изображения не являются тождественными. Покажем, что в зависимости от распределения интенсивности и фазы в плоскости предмета частотные спектры могут отличаться друг от друга. С этой целью рассмотрим два предмета A и B с одинаковыми функциями пропускания по интенсивности $T(x)$, но с различными распределениями фазы. Пусть функция $T(x)$ имеет вид $T(x) = \cos^2 2\pi i x$, причем пространственная частота $u_0 / 2 \ll u < u_0$, где u_0 – предельная частота при когерентном освещении. Функция пропускания по амплитуде для двух предметов пусть будет $t(x) = \cos 2\pi i x$ – для предмета A и $t(x) = |\cos 2\pi i x|$ – для предмета B , т.е. предметы отличаются только периодическим распределением фазы. На рис. 12.13 показаны результаты расчета спектра интенсивности I изображения для предмета A при некогерентном (рис. 12.13, a) и когерентном (рис. 12.13, b) освещении. Как видим, контраст распределения интенсивности изображения при некогерентном освещении хуже, чем при когерентном. Следовательно, можно

считать, что для предмета A лучше когерентное освещение. Что касается предмета B , то его амплитуда представляет собой периодическую функцию с частотой основной гармоники $2u$. Но так как $2u > u_0$, то в случае когерентного освещения интенсивность изображения меняться не будет, тогда как некогерентная система формирует то же самое изображение, что и для предмета A . Следовательно, для предмета B лучше некогерентное освещение.

Рассмотренный выше пример показывает, что в отношении спектрального состава изображения нельзя однозначно сказать, какой тип освещения предпочтительнее. Результат в значительной степени зависит от структуры предмета и, в частности, от распределения фазы в точках предмета. Нельзя утверждать, что один вид освещения предпочтительнее во всех случаях.

Другой метод сравнения изображения при различных видах освещения основан на способности оптической системы различать два близко расположенных точечных источника света, т.е. на разрешающей способности оптической системы.

Согласно критерию Рэля, два некогерентных точечных источника

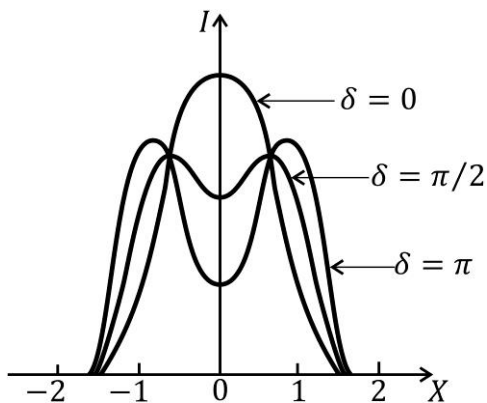


Рис. 12.14

разрешаются оптической системой, если центр диска Эри, созданного одним источником, совпадает с краем диска Эри, созданного другим источником. Минимальное разрешаемое расстояние между идеальными изображениями равно радиусу диска Эри $r_1 =$

$$1,22\lambda f / D.$$

Выясним, при каком освещении, когерентном или некогерентном, легче различать два точечных источника, разделенных расстоянием Эри r_1 . Предположим, что источники когерентны. Пусть $E_1(x)$ и $E_2(x)$ – комплексные амплитуды полей, создаваемых в точке с координатой x первым и вторым источником соответственно. Распределение интенсивности в этом случае будет определяться как

$$I(x) \sim \left| E_1(x) + E_2(x)e^{i\delta} \right|^2 = \\ = |E_1(x)|^2 + |E_2(x)|^2 + 2\operatorname{Re}\left[E_1(x)E_2^*(x)e^{-i\delta} \right],$$

где третий член представляет собой сумму двух комплексно сопряженных чисел $E_1(x)E_2^*(x)e^{-i\delta}$ и $E_1^*(x)E_2(x)e^{i\delta}$, δ – разность фаз волн, излучаемых двумя источниками. На рис. 12.14 приведены распределения интенсивности изображения двух точечных источников при трех значениях разности фаз δ , равных 0 , $\pi / 2$ и π . Мы видим, что распределение интенсивности в изображении для когерентных источников с разностью фаз $\delta = \pi / 2$ ($I(x) = I_1(x) + I_2(x)$) тождественно распределению интенсивности для некогерентных точечных источников (рис. 12.1). При разности фаз $\delta = 0$ провал на кривой интенсивности изображения отсутствует (имеет место интерференционный максимум, когда $I(x) = I_1(x) + I_2(x) + 2\sqrt{I_1(x)I_2(x)}$) и, следовательно, две точки труднее различать, чем в случае некогерентного освещения. Наконец, если точечные источники излучают в противофазе ($\delta = \pi$), то провал на кривой $I(x)$ оказывается больше, чем при некогерентном освещении (имеет место интерференционный минимум, когда $I(x) = I_1(x) + I_2(x) - 2\sqrt{I_1(x)I_2(x)}$). В этом случае две точки различить легче, чем при некогерентном освещении.

Таким образом, и с помощью разрешающей способности нельзя сделать вывод о том, какой вид излучения для

формирования изображения предпочтительнее. И в этом случае результат зависит от распределения фазы, связанной с предметом.

Сравнение формирования изображения при двух типах освещения можно произвести и по отклику некогерентной и когерентной систем на возмущение, распространяющееся от резко отграниченного края (т.е. по распределению интенсивности в изображении края полуплоскости). На рис. 12.15 снова воспроизведен график распределения интенсивности в изображении края полуплоскости, создаваемой системой с прямоугольным зрачком при некогерентном освещении. Пунктирной кривой показано распределение той же интенсивности, соответствующее когерентному освещению. Видно, что эти распределения существенно отличаются друг от друга. В случае когерентного освещения резко выражены осцилляции интенсивности. Распределение интенсивности при дифракции от края при когерентном освещении имеет резкие переходы, тогда как изменения функции рассеяния края (ФРК) происходят более плавно.

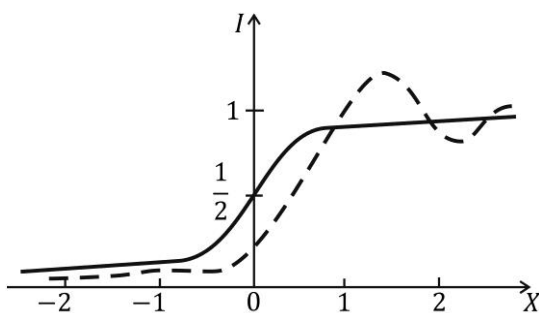


Рис. 12.15

Различие между изображениями при когерентном и некогерентном освещении проявляется и в эффекте зернистости, который наблюдается при высококогерентном освещении. Размер отдельных зерен приблизительно равен размеру изображения разрешаемого элемента. Эффект зернистости ухудшает качество изображения.

Видно также, что кажущиеся положения края, т.е. абсциссы точек, в которых интенсивность равна половине максимальной, различны для двух видов освещения, причем когерентное освещение вносит незначительную погрешность.

Следует отметить также и то, что освещение с высокой степенью когерентности чувствительно к оптическим дефектам, которые могут встретиться на пути света. Такими дефектами могут являться, например, мельчайшие частички пыли на линзе объектива. Дифракция света на этих дефектах приводит к образованию дифракционных колец, которые накладываются на изображение и ухудшают его качество. В то же время когерентное освещение дает больше возможностей для применения оптических систем в качестве систем обработки информации.

12.6. Изображение при частично когерентном квазимонохроматическом освещении

Выше мы рассмотрели формирование изображения в двух предельных случаях, когда изображаемый предмет освещался полностью когерентным и полностью некогерентным светом. В первом случае рассматривалось распространение через оптическую систему комплексной амплитуды, а во втором – интенсивности. Поскольку на практике обычно приходится иметь дело с источниками квазимонохроматического света, представляет интерес рассмотреть и более общий случай частично когерентного освещения квазимонохроматическим светом. Передаваемым оптической системой сигналом в этом случае будет взаимная интенсивность $I(x_1, y_1; x_2, y_2)$ для точек (x_1, y_1) и (x_2, y_2) . Напомним, что взаимной интенсивностью называют величину $\langle E_1(x_1, y_1, t) \cdot E_2^*(x_2, y_2, t) \rangle$, где $E_1(x_1, y_1, t)$ и $E_2(x_2, y_2, t)$ – напряженности поля в точках (x_1, y_1) и (x_2, y_2) соответственно, измеренные в один и тот же момент времени t .

Пусть $I(x_1, y_1; x_2, y_2)$ – взаимная интенсивность для произвольных точек (x_1, y_1) и (x_2, y_2) предмета, а $h(x, y; x', y')$ – импульсная характеристика (функция пропускания) оптической системы. Оптическая система предполагается изопланарной. Тогда взаимная интенсивность света для соответствующих точек

(x'_1, y'_1) и (x'_2, y'_2) плоскости изображения, выходящего из точек (x_1, y_1) и (x_2, y_2) предмета, найдется как

$$I'(x'_1, y'_1; x'_2, y'_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x_1, y_1; x_2, y_2) h(x'_1 - x_1; y'_1 - y_1) \times$$

$$\times h^*(x'_2 - x_2; y'_2 - y_2) dx_1 dy_1 dx_2 dy_2. \quad (12.28)$$

Интегрирование в (12.28) распространено на бесконечные пределы, так как вне предмета взаимная интенсивность $I(x_1, y_1; x_2, y_2) = 0$. Соотношение (12.28) показывает, что при частично когерентном освещении предмета квазимонохроматическим светом оптическая система является линейной по взаимной интенсивности.

Взяв фурье-преобразование от обеих частей равенства (12.28) и используя теорему свертки, получим

$$I'(u, v; u', v') = I(u, v; u', v') M(u, v; u', v'), \quad (12.29)$$

где $M(u, v; u', v')$ – преобразование Фурье произведения $h(x_1, y_1)h^*(x_2, y_2)$. Функцию $M(u, v; u', v')$ называют передаточной функцией или функцией частотного отклика оптической системы для частично когерентного квазимонохроматического освещения. Эта функция определяет влияние оптической системы на пространственный спектр сигнала, прошедшего через оптическую систему при частично когерентном квазимонохроматическом освещении. Соотношение (12.29) показывает, что если взаимную интенсивность в плоскости предмета и изображения представить суперпозицией четырехмерных пространственных гармоник всевозможных пространственных частот $(u, v; u', v')$, то каждая такая компонента взаимной интенсивности в изображении будет определяться соответствующей компонентой в предмете и соответствующим значением частотного отклика оптической системы. Как видим, влияние оптической системы на взаимную

интенсивность эквивалентно действию четырехмерного линейного фильтра.

Величина $h(x_1, y_1)h^*(x_2, y_2)$ равна обратному фурье-преобразованию от функции $M(u, v; u', v')$:

$$h(x_1, y_1)h^*(x_2, y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} M(u, v; u', v') \times \\ \times \exp[i2\pi(ux_1 + vy_1 + u'x_2 + v'y_2)] dudvdu'dv'. \quad (12.30)$$

Запишем также обратные фурье-преобразования функций $h(x_1, y_1)$ и $h^*(x_2, y_2)$:

$$h(x_1, y_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(u, v) \exp[-i2\pi(ux_1 + vy_1)] dudv, \\ h^*(x_2, y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H^*(u', v') \exp[-i2\pi(u'x_2 + v'y_2)] du'dv'.$$

Произведя в последнем интеграле замену переменных $u' \rightarrow -u'$, $v' \rightarrow -v'$, приведем его к виду

$$h^*(x_2, y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H^*(-u', -v') \exp[i2\pi(u'x_2 + v'y_2)] du'dv'.$$

Здесь $H(u, v)$ – передаточная функция когерентной оптической системы. Подставляя выражения $h(x_1, y_1)$ и $h^*(x_2, y_2)$ в левую часть соотношения (12.30) и поменяв затем местами левую и правую части, получим

$$M(u, v; u', v') = H(u, v) H^*(-u', -v').$$

Это соотношение связывает частотный отклик оптической системы при частично когерентном освещении с передаточной функцией системы при когерентном освещении.

Функцию частотного отклика можно связать и с функцией зрачка оптической системы. Величина $H(u, v)$ равна значению функции зрачка $P(\xi, \eta)$ в точке с координатами $\xi = \lambda l u$, $\eta = \lambda l v$. С учетом этого находим

$$M(u, v) = P(\lambda l u, \lambda l v) P^*(-\lambda l u', -\lambda l v')$$

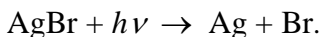
или

$$M\left(\frac{\xi}{\lambda l}, \frac{\eta}{\lambda l}, \frac{\xi'}{\lambda l}, \frac{\eta'}{\lambda l}\right) = P(\xi, \eta) P^*(-\xi', -\eta').$$

Это и есть формулы, связывающие частотный отклик частично когерентной системы с функцией ее зрачка. Здесь ξ, η, ξ', η' – координаты в плоскости входного зрачка системы. Так как функция зрачка оптической системы отлична от нуля только в пределах апертуры зрачка, то оптическая система пропустит только те спектральные компоненты, которые соответствуют пространственным частотам, не превышающим некоторых значений. Например, в случае оптической системы с круглым входным зрачком величина $P(\xi, \eta) P^*(-\xi', -\eta')$ отлична от нуля только в пределах круга радиуса, равного радиусу зрачка R , т.е. когда $\xi^2 + \eta^2 \leq R^2$ или $\xi'^2 + \eta'^2 \leq R^2$. Следовательно, рассматриваемая оптическая система пропустит спектральные компоненты взаимной интенсивности, соответствующие частотам $(u, v; u', v')$, удовлетворяющим условию $u^2 + v^2 \leq (R/\lambda R)^2$ или $u'^2 + v'^2 \leq (R/\lambda R)^2$, где λ – средняя длина волны в пространстве изображения. Как видим, максимальные пространственные частоты, пропускаемые частично когерентной оптической системой с круглым зрачком, как и такой же, но когерентной системой, равны $R/\lambda l$.

12.7. Фотографическое отображение сигналов

Запись сигнала можно получить путем воздействия света на соответствующий материал. Наиболее распространенным таким материалом является фотографический материал в виде фотопластинки или фотопленки, который представляет собой светочувствительный слой, нанесенный на стеклянную или гибкую подложку. Светочувствительный слой представляет собой суспензию мелких зерен галогенидов серебра (обычно бромид серебра AgBr) в нейтральном коллоиде (обычно в желатине). При освещении бромида серебра светом видимого диапазона происходит его фотолитическая диссоциация согласно уравнению



При малых временах экспозиции в отдельных зернах бромида серебра образуются только зародыши серебра, которые при дальнейшей химической обработке становятся каталитическими центрами процесса проявлений; при этом масса серебра увеличивается. Под экспозицией H (количеством освещения) понимают произведение интенсивности света I , падающего на светочувствительную среду, на длительность воздействия τ (длительность экспозиции):

$$H = I\tau.$$

Так определяется экспозиция в случае стационарного поля. Когда производится голографическая запись изменяющегося во времени поля, то экспозиция определяется соотношением

$$H = \int_{t_1}^{t_2} I(t) dt.$$

При освещении фотослоя образуется скрытое изображение сигнала, которое после фотохимического проявления преобразуется в почернение фотопленки.

Информацию, содержащуюся в изображении, можно записать либо поточно, раскладывая изображение на малые элементы и

последовательно нанося их на регистрирующий материал, либо интегрально на плоскости, либо в объемной регистрирующей среде. В классических методах используются первые два типа записи. Для записи изображения в голографии регистрирующая среда должна быть плоской или объемной.

Одной из наиболее употребительных характеристик фотопленки является коэффициент пропускания (по интенсивности), определяемый как отношение интенсивности света I , прошедшего через фотопленку к интенсивности света I_0 , падающего на фотопленку в каждой точке проявленной пленки:

$$T = I / I_0.$$

Очевидно, что $0 \leq T \leq 1$. Прозрачность (коэффициент пропускания по интенсивности) может быть различна в разных точках пленки, определяемых координатами x, y , т.е. представляет собой функцию $T(x, y)$ этих координат. Согласно закону Бугера интенсивность на выходе фотопленки

$$I = I_0 e^{-Kd},$$

где K – коэффициент поглощения фотопленки, d – толщина фотопленки. Это соотношение справедливо в предположении постоянства коэффициента поглощения вдоль толщины слоя. В этом предположении коэффициент пропускания

$$T = e^{-Kd}.$$

При обычных способах записи сигнала толщина d регистрирующей среды выбирается с учетом резкости изображения.

Другой наиболее употребительной характеристикой фотоматериала является оптическая плотность почернения D , определяемая как десятичный логарифм обратного коэффициента пропускания:

$$D = \lg \frac{1}{T} = -\lg T.$$

При когерентном освещении, наряду с коэффициентом пропускания по интенсивности T , вводится также амплитудный коэффициент пропускания $t = E / E_0$, где E_0 – распределение комплексной амплитуды в плоскости, непосредственно примыкающей к фотопластинке, E – распределение комплексной амплитуды в плоскости непосредственно за фотопластинкой. Амплитудный коэффициент пропускания является комплексной функцией координат x, y на поверхности пленки, определяющей изменение амплитуды и фазы световой волны при прохождении фотопленки, поэтому его можно представить в виде

$$t(x, y) = |t(x, y)| e^{i\varphi(x, y)}. \quad (12.31)$$

Модуль $|t(x, y)|$ ($0 \leq |t| \leq 1$) в выражении (12.31) называют амплитудной характеристикой фотопленки (она характеризует прозрачность пленки), аргумент комплексной величины $t(x, y)$, т.е. функцию $\varphi(x, y)$, называют фазовой характеристикой пленки. При прохождении когерентного света через экспонированную, проявленную и закрепленную фотопленку распределение амплитуд и фаз световой волны в различных точках непосредственно у задней поверхности фотопленки определяется функцией $t(x, y)$, поэтому эту функцию называют модуляционной характеристикой фотопленки. Следовательно, модуляционная характеристика определяет зависимость амплитудного коэффициента пропускания t от координат точки на пленке (x, y) . Фотопленка с записанным на ней сигналом может использоваться в качестве модулирующего экрана (транспаранта) в когерентных оптических системах обработки информации.

Коэффициенты T и t связаны между собой соотношением $T = |t|^2$. Это соотношение позволяет найти модуль амплитудного

коэффициента пропускания $|t|$ по известному коэффициенту пропускания по интенсивности T . Непосредственно измерить амплитудную характеристику нельзя, так как детекторы светового излучения реагируют не на амплитуду световой волны, а на квадрат ее модуля, т.е. на величину $|E|^2$. Так как $|t|$ и T – положительные величины, не превосходящие единицу, то $|t| = \sqrt{T}$ или (с учетом определения величины D)

$$|t| = 10^{-D/2}.$$

Фазовая характеристика фотопленки $\varphi(x, y)$ определяется величиной фазового набега световой волны при прохождении фотопленки. При угле падения света, близком к нормальному, фазовый набег определится как

$$\varphi(x, y) = \frac{2\pi}{\lambda_0} \Delta(x, y) = \frac{2\pi d(x, y)}{\lambda_0} [n(x, y) - 1], \quad (12.32)$$

где $\Delta(x, y) = d(x, y)[n(x, y) - 1]$ – изменение оптического пути светового луча в пленке, $d(x, y)$ – толщина фотопленки в точке (x, y) , $n(x, y)$ – показатель преломления фотопленки в точке (x, y) .

Все введенные характеристики светочувствительной среды зависят от экспозиции H . Под действием света происходит также и изменение комплексного показателя преломления $N = n + in'$, действительная часть которого n представляет собой обычный показатель преломления, а мнимая n' определяет поглощение света в веществе.

При регистрации света с помощью фотоэмульсии под действием световой энергии происходит изменение поглощения материала, вследствие чего прозрачность светочувствительного слоя определяется экспозицией. Такая запись называется амплитудной. С изменением коэффициента поглощения

происходит изменение и действительной части комплексного показателя преломления, так как вблизи полосы поглощения имеет место аномальная дисперсия. Если под действием света происходит изменение только показателя преломления, то запись называется фазовой.

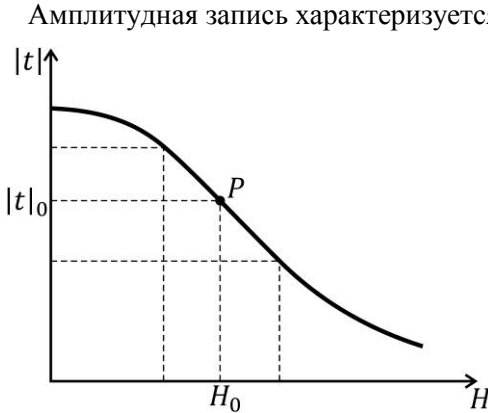


Рис. 12.16

нением амплитудного коэффициента пропускания $|t|$. Эти изменения определяются экспозицией. Считая в рабочей области (рис. 12.16) эту зависимость линейной, аппроксимируем ее выражением

$$|t| = |t|_0 + \left(\frac{d|t|}{dH} \right)_{|t|_0} (H - H_0).$$

Кривая, изображенная на рис. 12.16, называется амплитудной характеристикой записи сигнала на фотопленку, а точка $P(|t|_0, H_0)$ – рабочей точкой. Производная $\left(\frac{d|t|}{dH} \right)_{|t|_0}$ определяет крутизну

характеристики вблизи рабочей точки. Эта производная отрицательна, поскольку большему количеству освещения соответствует меньшее пропускание. Обозначив ее через β_1 , можно записать

$$|t| = |t|_0 + \beta_1(H - H_0)$$

или

$$|t| = |t|_0 + \beta(I - I_0), \quad (12.33)$$

где $\beta = \beta_1 \tau$. Для обеспечения линейности амплитудной характеристики системы записи необходимо применять специальные меры.

В соответствии с формулой (12.32) фазовую запись можно осуществить как путем пространственной модуляции коэффициента преломления $n(x, y)$, так и путем модуляции толщины пленки $d(x, y)$. Изменение толщины фотопленки обусловлено (в пренебрежении изменением толщины основы, имеющим случайный характер) изменением толщины фотоэмульсионного слоя, которое обусловлено изменением оптической плотности фотоэмульсии. При проявлении фотопленки в местах большей экспозиции откладывается больше металлического серебра, что и приводит к образованию на поверхности фотопленки рельефа, пропорционального записываемому сигналу. Для мелкозернистых эмульсий зависимость между оптической плотностью D и толщиной пленки d (в мкм) является линейной. Действительно, для мелкозернистых эмульсий коэффициент поглощения K от толщины эмульсии не зависит, поэтому

$$D = -\lg T = -\lg e^{-Kd} = Kd \lg e.$$

Откуда

$$d = D/(K \lg e) = \mu D,$$

где $\mu = (K \lg e)^{-1}$ — коэффициент пропорциональности. Ограничиваясь случаем линейного отклика фотопленки, т.е. линейной зависимостью фазы от количества освещения (экспозиции), можно записать

$$\varphi = \varphi_0 + \left(\frac{d\varphi}{dH} \right)_{\varphi_0} (H - H_0). \quad (12.34)$$

Это соотношение выражает собой так называемую фазовую характеристику записи сигнала. Объединяя выражение (12.34) с выражением (12.31), для случая небольших изменений среды (т.е. в линейном приближении) получим

$$t = |t| \left[1 + i \left(\frac{d\varphi}{dH} \right)_{\varphi_0} (H - H_0) \right] e^{i\varphi_0},$$

где показательную функцию $\exp \left[i \left(\frac{d\varphi}{dH} \right)_{\varphi_0} (H - H_0) \right]$ заменили

первыми двумя членами разложения в ряд.

На рис. 12.17 показана кривая зависимости оптической плотности D от десятичного логарифма экспозиции H для области,

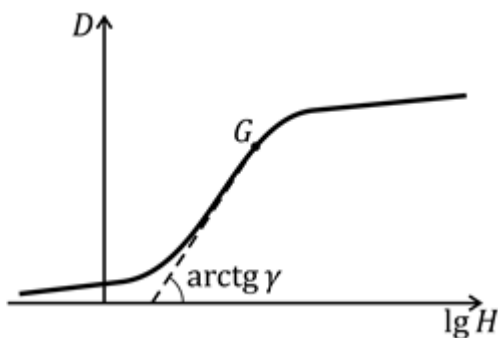


Рис. 12.17

важной с точки зрения практического применения. Эта кривая называется характеристической кривой (или кривой почернения) фотопленки. Она определяет амплитудную характеристику системы записи. На

рис. 12.17 видно, что начальный участок характеристической кривой имеет нелинейный характер, а затем (в окрестности точки перегиба G) зависимость D от $\lg H$ становится почти линейной. При дальнейшем увеличении экспозиции скорость нарастания оптической плотности уменьшается, а затем наблюдается даже

некоторое снижение оптической плотности. Участок кривой в окрестности точки перегиба G можно аппроксимировать прямой

$$D = \gamma \lg H + C,$$

где γ и C – постоянные коэффициенты. Угловым коэффициентом прямой γ называется коэффициент контрастности; он зависит от свойств фотоматериала, времени и способа его обработки. Линейная аппроксимация предполагает, что регистрируемый сигнал преобразуется в пространственное изменение амплитудного коэффициента пропускания без искажения. Например, гармонический сигнал преобразуется в гармонический же сигнал. Принимая во внимание связь $|t|$ с D , можно записать

$$|t| = 10^{-(\gamma \lg H + C)/2} = k_1 H^{-\gamma/2}$$

или

$$|t| = k I^{-\gamma/2},$$

где $k_1 = 10^{-C/2} = e^{-C \ln 10 / 2}$ – характерная для каждого фотоматериала определенная постоянная, $k = k_1 \tau^{-\gamma/2}$. Полученные зависимости также называются амплитудными характеристиками записи сигнала. Рассчитанная по этой формуле амплитудная характеристика записи приведена на рис. 12.17. Прямолинейный участок в окрестности точки перегиба G соответствует рабочему участку характеристической кривой.

Для практических целей наиболее оптимальным оказывается коэффициент контрастности $\gamma = -2$. В этом случае амплитудный коэффициент пропускания пропорционален количеству освещения: $|t| = k_1 H$.

При записи сигнала на фотопленку амплитудный коэффициент пропускания $t(x, y)$ не повторяет в точности закон изменения амплитуды сигнала $E(x, y)$. Для линейной системы записи эти величины связаны между собой преобразованием свертки

$$t(x, y) = \int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} E(\xi, \eta) h_n(x - \xi; y - \eta) d\xi d\eta,$$

где интегрирование ведется по плоскости пленки. Функция $h_n(x, y; \xi, \eta)$ называется импульсной переходной функцией фотопленки. Таким образом, фотопленку можно характеризовать импульсной переходной функцией. Фотопленку можно характеризовать и амплитудно-частотной характеристикой $H_n(u, v)$, которая представляет собой зависимость амплитуды гармонических сигналов, записанных на фотопленке, от пространственных частот u и v . Она связана преобразованием Фурье с импульсной переходной функцией:

$$H_n(u, v) = \int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} h_n(x, y) \exp[-i2\pi(ux + vy)] dx dy.$$

Импульсная переходная функция фотопленки $h_n(x, y)$ обладает свойством симметрии, поэтому ее фурье-спектр $H_n(u, v)$ является действительной величиной.

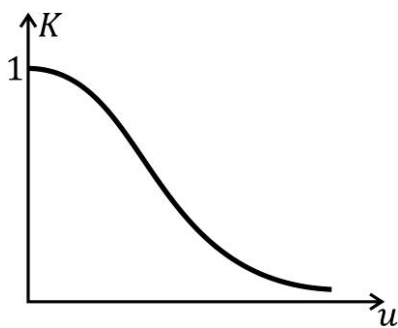


Рис. 12.18

Свойства переходных характеристик пленки определяются в основном рассеянием света в эмульсии и действием местных фотографических эффектов при проявлении.

Еще одной важной характеристикой регистрирующего материала, особенно в отношении использования его в голографии, является частотно-контрастная характеристика (ЧКХ), выражающая собой зависимость контраста полос от пространственной частоты.

Частотно-контрастная характеристика $K(u)$ регистрирующего материала определяется зависимостью контраста амплитудного пропускания материала от пространственной частоты регистрируемой синусоидальной картины. Пусть распределение интенсивности в регистрируемой картине описывается выражением

$$I(u) = I_0 + I_1(u) \cos 2\pi ux,$$

где I_0 – среднее значение интенсивности, I_1 – амплитуда колебания интенсивности, u – частота гармонического распределения. Контраст полос, зарегистрированных фотоматериалом, определяется как отношение

$$K_I(u) = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}.$$

Подставив сюда максимальное и минимальное значения интенсивности, получим $K_I = I_1 / I_0$. Амплитудное пропускание имеет вид

$$t(u) = t_0 + t_1(u) \cos 2\pi ux,$$

а контраст

$$K_t(u) = \frac{t_{\max} - t_{\min}}{t_{\max} + t_{\min}} = \frac{t_1(u)}{t_0}.$$

Частотно-контрастная характеристика регистрирующего материала определяется как отношение

$$K(u) = K_t(u) / K_I(u).$$

Типичная форма ЧКХ регистрирующего материала представлена на рис. 12.18. Как видно на этом рисунке, с ростом частоты u кривая монотонно убывает от максимального значения, равного единице, при нулевой частоте, до минимального значения, равного нулю.

12.8. Оптическая обработка пространственной информации

Возможность применения оптических систем в целях обработки информации заключается в следующем. Если необходимая для обработки информация зашифрована в виде двумерного распределения коэффициента пропускания $t(x, y)$ транспаранта, который устанавливается во входную плоскость системы, то с помощью специально подобранных фильтров с заданным распределением коэффициента пропускания можно на выходе получить сигнал, представляющий результат воздействия фильтра на входной сигнал.

Рассмотрим два вида обработки пространственной информации: пространственную фильтрацию и восстановление искаженных изображений.

12.8.1. Пространственная фильтрация

Вообще говоря, пространственная фильтрация осуществляется практически в любой оптической системе, так как в любом случае ограниченные размеры входного зрачка оптической системы приводят к пространственной фильтрации, ибо составляющие пространственного спектра, не попадающего в площадь апертуры, отфильтровываются на входе. Здесь под пространственной фильтрацией понимается целенаправленные воздействия на пространственные составляющие спектра сигнала. Пространственная фильтрация в когерентном свете основывается на том, что линза выполняет преобразование Фурье двумерного оптического сигнала. Если предмет с распределением комплексной амплитуды во входной плоскости $E(x, y)$ поместить в переднем фокусе линзы и осветить плоской монохроматической волной, то, как было показано в п. 10.6, в ее задней фокальной плоскости образуется фурье-образ этого распределения: $E(u, v) =$

$F\{E(x, y)\}$ (с коэффициентом перед интегралом $1/(\lambda f)$), где пространственные частоты $u = x_1/(\lambda f)$, $v = y_1/(\lambda f)$; x_1 , y_1 – координаты в задней фокальной плоскости линзы, f – фокусное расстояние линзы. Плоская монохроматическая волна формируется точечным монохроматическим источником, помещенным в передний фокус линзы, расположенной перед предметом.

В основе метода когерентной пространственной фильтрации светового поля с целью воздействия на изображение, даваемое оптической системой, лежит описанный в п. 11.7 опыт Аббе – Портера, в котором путем введения различных диафрагм и экранов в фокальную плоскость линзы, строящей изображение предмета, достигалось изменение структуры его изображения.

Любой объект можно представить в виде совокупности элементарных гармонических амплитудных решеток. На каждой такой решетке возникает дифракция света, в результате которой в фокальной плоскости линзы образуются три пятна, соответствующие сфокусированным дифрагировавшим пучкам нулевого и двух первых порядков. В зависимости от величины пространственной частоты решетки первые дифракционные порядки удалены от нулевого порядка на большее или меньшее расстояние. Поэтому каждой точке фокальной плоскости соответствует одна элементарная гармоническая решетка предмета. Помещая в фокальную плоскость маску с отверстиями, можно отфильтровывать определенные элементарные решетки, а тем самым и соответствующие элементы в изображении.

Как отмечалось в п. 11.7, крупные детали объекта, которым соответствуют низкие частоты, дают дифрагировавшие пучки мало отклоненные, а мелкие, которым соответствуют высокие частоты, – сильно отклоненные от направления прямого пучка. Поэтому если поместить в фокальную плоскость экран с небольшим отверстием в центре, то он пропустит только свет, дифрагировавший на крупных деталях объекта, а если, наоборот, экранировать центральный участок фокальной плоскости, то дифрагировавшие лучи, соответствующие крупным деталям,

будут задерживаться, а пучки, дифрагировавшие на мелких деталях, будут проходить. (Во втором случае возникает впечатление увеличения резкости фотоснимка.) Таким образом, с помощью экрана, помещенного в фокальную плоскость объектива, можно осуществить фильтрацию пространственных частот объектов и тем самым произвести воздействие на структуру его изображения.

Вообще с помощью масок, помещаемых в частотную плоскость системы фильтрации, можно производить следующие виды пространственной фильтрации: 1) низкочастотная фильтрация, позволяющая ослабить шум, устранить растр на изображении, осуществить операцию сглаживания; 2) высокочастотная фильтрация, позволяющая повысить контраст изображения, но ухудшить его разрешение, выделить контуры на изображении; 3) полосовая фильтрация, позволяющая выявить скрытые периодические структуры на изображении, или структуру заданного размера; 4) направленная фильтрация, позволяющая выделить линии с требуемой пространственной ориентацией; 5) мультипликация изображений при помощи одномерных и двумерных решеток.

В соответствии с опытами Аббе в качестве двумерного фильтра нижних частот можно использовать бесконечный экран с вырезанным прямоугольным отверстием (прямоугольная диафрагма), помещенный в заднюю фокальную плоскость линзы. Если размеры отверстия $a \times b$, то граничные частоты будут равны $u_0 = a/(2\lambda f)$, $v_0 = b/(2\lambda f)$. (Эти значения получаются из равенств (11.25), если фигурирующее в них расстояние l от плоскости зрочка до плоскости изображения заменить на фокусное расстояние линзы f .) Следовательно, этот фильтр будет пропускать пространственные частоты, заключенные в прямоугольнике со сторонами $a/(2\lambda f)$, $b/(2\lambda f)$, и не пропускать более высокие частоты. Фильтр верхних частот, наоборот, должен не пропускать колебаний низких частот и пропускать колебания высоких частот. В оптике таким фильтром является прямоугольный экран, помещенный в фокальную

плоскость линзы. Подобным образом могут быть построены и более сложные фильтры.

При математическом описании метода оптической фильтрации вводят импульсную характеристику пространственного фильтра $h(x, y)$ и исходят из того, что с точки зрения передачи информации оптическая система является линейной и пространственно инвариантной, т.е. отклик $E_1(x, y)$ на какой-либо сигнал $E(x, y)$ на входе системы описывается сверткой

$$E_1(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(\xi, \eta) h(x - \xi, y - \eta) d\xi d\eta. \quad (12.35)$$

Импульсная характеристика пространственного фильтра, как и импульсная характеристика изображающей системы, описывает изображение, формируемое в плоскости изображения, когда в предметной плоскости помещен точечный источник, описываемый δ -функцией. Она определяется фильтрующими свойствами системы.

С помощью теоремы свертки выражение (12.35) можно преобразовать к виду

$$E_1(u, v) = E(u, v)H(u, v), \quad (12.36)$$

где $H(u, v)$ – функция, описывающая заданный пространственно-частотный фильтр, называемая модуляционной характеристикой пространственного фильтра или функцией пропускания фильтра и являющаяся фурье-образом его импульсной характеристики $h(x, y)$, $H(u, v) = F\{h(x, y)\}$. Отсюда следует, что в случае оптической системы, осуществляющей пространственную фильтрацию, модуляционная характеристика фильтра $H(u, v)$ есть передаточная функция системы, импульсная характеристика которой $h(x, y)$. Подбирая пространственные фильтры с соответствующими модуляционными характеристиками, можно получить нужную передаточную функцию оптической системы и тем самым синтезировать оптическую систему, осуществляющую

требуемое преобразование оптического сигнала: $E(x, y) \rightarrow E_1(x_1, y_1)$.

Соотношение (12.36) описывает общий процесс фильтрации. Поместив в фокальной плоскости линзы транспарант с функцией пропускания $t(x, y) = H(kx/f, ky/f)$, можно осуществить фильтрацию пространственных частот объекта.

Функция пропускания фильтра $H(u, v)$ в общем случае является комплексной. Это означает, что пространственно-частотный фильтр может преобразовывать как амплитуду, так и фазу спектра сигнала. Если фильтр преобразует только амплитуду спектра, он называется амплитудным, а если только фазу – фазовым. С помощью амплитудного фильтра, установленного в плоскости пространственных частот (фокальной плоскости линзы), можно устранить или ослабить некоторые частоты передаваемого изображения или некоторых его деталей, а с помощью фазового фильтра – изменить фазы некоторых гармоник его пространственного спектра. В соответствии с соотношением (12.36), для проведения операции фильтрации требуется перемножить частотный спектр фильтруемого сигнала с частотным спектром фильтра – некоторой заданной функции (заданного импульсного отклика).

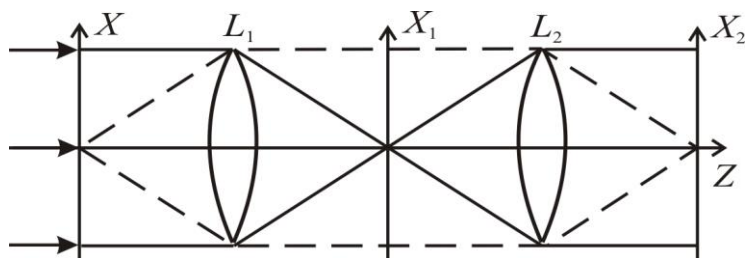


Рис. 12.19

Пространственная фильтрация в когерентном свете осуществляется по методу двойной дифракции или, иначе, по методу двойного преобразования Фурье. Классическая двухлинзовая схема (называемая схемой $4f$) показана на рис.

12.19. Объект (транспарант с записанным изображением предмета) помещается в передней фокальной плоскости первой линзы L_1 (в координатной плоскости X_1Y_1) и освещается плоской монохроматической волной. Линза L_1 осуществляет фурье-преобразование входного сигнала $E(x, y) = E_0 t(x, y)$, где $t(x, y)$ – функция пропускания транспаранта, E_0 – амплитуда плоской волны, освещающей транспарант, давая в ее задней фокальной плоскости (координатной плоскости X_1Y_1) распределение комплексной амплитуды $\frac{1}{\lambda f} E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right)$. Это фурье-

преобразование есть не что иное, как картина фраунгоферовской дифракции на транспаранте. Следовательно, при первом процессе дифракции, происходящем на апертуре первой линзы L_1 , образуется фурье-спектр оптического сигнала, расположенного в плоскости X_1Y_1 (с коэффициентом $1/(\lambda f)$). Поэтому плоскость X_1Y_1 является плоскостью фильтрации. В этой плоскости помещают фильтр с заданным распределением комплексного коэффициента пропускания – передаточной функцией $H(u, v) = H\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right)$, синтезируемой оптической системы, имеющей импульсную характеристику $h(x, y)$.

Пройдя через фильтр и испытав фильтрацию согласно (12.36), сигнал попадает на вторую линзу L_2 , которая расположена так, что ее передняя фокальная плоскость совпадает с задней фокальной плоскостью первой линзы. Фокусные расстояния обеих линз одинаковы и равны f . В результате второй дифракции, осуществляемой второй линзой L_2 , или просто в результате распространения света в свободном пространстве между задней фокальной плоскостью первой линзы и плоскостью изображения X_2Y_2 формируется результирующий сигнал системы – изображение объекта, подвергшегося фильтрации. Линза L_2

осуществляет, таким образом, фурье-преобразование произведения

$$\frac{1}{\lambda f} E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right) H\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right),$$

так что в соответствии с тем, что при двойным фурье-преобразованием $F\{F\{f(x, y)\}\} = f(-x, -y)$, в плоскости X_2Y_2 образуется сигнал

$$E_2(-x_2, -y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E_1(x_1, y_1) h(-x_2 - x_1, -y_2 - y_1) dx_1 dy_1.$$

Знак минус, который не позволяет считать такую систему пространственно инвариантной, можно исключить, произведя инверсию координат в плоскости изображения X_2Y_2 . Как видим, фурье-преобразование, выполняемое линзой L_2 , восстанавливает в плоскости изображения пространственный сигнал $E_2(x_2, y_2)$, соответствующий спектру $E_1(u, v)$.

Таким образом, пространственная фильтрация в схеме с двойным фурье-преобразованием состоит из следующих трех операций: 1) прямого фурье-преобразования, осуществляемого первой линзой; в результате этой операции

$$E(x, y) \rightarrow E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right) = F\{E(x, y)\};$$

2) фильтрации пространственных частот, при которой

$$E(u, v) \rightarrow E_1(u, v) = E(u, v) H(u, v);$$

3) фурье-преобразования, осуществляемого второй линзой:

$$\begin{aligned} E_1(u, v) &\rightarrow E_2(x_2, y_2) = F\{E_1(u, v)\} \\ (E_2(-x_2, -y_2) &= F^{-1}\{E_2(x_2, y_2)\}). \end{aligned}$$

Пространственную фильтрацию можно осуществить и по двухлинзовой схеме, показанной на рис. 12.20. Здесь объект помещается непосредственно за линзой L_1 , на которую падает сферическая волна от точечного источника S . Плоскость фильтрации располагается в плоскости изображения источника S (плоскость X_1). Изображение объекта во второй схеме формируется одной

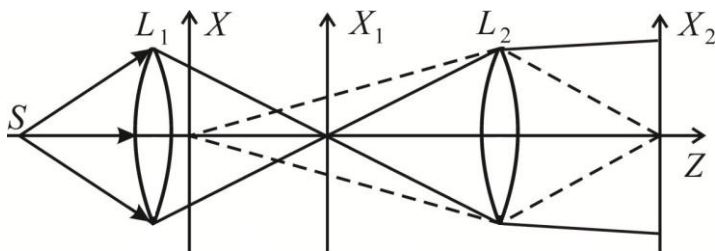


Рис.12.20

линзой L_2 , тогда как в первой схеме – системой двух линз L_1 и L_2 . Плоскость изображения во второй схеме, как и в первой, совмещена с координатной плоскостью X_2 .

Типичным примером селективной пространственной фильтрации по методу двойной дифракции является осуществление метода фазового контраста, предложенного Цернике и используемого в микроскопии для наблюдения почти прозрачных объектов, неподдающихся наблюдению в обычный микроскоп.

Рассмотрим чисто фазовый предмет с функцией пропускания

$$t(x, y) = \exp[i\varphi(x, y)].$$

Падающая на предмет по нормали плоская волна с амплитудой E_0 после прохождения его окажется промодулированной в пространстве по закону

$$E(x, y) = E_0 \exp[i\varphi(x, y)].$$

Предположим, что фаза $\varphi(x, y)$ имеет достаточно малые значения ($\varphi(x, y) \ll 2\pi$). Тогда поле входного сигнала можно представить в виде

$$E(x, y) \approx E_0[1 + i\varphi(x, y)].$$

Интенсивность сигнала на выходе

$$I(x, y) = |E_0|^2 |1 + i\varphi(x, y)|^2 = |E_0|^2 \{1 + [\varphi(x, y)]^2\} \approx |E_0|^2,$$

т.е. полезный сигнал на выходе оказывается ничтожно малым по сравнению с фоном.

Цернике предложил изменить фазу фона, поместив в фокальную плоскость первой линзы фильтр в виде прозрачной пластинки, в центре которой нанесен тонкий слой прозрачного вещества, изменяющего фазу проходящей волны на $\pi/2$ (оптическая толщина ее должна быть равна $(n-1)\lambda/4$ – пластинка в четверть волны, где n – показатель преломления вещества).

Фурье-образ функции $E(x, y)$, который создается в задней фокальной плоскости линзы L_1 (рис. 12.20) после первой дифракции, будет иметь вид

$$E(u, v) = \frac{1}{\lambda f} E_0[\delta(0) + iF\{\varphi(x, y)\}].$$

Установим в фокальной плоскости линзы L_1 указанный выше фильтр Цернике так, чтобы четвертьволновая часть покрывала центральную область спектра. Его передаточную функцию можно представить в виде

$$H(u, v) = i = \exp\left(i\frac{\pi}{2}\right)$$

для области $\delta(0)$ и

$$H(u, v) = 1$$

для остальной части спектра. Амплитудный коэффициент пропускаения четвертьволновой области полагается равным единице. Отфильтрованный спектр будет

$$E'(u, v) = E(u, v)H(u, v) = \frac{1}{\lambda f} E_0 [i\delta(0) + iF\{\varphi(x, y)\}].$$

После второй дифракции (второго преобразования Фурье) образуется изображение

$$\begin{aligned} E'(-x', -y') &= F\{E'(u, v)\} = E_0 [i + i\varphi(-x', -y')] = \\ &= E_0 i [1 + \varphi(-x', -y')], \end{aligned}$$

так как $F\{F\{f(x, y)\}\} = f(-x, -y)$. Интенсивность изображения равна

$$\begin{aligned} I'(-x', -y') &= |E_0|^2 |i[1 + \varphi(-x', -y')]|^2 \approx \\ &\approx |E_0|^2 [1 + 2\varphi(-x', -y')]. \end{aligned} \quad (12.37)$$

Таким образом, интенсивность изображения становится линейной по фазе φ (контраст интенсивности изображения пропорционален фазовому контрасту предмета) и может быть зафиксирована на уровне фона.

Можно в качестве фильтра использовать и пластинки, изменяющие фазы лучей на $3\pi/2$. В этом случае

$$H(u, v) = -i = \exp\left(i\frac{3\pi}{2}\right)$$

для центральной области $\delta(0)$,

$$E'(-x', -y') = E_0 [-i + i\varphi(-x', -y')],$$

а интенсивность изображения

$$I'(-x', -y') \approx |E_0|^2 [1 - 2\varphi(-x', -y')]. \quad (12.38)$$

Когда интенсивность изображения определяется выражением (12.37), говорят о положительном фазовом контрасте, а когда выражением (12.38) – об отрицательном фазовом контрасте. Знак минус перед координатами x' , y' в выражениях (12.37) и (12.38) можно исключить, если повернуть систему координат $X'Y'$ на 180 градусов.

Контраст изображения можно улучшить, сделав фазосдвигающий элемент поглощающим и уменьшив тем самым уровень фона.

12.8.2. Восстановление искаженных изображений

Идеи пространственной фильтрации также широко используются в фотографии, где дефекты негатива можно устранить, поместив в фокальную плоскость фильтрующей оптической системы фазосдвигающие и поглощающие пластинки. Задача восстановления искаженных изображений является одной из наиболее сложных задач обработки пространственной информации. Часто эту задачу называют задачей обращения свертки. В предположении, что процесс искажения изображения является линейным, она состоит в решении интегрального уравнения

$$E'(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(\xi, \eta) h(\xi, \eta, x, y) d\xi d\eta, \quad (12.39)$$

где функции $E(x, y)$ и $E'(x, y)$ описывают соответственно неискаженное и искаженное изображения; ядро этого интегрального преобразования – функция $h(\xi, \eta, x, y)$ описывает процесс искажения изображения (импульсный отклик процесса). В более простом случае, когда оптическая система является

пространственно-инвариантной, функцию $h(\xi, \eta, x, y)$ в подынтегральном выражении (12.39) следует заменить функцией $h(x - \xi, y - \eta)$. Тогда получим

$$E'(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(\xi, \eta) h(x - \xi, y - \eta) d\xi d\eta.$$

В символической форме

$$E'(x, y) = E(x, y) * h(x, y).$$

В плоскости пространственных частот системы фильтрации будем иметь фурье-образ искаженного изображения:

$$E'(u, v) = E(u, v)H(u, v). \quad (12.40)$$

Умножив обе части этого соотношения на $H^{-1}(u, v)$ и учтя при этом, что

$$H(u, v)H^{-1}(u, v) = 1,$$

получим исходный (неискаженный) сигнал в плоскости пространственных частот:

$$E(u, v) = E'(u, v)H^{-1}(u, v) = E'(u, v) / H(u, v). \quad (12.41)$$

Операцию, представленную этим выражением, часто называют операцией антисвертки, поскольку она определяет искомый первоначальный сигнал, т.е. решает поставленную задачу в спектральной плоскости системы пространственной фильтрации. Пространственный фильтр с модуляционной характеристикой

$$K(u, v) = H^{-1}(u, v)$$

называют инверсным фильтром. Умножим теперь числитель и знаменатель правой части равенства (12.41) на комплексно сопряженную функцию $H^*(u, v)$. Тогда получим

$$E(u, v) = E'(u, v)H^*(u, v)/|H(u, v)|^2. \quad (12.42)$$

Выражение (12.42) показывает, что неискаженный исходный пространственный сигнал $E(x, y)$ можно восстановить, если в плоскость пространственных частот системы фильтрации поместить два фильтра, один из которых имеет амплитудное пропускание, являющееся комплексно сопряженной величиной фурье-образа импульсного отклика процесса искажения, а второй имеет амплитудное пропускание, обратное величине квадрата модуля этого фурье-образа.

С помощью фильтров могут быть скорректированы такие, например, искажения изображения, как смазанность изображения, обусловленная движением фотографируемого объекта во время съемки; нерезкость, обусловленная дефокусировкой, и др.

Таким образом, если известны причины, приводящие к искажению изображения (потере информации на снимке), то в ряде случаев оказывается возможным (используя голографические методы) построить пространственный фильтр, который бы устранил эти недостатки.

Выше мы рассмотрели когерентную оптическую обработку информации. Она имеет две дуальные формы: в плоскости изображения и в плоскости пространственных частот. Для ее осуществления требуется источник света, излучение которого когерентно в пространстве и времени, т.е. он должен иметь конечные размеры и быть монохроматическим. Основными особенностями такой обработки являются возможность мгновенного (со скоростью света) осуществления двумерной свертки, возможность визуализации фурье-образа распределения комплексных амплитуд и мгновенного его видоизменения. Эти особенности являются следствием пространственной инвариантности оптической системы с двойной дифракцией.

В случае пространственно-некогерентного освещения объекта система является линейной по интенсивности. Распределение интенсивности $I_1(x, y)$ в произвольной плоскости PP' в этом случае будет описываться сверткой распределения интенсивности

на объекте $I(x, y)$ с импульсным откликом оптической системы по интенсивности $F_0(x, y)$:

$$I_1(x, y) = \int_{-\infty-\infty}^{\infty} \int_{-\infty}^{\infty} I(\xi, \eta) F_0(x - \xi, y - \eta) d\xi d\eta.$$

Это соотношение лежит в основе обработки информации методами некогерентной оптики. Как видим, оно имеет тот же вид, что и при описании когерентной пространственной фильтрации. Но в него входят теперь другие величины: вместо распределения комплексных амплитуд на объекте (которое для некогерентного объекта не имеет смысла) стоит распределение интенсивности, а вместо импульсного отклика когерентной системы – импульсный отклик некогерентной системы, равный квадрату модуля импульсного отклика когерентной системы:

$$F_0(x, y) = |h(x, y)|^2. \quad (12.43)$$

Выполнив фурье-преобразование левой и правой частей этого равенства, можно переписать его в следующем виде:

$$K(u, v) = H(u, v) * H(-u, -v), \quad (12.44)$$

так как $H^*(u, v) = H(-u, -v)$, $K(u, v) = F\{F_0(x, y)\}$ – передаточная функция некогерентной системы. Соотношение (12.44) показывает, что величина $K(u, v)$ является автокорреляцией функции $H(u, v)$. В символической форме ее можно записать как

$$K(u, v) = H(u, v) \otimes H(u, v).$$

Таким образом, передаточная функция некогерентной системы пространственной фильтрации равна функции автокорреляции передаточной функции когерентной системы пространственной фильтрации.

Глава 13

ГОЛОГРАФИЯ

13.1. Физические принципы голографии

Голографией называют интерференционный метод записи и последующего восстановления изображения без линз (открытый Д. Габором в 1947 г.). Термин «голография» переводится как «полная запись» и означает, что в этом методе фиксируется полная информация о распределении амплитуды и фазы электромагнитных волновых полей, исходящих из объекта. В основе физических принципов голографии лежит учение о волновых процессах, интерференции и дифракции когерентных пучков света. Поэтому для практической реализации голографии необходимы источники с высокой пространственной и временной когерентностью. Такими источниками являются лазеры, созданные в 1960 году. Тогда (в 1960 – 1963 гг.) и были получены (Лейтом и Упатниексом) первые высококачественные голограммы двумерных и трехмерных объектов.

Суть метода голографии можно понять, если рассмотреть процесс записи изображений в фотографии. При фотографировании каждая точка объекта является центром рассеяния падающего излучения; она посылает расходящуюся сферическую волну, которая фокусируется с помощью линзы в небольшое пятнышко на светочувствительной поверхности фотопластинки. Так как отражательная способность объекта меняется от точки к точке, интенсивность света, попадающего на соответствующие элементы фотопластинки, оказывается различной. Вследствие этого на фотопластинке и записывается изображение объекта. При этом каждый элемент светочувствительной поверхности дает изображение соответствующей точки объекта. Трехмерные объекты на светочувствительной поверхности фотопластинки регистрируются в виде плоских двумерных изображений.

В отличие от фотографии, голография позволяет записывать и восстанавливать не двумерное распределение в плоскости снимка интенсивности излучения, рассеянного предметом (если изображаемый предмет непрозрачный) или прошедшего через предмет (если предмет прозрачный), а амплитуду и фазу волны, идущей от предмета (эту волну называют предметной или объектной волной). В результате при восстановлении голограммы получается точная копия записанной волны, которая несет информацию об изображаемом предмете. Попадая в глаз наблюдателя, эта волна вызывает такие же зрительные ощущения, как и при непосредственном рассматривании предмета, и создает полную иллюзию реальности наблюдения предмета.

Таким образом, задачей голографии, как и фотографии, является запись, хранение и воспроизведение зрительных образов. Но если фотография позволяет записывать лишь подобия плоских проекций поверхностной освещенности предмета, то голография дает возможность точно воссоздать пространственную структуру светового поля, т.е. создать его оптическую копию, визуально не отличимую от оригинала. Фотографическую запись, которая, будучи освещенной надлежащим образом, восстанавливает волну, исходившую от предмета, называют голограммой. Внешне голограмма почти ничем не отличается от обычной равномерно засвеченной фотопластинки.

Возможность регистрации амплитуды и фазы световой волны обусловлена тем, что при интерференции волны с другой, когерентной с ней, волной результат интерференции зависит от разности фаз $\delta = \varphi_2 - \varphi_1$ этих волн, а амплитуда результирующего колебания E_0 в этом случае определяется соотношением

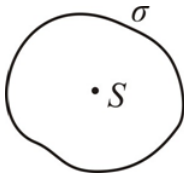
$$E_0^2 = E_{01}^2 + E_{02}^2 + 2E_{01}E_{02} \cos \delta.$$

Следовательно, если в поле интерференции поместить фотопластинку, то плотность почернения каждого ее участка будет содержать информацию об амплитудах и фазовых соотношениях интерферирующих волн.

Метод голографии основан на принципе Гюйгенса – Френеля, который применительно к голографии можно сформулировать следующим образом. Колебания светового поля

$$E(x, y, z, t) = E(x, y, z)\exp[-i(\omega t - \varphi(x, y, z))], \quad (13.1)$$

созданные когерентным источником S (рис. 13.1) на поверхности σ , окружающей источник S , однозначно определяют волновое поле за поверхностью σ , если там нет других источников. Поле в



$P(x, y, z)$

любой точке $P(x, y, z)$ одинаково как в случае реального источника, так и в случае, когда источника нет, но при условии, что поверхность, охватывавшая источник, создает такое же

Рис. 13.1

колебание (13.1), какое создавал источник.

Пусть некоторый объект O (рис. 13.2) освещается монохроматическим светом от точечного источника S , вследствие чего объект создает отраженную и дифрагировавшую на нем монохроматическую световую электромагнитную волну, которая регистрируется в точке P . Выделим между объектом и указанной точкой воображаемую плоскость Π . Волна,

распространяющаяся от объекта O в точку P , пересекает эту плоскость. Свяжем с этой плоскостью Π координатную плоскость XU . Объект создает в каждой точке плоскости Π световое поле с амплитудой $E(x, y)$ и фазой $\varphi(x, y)$, совокупность значений которых полностью

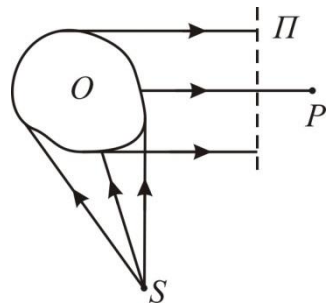


Рис. 13.2

определяет состояние отраженно-рассеянной объектом волны при ее прохождении через эту

плоскость. Таким образом, можно считать, что поле в точке P создается объектом O , но также можно считать, что это поле создается воображаемой плоскостью Π . Уберем теперь объект и его освещение и создадим каким-либо образом в плоскости Π то же распределение амплитуд $E(x, y)$ и фаз $\varphi(x, y)$. Согласно принципу Гюйгенса – Френеля результирующий эффект в точке P в этом случае будет таким же, как и при наличии объекта и его освещения, так как в обоих случаях будет одно и то же распределение вторичных источников, размещенных на плоскости Π с амплитудой $E(x, y)$ и фазой $\varphi(x, y)$ ¹². Наблюдатель, находящийся справа от плоскости Π , наблюдая только результат прохождения волны через эту плоскость, также не может установить, каким образом образовалось данное распределение амплитуд $E(x, y)$ и фаз $\varphi(x, y)$, поскольку волна, распространяющаяся от плоскости Π , несет такую же информацию об объекте, что и волна, исходящая непосредственно от самого объекта. Поэтому у наблюдателя на сетчатке глаза сформируется изображение объекта, полностью идентичное самому объекту. Наблюдатель смотрит сквозь голограмму, как сквозь окно, и видит объект, хотя его в действительности нет, и к тому же он видит его как реальный объект, т.е. с его объемностью, рельефом, параллаксом и глубиной резкости. Объемность создается благодаря способности наблюдателя воспринимать разницу в удаленности от него различных частей объекта. Это связано с тем, что в волне наряду с амплитудами воспроизводятся и фазы (точнее, соотношения фаз волн, отраженных от разных элементов освещаемого объекта). В окрестности каждой точки голограммы в интерференционной форме представлены все элементы объекта, видимые из этой точки, поэтому, перемещая взгляд вдоль голограммы, наблюдатель видит, как выглядит объект с разных направлений

¹² Данное утверждение предполагает единственность решения обратной задачи дифракции.

наблюдения. Это приводит к совершенной иллюзии наблюдения реального объекта. Полученное изображение можно сфотографировать обычными методами.

Таким образом, если во втором случае мы сможем добиться того, чтобы распределение поля на воображаемой плоскости P было таким же, как и при наличии объекта и его освещения, то мы можем реализовать голографическую запись. Чтобы получить такую запись, недостаточно расположить в плоскости P фотопластинку, преобразующую изменения комплексной амплитуды падающей волны в изменения комплексного амплитудного пропускания за счет ослабления или фазового сдвига, так как фотоэмульсия является квадратичным детектором, чувствительным к интенсивности света:

$$I(x, y) = |E(x, y) \exp \{-i[\omega t + \varphi(x, y)]\}|^2 = |E(x, y)|^2.$$

Следовательно, при обычном фотографическом методе регистрации информация о фазе $\varphi(x, y)$ волны безвозвратно теряется.

Следует, однако, отметить, что и в обычной фотографии фазовая информация в волне света играет, хотя и неявную, роль. Она соответствует расстояниям от точки предмета до фотопластинки, поэтому если удалить фазовую информацию, то на фотопластинке никакого изображения не получится. Проблема регистрации фазовой информации в обычной фотографии устраняется косвенно системой наводки на фокус. Изображение точки становится резким на негативе, потому что в этих условиях сферическая волна стягивается в точку.

Для того чтобы сохранить информацию и об амплитуде, и о фазе, т.е. произвести голографическую запись объекта, распределение поля (13.1) преобразуют в стоячую волну путем сложения с когерентным фоном. Таким способом можно получить сетку интерференционных полос, образующих голограмму. Локальные изменения контрастности полос и интервалов между ними в интерферограмме выражают соответствующие изменения амплитуды и фазы волны (13.1).

Из сказанного выше следует, что создать голограмму – значит восстановить на плоскости Π колебания $E(x, y, t)$ с таким же распределением амплитуды $E(x, y)$ и фазы $\varphi(x, y)$, которое давала первичная волна от изображаемого объекта. При этом в направлении за плоскость Π будет распространяться волновое поле, которое исходило из реального объекта.

Схема получения изображения предмета в голографии выглядит следующим образом. Изображаемый объект O освещается (или просвечивается в случае транспаранта) пучком света лазера L , предварительно уширенным оптическим устройством (рис. 13.3). Получающаяся при дифракции предметная волна падает на фотопластинку $\Phi\Pi$. Вместе с этой волной на фотопластинку попадает часть света от того же лазера после отражения от зеркала M , которая и является указанным выше когерентным фоном. Эту, не испытавшую рассеяния (дифракции) волну называют опорной или референтной волной. В качестве опорной волны

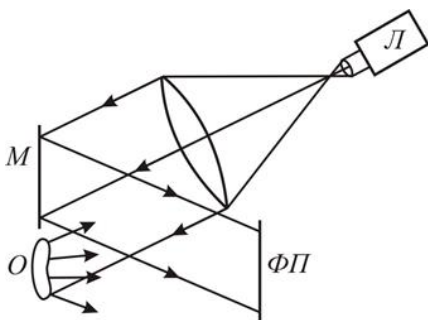


Рис. 13.3

обычно берут волну с простым фронтом (плоскую или сферическую), хотя можно использовать и волну, рассеянную другим объектом. Поскольку предметная и опорная волны когерентны, то на поверхности фотопластинки происходит их интерференция.

Интерференционная картина регистрируется фотопластинкой в виде своеобразной нерегулярной дифракционной решетки. Экспонированная таким способом, а затем проявленная и

закрепленная фотопластинка и есть голограмма изображаемого предмета. Как видим, на голограмме регистрируется не оптическое изображение предмета, а интерференционная картина, возникающая при наложении световой волны, исходящей от предмета, и когерентной с ней опорной волны. Если теперь, убрав предмет, осветить полученную голограмму одной только опорной волной (рис. 13.4) или другой волной, фронт которой подобен фронту опорной волны, то эта волна испытает дифракцию, как на дифракционной решетке. При этом одна из дифрагировавших волн будет совпадать с волной, рассеянной изображаемым предметом (или прошедшей через изображаемый предмет) при получении голограммы, т.е. происходит восстановление предметной волны. Если смотреть навстречу этой волне, то за голограммой можно увидеть трехмерное изображение предмета. Это мнимое изображение предмета (на рис. 13.2 показано пунктиром) будет располагаться в том же месте, где он был на стадии изготовления голограммы. Мнимое изображение наблюдается через голограмму, как через окно.

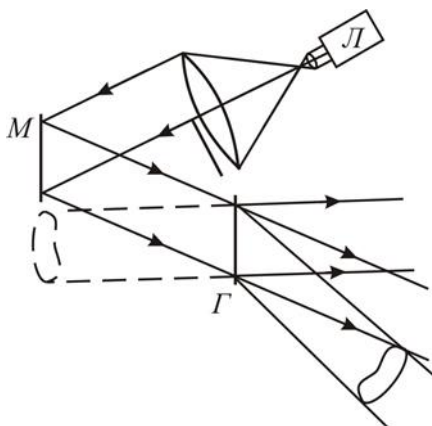


Рис. 13.4

Как видим, при считывании информации с голограммы ее так же, как и при считывании информации с обычной фотографии, освещают источником света. Но, в отличие от обычной фотографии, голограмму освещают когерентным светом. Информация при этом воспринимается глазом не непосредственно с помощью отраженных от голограммы лучей света, как в случае фотографии; здесь процесс считывания информации гораздо более сложен.

Освещение голограммы восстанавливающей волной, в результате дифракции которой восстанавливается изображение предмета, называется реконструкцией голограммы.

Следует отметить, что для изготовления голограмм требуются специальные высокоразрешающие фотоматериалы с разрешающей способностью 2000 – 5000 линий на 1 мм (напомним, что обычные фотопленки и фотопластинки позволяют разрешать около 100 линий на 1 мм). Создание таких мелкозернистых и одновременно достаточно чувствительных фотоматериалов представляет собой довольно сложную задачу.

13.2. Голография простейших объектов

13.2.1. Плоские предметная и опорная волны

Рассмотрим сначала простейший случай плоской предметной и плоской опорной волн. Плоская предметная волна формируется, например, рассеивающим падающий когерентный свет бесконечно удаленным точечным объектом или когерентно освещенным точечным объектом, находящимся в переднем фокусе собирающей линзы, расположенной перед фотопластинкой. Пусть плоский предметный пучок 1 () падает под углом ψ к направлению нормали (оси Z , рис. 13.5) к фотопластинке (ФП), расположенной в плоскости $z = 0$. Опорный пучок 2 падает нормально к фотопластинке. Светочувствительный слой фотопластинки способен зарегистрировать лишь среднее за время экспозиции распределение освещенности (интенсивности). По степени почернения

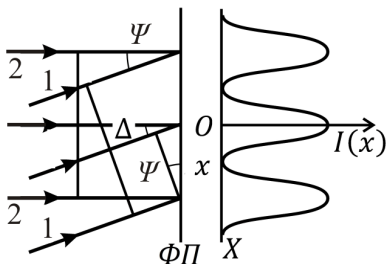


Рис. 13.5

пластинки можно судить лишь об амплитуде световых колебаний, поэтому если бы на фотопластинку падала одна только предметная волна, то информация об ее фазе была бы полностью утрачена. При плоской предметной волне

$$E_{\text{пр}} = E_1 \exp[ik(x \sin \psi + z \cos \psi)],$$

где E_1 – амплитуда предметной волны, x – абсцисса в плоскости фотопластинки, фотопластинка была бы зачернена равномерно: $I = |E_1|^2 = \text{const}$. При одновременном падении на фотопластинку предметной и плоской опорной волны

$$E_{\text{оп}} = E_2 \exp(ikz)$$

будет происходить их интерференция. Результирующая амплитуда в плоскости фотопластинки (плоскости $z = 0$) определится суммой комплексных амплитуд объектной и референтной волн:

$$E(x) = E_1 \exp(ikx \sin \psi) + E_2.$$

В результате интерференции предметного и опорного пучков на поверхности фотопластинки установится стационарное распределение интенсивности $I(x) = |E(x)|^2$ (рис. 3.2) в виде системы параллельных полос. Это распределение будет иметь вид

$$I(x) = E_1^2 + E_2^2 + E_1 E_2 \exp(ikx \sin \psi) + E_1 E_2 \exp(-ikx \sin \psi).$$

Интерференционная картина фиксируется фотопластинкой. Если фотопластинка проэкспонирована и проявлена таким образом, что рабочий диапазон амплитудной характеристики

записи сигнала на фотопленку $|t| = f(H) = f_1(I)$ (рис. 12.16) не выходит за пределы линейного участка, то амплитудная функция пропускания фотопластинки будет определяться зависимостью (12.33):

$$t = t_0 + \beta(I - I_0),$$

где I_0 – среднее (рабочее) значение интенсивности. Подставляя сюда выражение $I(x)$ и принимая $I_0 = E_1^2 + E_2^2$, для амплитудного пропускания фотопластинки будем иметь

$$t(x) = t_0 + \beta E_1 E_2 \exp(ikx \sin \psi) + \beta E_1 E_2 \exp(-ikx \sin \psi) \quad (13.2)$$

или

$$t(x) = t_0 + 2\beta E_1 E_2 \cos(kx \sin \psi). \quad (13.3)$$

Выражения (13.2) и (13.3) называют уравнением голограммы. Физический смысл уравнения голограммы состоит в том, что в функции амплитудного пропускания голограммы заключена информация об амплитуде (E_1) и фазе ($kx \sin \psi$) объектной волны. Особенностью уравнения голограммы является то, что точное значение коэффициента β не влияет на процесс получения голограммы.

Полученная голограмма будет представлять собой дифракционную решетку, прозрачность которой изменяется по гармоническому закону. Период этой решетки будет равен расстоянию между соседними интерференционными полосами: $d = h = \lambda / \sin \psi$ (формула (3.5)). Структура зарегистрированных на голограмме интерференционных полос содержит, таким образом, информацию о распределении амплитуд и фаз в предметной волне, а также о ее направлении (так как h зависит от угла ψ).

При освещении голограммы нормально падающей (т.е. распространяющейся в направлении опорной волны) восстанавливающей плоской волной эта волна, проходя сквозь

голограмму, дифрагирует на ее тонкой структуре. Если восстанавливающая волна идентична опорной волне, то распределение светового поля непосредственно за голограммой (в плоскости $z = 0$) будет описываться функцией $E'(x) = E_2 t(x)$. Подставляя в это соотношение выражение (13.2), получим

$$E'(x) = E_2 t_0 + \beta E_1 E_2^2 \exp(ikx \sin \psi) + \beta E_1 E_2^2 \exp(-ikx \sin \psi). \quad (13.4)$$

Это распределение амплитуд и фаз в плоскости $z = 0$ однозначно определяет ход волнового процесса в области $z > 0$ (рис. 13.6). Таким образом, за голограммой распределение комплексной амплитуды светового поля описывается суммой трех слагаемых (так и должно быть при дифракции на гармонической дифракционной решетке). Первое слагаемое $E_2 t_0$ в (13.4) представляет собой с точностью до постоянного множителя t_0 амплитуду плоской волны, используемой для восстановления (волна 2 на рис. 13.6.). Эта плоская волна прошла сквозь голограмму без изменения направления, но ослабленная (так как $t_0 < 1$). Второй член ($\beta E_1 E_2^2 \exp(ikx \sin \theta)$) определяет комплексную амплитуду волны, распространяющейся под тем же самым углом θ к направлению оси Z , что и предметная волна при записи голограммы, т.е. с точностью до постоянного множителя представляет собой амплитуду предметной волны. Третий член $\beta E_1 E_2^2 \exp(-ikx \sin \psi)$, являющийся комплексно сопряженным второму члену, представляет собой амплитуду волны, распространяющейся под углом $-\psi$ к направлению нормали к голограмме (к оси Z). Следовательно, при освещении голограммы плоской монохроматической волной, идентичной волне, используемой в качестве опорной, происходит восстановление предметной волны.

Первая (не отклоненная) плоская волна, соответствующая дифракционному максимуму нулевого порядка ($m = 0$), не

попадает в глаз наблюдателя при надлежащем его расположении. Вторая

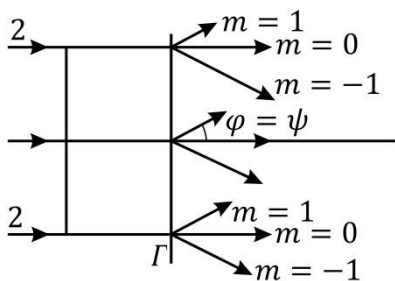


Рис. 13.6

плоская волна, распространяющаяся под углом ψ к оси Z и соответствующая максимуму первого порядка ($m = 1$), дает мнимое изображение предмета. Третья плоская волна, распространяющаяся под углом $-\psi$ к оси Z и соответствующая максимуму

минус первого порядка ($m = -1$), формирует еще одно – действительное – изображение предмета.

Легко убедиться, что если в качестве восстанавливающей волны использовать плоскую монохроматическую волну, распространяющуюся в том же направлении, что и предметная волна, то произойдет восстановление опорной волны. Действительно, если в качестве восстанавливающей волны использовать плоскую волну, распространяющуюся в направлении предметной волны с амплитудой E_1 , т.е. волну вида $E_1 \exp(ikx \sin \psi)$, то распределение амплитуды поля в плоскости непосредственно за голограммой будет иметь вид

$$E'(x) = E_1 t_0 \exp(ikx \sin \psi) + \beta E_1^2 E_2 + \beta E_1^2 E_2 \exp(i2kx \sin \psi).$$

Второй член в этом выражении и представляет собой с точностью до постоянного амплитудного множителя плоскую опорную волну, прошедшую через голограмму без отклонения.

Это общее свойство любых голограмм обычно формулируют так: опорный и предметный пучки полностью равноправны. На этом свойстве основан метод коррекции волновых фронтов. В этом методе лазерный пучок со сложным волновым фронтом с помощью изготовленной заранее голограммы (пучок является для голограммы предметным пучком) преобразуется в пучок с плоским волновым фронтом, совпадающим с фронтом того

опорного пучка, который использовался при изготовлении голограммы.

Произвольное волновое поле, исходящее из предмета, освещаемого (или просвечиваемого) когерентным лазерным пучком при голографировании (создании голограммы), в соответствии с методом Рэлея, можно представить в виде совокупности плоских волн. Каждая такая волна при интерференции с опорной волной, получаемой из того же лазерного пучка, создает на фотопластинке свою систему интерференционных полос с характерной для нее ориентацией и периодом. После проявления и закрепления на голограмме образуется совокупность дифракционных решеток с гармоническими функциями пропускания. Каждая из этих решеток на этапе восстановления при дифракции на ней пучка света, идентичного с опорным, формирует соответствующую ей исходную плоскую волну. Все восстановленные элементарные плоские волны находятся в таких же амплитудных и фазовых соотношениях, как и при записи голограммы. Их совокупность и воссоздает полное, рассеянное объектом, световое поле.

13.2.2. Точечный объект, плоская нормально падающая опорная волна

Пусть точечный монохроматический источник S (рис. 13.7) находится на оси Z на расстоянии a от фотопластинки. Источник создает сферическую волну 1 (расходящуюся из точки S), световое поле которой в плоскости фотопластинки ($z = 0$) будет

$$E_{\text{сф}} = \frac{E_{01}}{r} \exp(ikr) = E_1 \exp(ikr),$$

где $r = \sqrt{x^2 + y^2 + a^2}$; начальную фазу волны полагаем равной нулю. Одновременно с этой предметной сферической волной на фотопластинку падает когерентная с ней плоская опорная волна 2:

$$E_{\text{он}} = E_2 \exp[ik(z + a)].$$

Поле, создаваемое этой волной в плоскости фотопластинки,

$$E_{\text{он}} = E_2 \exp(ika).$$

Результирующее поле в плоскости фотопластинки

$$E(r) = E_{\text{пр}} + E_{\text{он}} = E_1 \exp(ikr) + E_2 \exp(ika),$$

а распределение интенсивности

$$I(r) = E_1^2 + E_2^2 + E_1 E_2 \exp[ik(r - a)] + \\ + E_1 E_2 \exp[-ik(r - a)].$$

Если условия экспонирования и проявления фотопластинки соответствовали прямолинейному участку кривой $|t| = f(H) = f_1(I)$, то для функции амплитудного пропускания голограммы (уравнения голограммы) будем иметь

$$t(x, y) = t_0 + \beta E_1 E_2 \exp[ik(r - a)] + \beta E_1 E_2 \exp[-ik(r - a)].$$

Используя параксиальное приближение, уравнение голограммы можно представить в виде

$$t(x, y) = t_0 + \beta E_1 E_2 \exp\left(ik \frac{x^2 + y^2}{2a}\right) + \beta E_1 E_2 \exp\left(-ik \frac{x^2 + y^2}{2a}\right)$$

или

$$t(\rho) = t_0 + 2\beta E_1 E_2 \left(k \frac{\rho^2}{2a}\right),$$

где $\rho = \sqrt{x^2 + y^2}$. Обозначив $t_1 = 2\beta E_1 E_2$, придем к выражению

(3.17), $I(\rho) = \alpha + \beta \cos k \frac{\rho^2}{2a}$. Таким образом, на фотопластинке

после ее проявления и закрепления будет зафиксирована интерфе

ренционная картина в виде зонной решетки Габора с косинусоидальной функцией пропускания. Интерференционная картина будет представлять собой систему концентрических чередующихся светлых и темных колец. При этом радиусы светлых колец, как было показано в п. 3.2.3, определяются формулой $\rho_m = \sqrt{2m a \lambda}$, а темных – формулой $\rho_m = \sqrt{(2m + 1) a \lambda}$, где $m = 0, 1, 2, \dots$. Полученную голограмму называют голограммой точки.

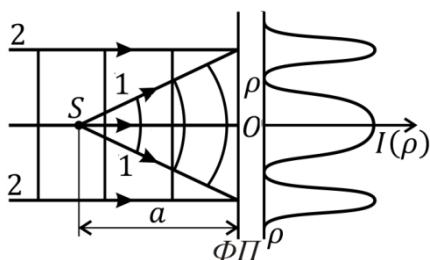


Рис. 13.7

На стадии восстановления голограмму точки освещают нормально падающей плоской волной, идентичной опорной волне 2. Тогда в результате дифракции этой волны на голограмме в плоскости непосредственно за голограммой возникает

распределение поля $E'(\rho) = E_2 t(\rho)$. Подставив сюда первое уравнение голограммы, получим

$$E'(x, y) = E_2 t_0 + \beta E_1 E_2^2 \exp\left(ik \frac{x^2 + y^2}{2a}\right) + \beta E_1 E_2^2 \exp\left(-ik \frac{x^2 + y^2}{2a}\right).$$

Как видим, при дифракции плоской восстанавливающей волны на голограмме точки, как и на зонной решетке Габора, световое поле в плоскости $z = 0$ представляется в виде суммы трех слагаемых. Соответственно, в области $z > 0$ возникают три дифрагировавшие волны: плоская волна с амплитудой $E_2 t_0$, распространяющаяся в направлении падающей восстанавливающей волны, и две

на помещенном в это место экране в виде светового дифракционного кружка. Это изображение формируется без помощи линз и зеркал.

Протяженный предмет, рассеивающий когерентный лазерный свет, можно рассматривать как совокупность точечных источников, испускающих сферические волны. В результате их интерференции с опорной волной на голограмме возникает сложный узор в виде наложения элементарных зонных решеток Габора. На этапе восстановления при дифракции опорной волны каждая такая решетка формирует сферическую волну, расходящуюся от центра, где находилась соответствующая рассеивающая свет точка предмета при записи голограммы. Совокупность этих волн, создающих мнимое изображение всего предмета, неотличима от волны, рассеянной предметом при записи, так как характеризуется тем же распределением амплитуд и фаз. Совокупность сходящихся волн, возникающих при дифракции восстанавливающей волны на голограмме, без линз формирует действительное изображение предмета.

Следует отметить, что создаваемое голограммой изображение точки (действительное и мнимое), как и любой другой оптической изображающей системой, имеет вид дифракционного пятна (кружка Эри). Это связано с тем, что сферический волновой фронт ограничен здесь размерами голограммы (и, следовательно, имеет место дифракция на краях голограммы, подобно дифракции на оправе линзы), поэтому сходящаяся или расходящаяся сферическая волна, формируемая голограммой, ничем не отличается от сферической волны, возникающей за линзой, диаметр D которой равен размеру голограммы. Размер этого пятна определяется той же формулой $r_1' = 1,22\lambda b / D$, что и в случае линзы, в которой под D следует понимать размер голограммы, а под b – расстояние от голограммы до изображения. Полученная на основе этой формулы разрешающая способность линзы применима и для оценки разрешающей способности голограммы: $R = D / (1,22\lambda)$. Минимально разрешимое расстояние между точками голограммы также может быть

определено с помощью полученного в п. 12.1 соотношения $\Delta x = 1,22\lambda a / D$. Здесь a – расстояние от точечного источника до голограммы в процессе записи.

13.2.3. Точечный объект, плоская наклонная опорная волна

Как видно на рис. 13.7, при использовании нормально падающей плоской опорной волны все три порядка дифракции на голограмме восстанавливающей волны перекрываются, в результате чего на восстановленном изображении появляются паразитные искажения. Перекрывание наблюдается даже в том случае, если голограмму осветить наклонной плоской волной. Для разделения изображений Лейт и Упатниекс предложили при записи голограммы использовать наклонную опорную волну.

Пусть на фотопластинку под углом $-\theta$ к нормали (к оси Z) одновременно с объектной падает плоская опорная волна

$$E_{\text{он}}(x, z) = E_2 \exp[ik(a - x \sin \theta + z \cos \theta)],$$

когерентная с предметной (рис. 13.9). Поле, создаваемое опорной волной в плоскости фотопластинки ($z = 0$),

$$E_{\text{он}}(x) = E_2 \exp[ik(-x \sin \theta + a)].$$

Точечный источник, как и прежде, располагается в точке $(0, 0, -a)$. Используя, как и ранее, параксиальное приближение, для результирующего распределения интенсивности в плоскости фотопластинки в этом случае будем иметь

$$I(x, y) = E_1^2 + E_1^2 + E_1 E_2 \exp(ikx \sin \theta) \exp\left(ik \frac{x^2 + y^2}{2a}\right) + E_1 E_2 \exp(-ikx \sin \theta) \exp\left(-ik \frac{x^2 + y^2}{2a}\right).$$

Уравнение голограммы

$$t(x, y) = t_0 + \beta E_1 E_2 \exp(ikx \sin \theta) \exp\left(ik \frac{x^2 + y^2}{2a}\right) + \beta E_1 E_2 \exp(-ikx \sin \theta) \exp\left(-ik \frac{x^2 + y^2}{2a}\right). \quad (13.5)$$

Осветим полученную голограмму нормально падающей плоской волной с амплитудой E_0 . Получим за голограммой распределение светового поля в виде

$$E'(x, y) = E_0 t_0 + \beta E_0 E_1 E_2 \exp(ikx \sin \theta) \exp\left(ik \frac{x^2 + y^2}{2a}\right) + \beta E_0 E_1 E_2 \exp(-ikx \sin \theta) \exp\left(-ik \frac{x^2 + y^2}{2a}\right). \quad (13.6)$$

Первый член в этом выражении, как и в предыдущих аналогичных выражениях, представляет собой с точностью до постоян-

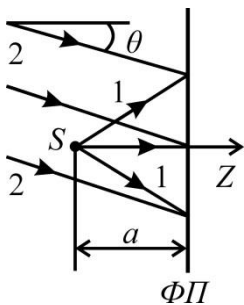


Рис. 13.9

ного множителя плоскую волну, используемую для восстановления. Второй член дает с точностью до множителя $\beta E_0 E_2 \exp(ikx \sin \theta)$ комплексную амплитуду волны, испускаемой точечным источником. Это – расходящаяся сферическая волна, источником которой служит мнимый источник S' , расположенный на расстоянии a перед голограммой. Фазовый множитель $\exp(ikx \sin \theta)$ указывает на то, что среднее направление этой волны составляет угол θ с нормалью к плоскости голограммы. Преобразовав произведение фазовых

множителей $\exp(ikx \sin \theta) \exp\left(ik \frac{x^2 + y^2}{2a}\right)$ во втором члене

выражения (13.6), приведем это произведение к виду $A \exp\left[i \frac{k}{2a} [(x + a \sin \theta)^2 + y^2]\right]$, где $A = \exp\left(-ik \frac{1}{2} a \sin^2 \theta\right)$

– комплексная постоянная. Отсюда следует, что мнимый источник S' расположен в точке с координатами $x_0 = -a \sin \theta$, $y_0 = 0$ на расстоянии a слева от голограммы.

Третий член в выражении (13.6) пропорционален амплитуде комплексно сопряженной волны $E_1 \exp\left(-ik \frac{x^2 + y^2}{2a}\right)$. Это –

сходящаяся сферическая волна, образующая действительное изображение S'' источника, расположенного на расстоянии a справа от голограммы. Фазовый множитель $\exp(-ikx \sin \theta) = \exp[ikx \sin(-\theta)]$ свидетельствует о том, что среднее направление волны, образующей изображение S'' , составляет угол $-\theta$ с нормалью к плоскости голограммы. Произведение фазовых множителей третьего члена приводится к виду

$A \exp\left\{-\left[i \frac{k}{2a} (x + a \sin \theta)^2 + y^2\right]\right\}$. Отсюда следует, что

источник S'' имеет те же x - и y -координаты, что и источник S' , но расположен справа от голограммы, т.е. симметрично источнику S' относительно плоскости голограммы, на расстоянии a от нее. Все три дифрагировавшие на голограмме волны показаны на рис. 13.10.

Таким образом, введение наклонного опорного пучка приводит к смещению мнимого и действительного изображений так, что они не создают помех друг другу при их наблюдении. При таком способе и когерентный фон восстанавливающей плоской волны не создает помех при наблюдении изображений S' и S'' .

Сравнивая квадратичные фазовые множители во втором и третьем членах уравнения голограммы (13.5) с функцией пропускания линзы, мы видим, что они совпадают с этими функциями, если положить расстояние a равным фокусному расстоянию линзы f . При этом фазовый множитель во втором члене совпадает с функцией пропускания рассеивающей линзы, а в третьем – с функцией пропускания собирающей линзы. Наличие линейных фазовых множителей в указанных членах приводит к отклонению светового пучка либо вверх, либо вниз в зависимости от знака показателя экспоненты. Действие этих множителей аналогично действию призмы. Поэтому действия второго и третьего членов уравнения голограммы на плоскую восстанавливающую волну аналогичны совместным действиям трехгранной призмы и линзы. Второй член не только отклоняет пучок вверх, но и действует как рассеивающая линза. Падающая плоская волна поворачивается вверх и превращается в расходящуюся из точки S' сферическую волну. Аналогично третий член в (13.5) действует на параллель-

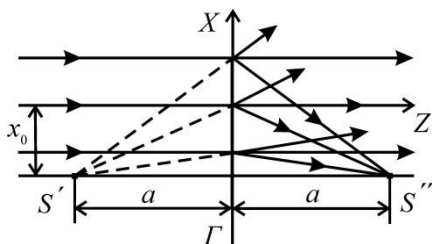


Рис. 13.10

ный пучок не только как отклоняющая вниз призма, но и как собирающая линза. Повернутый вниз пучок превращается в сходящуюся в точку S'' сферическую волну.

Голограмма, полученная при нормально падающем плоском опорном пучке, действует как рассеивающая и собирающая линза одновременно.

Если в качестве восстанавливающей волны использовать плоскую волну, совершенно идентичную опорной волне, то за голограммой получим распределение амплитуды светового поля

$$E'(x, y) = t_0 E_2 \exp(-ikx \sin \theta) + \beta E_1 E_2^2 \exp\left(ik \frac{x^2 + y^2}{2a}\right) +$$

$$+ \beta E_1 E_2^2 \exp(-2ikx \sin \theta) \exp\left(-ik \frac{x^2 + y^2}{2a}\right).$$

Здесь первый член с точностью до постоянного множителя представляет собой амплитуду восстанавливающей волны, ослабленную голограммой. Расходящаяся сферическая волна, комплексная амплитуда которой с точностью до постоянного множителя представляется вторым членом, создает мнимое изображение источника S и имеет центром точку S' , расположенную точно в том же месте, где располагался реальный источник S при его голографировании. Для определения положения действительного изображения S'' источника, описываемого третьим членом, приведем его общий фазовый множитель к виду

$$B \exp\left[i \frac{k}{2a} [(x + 2a \sin \theta)^2 + y^2]\right],$$

где $B = \exp(-ik2a \sin^2 \theta)$ – комплексная постоянная. Отсюда видно, что действительное изображение расположено в точке с координатами $x_0 = -2a \sin \theta$, $y_0 = 0$ на расстоянии a слева от голограммы. Как видим, и в этом случае изображения S' и S'' не создают помех друг другу при их наблюдении.

Если в качестве восстанавливающей волны взять комплексно сопряженную плоскую наклонную волну, т.е. волну вида $E_v = E_2 \exp(ikx \sin \theta)$, то, как легко убедиться, действительное изображение S'' будет располагаться симметрично точке, в которой располагался точечный источник S при записи голограммы, а мнимое изображение будет смещено вверх от оси голограммы (оси Z) на расстояние $x_0 = 2a \sin \theta$.

13.3. Голография произвольного предмета

Рассмотрим теперь восстановление изображения произвольного предмета непосредственно. Пусть излучение, рассеянное предметом, создает на плоскости фотопластинки поле с амплитудой $E_1(x, y)$ и фазой $\varphi(x, y)$, где x и y – координаты в плоскости фотопластинки. Комплексная амплитуда предметной волны в плоскости фотопластинки определится выражением

$$E_{\text{пр}}(x, y) = E_1(x, y)\exp[i\varphi(x, y)].$$

Одновременно с предметной на ту же фотопластинку под углом $-\theta$ к нормали (к оси Z) падает когерентная ей плоская опорная волна, комплексная амплитуда которой в плоскости фотопластинки ($z = 0$)

$$E_{\text{оп}}(x) = E_2\exp(-ikx\sin\theta).$$

Опорный пучок является частью широкого монохроматического пучка лазера, освещающего предмет, отклоненный с помощью прямоугольной призмы (рис. 13.11).

Общая комплексная амплитуда светового поля в плоскости фотопластинки определится суммой

$$E(x, y) = E_1(x, y)\exp[i\varphi(x, y)] + E_2\exp(-ikx\sin\theta).$$

Распределение интенсивности в плоскости фотопластинки представится в виде

$$\begin{aligned} I(x, y) &= |E(x, y)|^2 = E_1^2(x, y) + E_2^2 + \\ &+ E_1(x, y)E_2\exp\{ik[x\sin\theta + \varphi(x, y)]\} + \\ &+ E_1(x, y)E_2\exp\{-ik[x\sin\theta + \varphi(x, y)]\}. \end{aligned}$$

Это распределение фиксируется фотопластинкой. При соответ-

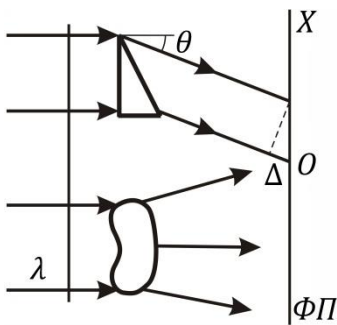


Рис. 3.11

в процессе экспонирования и химической обработке амплитудный коэффициент пропускания голограммы будет линейно связан с интенсивностью (формула (12.33)). Подставляя в эту формулу полученное выражение для $I(x, y)$ и положив $I_0 = E_2^2$, для функции амплитудного пропускания голограммы будем иметь

$$t(x, y) = t_0 + \beta E_1^2(x, y) + \beta E_1(x, y) E_2 \exp\{ik[x \sin \theta + \varphi(x, y)]\} + \beta E_1(x, y) E_2 \exp\{-ik[x \sin \theta + \varphi(x, y)]\}$$

или

$$t(x, y) = t_0 + \beta E_1^2(x, y) + 2\beta E_1(x, y) E_2 \cos\{k[x \sin \theta + \varphi(x, y)]\}.$$

Полученные выражения называются уравнениями голограммы с наклонным пучком. Голограмма с таким уравнением несет полную информацию как об амплитуде, так и о фазе волны, рассеянной предметом. Это следует из того, что в уравнение голограммы (в экспоненциальной форме) амплитуда $E_1(x, y)$ и фаза $\varphi(x, y)$ рассеянной предметом волны входят естественным образом, т.е. как $E_1(x, y) \exp[i\varphi(x, y)]$ в третий член и как $E_1(x, y) \exp[-i\varphi(x, y)]$ в четвертый. Мы видим, что благодаря использованию опорного пучка удастся сохранить и амплитуду, и фазу рассеянной предметом волны.

При освещении голограммы нормально падающей плоской монохроматической волной $E_b = E_0 \exp(ikz)$, за голограммой

вследствие дифракции возникнет световое поле, амплитуда которого $E'(x, y) = E_0 t(x, y)$ будет иметь вид

$$E'(x, y) = t_0 E_0 + \beta E_0 E_1^2(x, y) + \\ + \beta E_0 E_1(x, y) E_2 \exp(ikx \sin \theta) \exp[i\varphi(x, y)] + \\ + \beta E_0 E_1(x, y) E_2 \exp(-ikx \sin \theta) \exp[-i\varphi(x, y)].$$

(13.7)

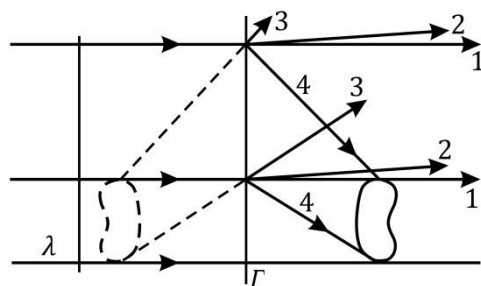


Рис. 13.12

Полученное выражение описывает преобразование плоской волны, дифрагировавшей на голограмме. Оно содержит четыре слагаемых, которым соответствуют четыре волны, распространяющиеся в различных направлениях (рис. 13.12). Следовательно, в результате дифракции плоской восстанавливающей волны на голограмме возникают четыре световые волны, различающиеся направлением распространения и пространственным изменением их амплитуды и фазы.

Первая волна (волна 1) распространяется в направлении восстанавливающей волны. Информации об объекте эта волна не несет. Вторая волна (2) представляет собой световое поле, промодулированное предметом. Направление этой волны вследствие дифракции на краях голограммы лишь немного отличается от направления первой волны. Информации о фазе рассеянной объектом волны она также не несет. По существу, первую и вторую волны можно рассматривать как одну

дифрагировавшую волну нулевого порядка. Третья волна (3) повторяет с точностью до амплитудного множителя $\beta E_0 E_2 \exp(ikx \sin \theta)$ волну, распространяющуюся от объекта, и создает его мнимое изображение. Умножение на $\exp(ikx \sin \theta)$ вызывает фазовый сдвиг, выражающийся в отклонении изображения от оси голограммы в положительном направлении (против часовой стрелки) на угол θ . Четвертая волна (4) с комплексной амплитудой, соответствующей четвертому слагаемому в правой части соотношения (13.7), является комплексно-сопряженной объектной волне. Она образует действительное изображение объекта. Умножение на $\exp(-ikx \sin \theta)$ вызывает отклонение изображения от оси голограммы в отрицательном направлении (по часовой стрелке) на угол θ .

Комплексно сопряженная волна является точной копией исходной волны, но распространяющейся в направлении, соответствующем обратному ходу лучей при обращении времени, как бы возвращаясь к исходному трехмерному предмету. Она несет информацию, очень близкую к информации об объекте, но существенно отличающуюся от объектной волны обратной фазой. Поэтому действительное изображение обращено таким образом, что становится видимым изнутри, т.е. является псевдоскопическим или вывернутым. На нем выпуклые места предмета глаз видит как вогнутые и наоборот. Однако псевдоскопичность проявляться не будет, если действительное изображение спроецировать на экран. Заметим, что в рассматриваемом случае, когда направления опорного и восстанавливающего пучков не совпадают, действительное S' и мнимое S'' изображения расположены симметрично относительно плоскости H , перпендикулярной направлению восстанавливающего пучка и проходящей через центр голограммы. Если бы восстанавливающий пучок (0) был идентичен опорному пучку, использованному при записи голограммы, то положение действительного изображения

предмета было бы зеркально симметрично с положением самого предмета, а мнимое изображение находилось бы на том же месте, где находился предмет, и было полностью ему подобно.

Третья и четвертая волны – волны первого дифракционного порядка.

Четвертый член в выражении (13.7) и соответствующая ему комплексно-сопряженная волна, так же как и первый и второй члены и соответствующие им первая и вторая волны, появляются из-за квадратичного характера детектирования света. Хотя при восстановлении появляются два изображения, они разнесены в пространстве относительно друг друга, а также относительно когерентного фона (плоской волны) и не создают помех друг другу при их наблюдении.

Максимально интенсивное мнимое изображение получается, если при восстановлении голограмму осветить волной, совершенно идентичной той, которая была использована в качестве опорной при записи. Для получения правильного неискаженного действительного изображения и удобства его наблюдения нужно при восстановлении осветить голограмму волной, сопряженной той, которая была использована при записи голограммы. В этом случае имеем следующее распределение комплексной амплитуды в плоскости непосредственно за голограммой:

$$E'(x, y) = [t_0 + \beta E_1^2(x, y)]E_2 \exp(ikx \sin \theta) + \\ + \beta E_1(x, y)E_2^2 \exp(i2kx \sin \theta) \exp[i\varphi(x, y)] + \\ + \beta E_1(x, y)E_2^2 \exp[-i\varphi(x, y)],$$

и в третьем члене получаем с точностью до постоянного множителя βE_2^2 амплитуду комплексно-сопряженной предметной волны в чистом виде.

Качество голографического изображения во многом определяется разрешающей способностью голограмм. Мы видели, что голограмму можно рассматривать как дифракционную

решетку, поэтому критерий разрешения Аббе, а также вторую и третью формулы (12.4) можно использовать и в голографии при определении разрешения голограммы. Так, разрешимый интервал (предел разрешения) голограммы определится формулой $\delta x = \lambda z_0 / L$, а угловое разрешение – формулой $\delta \varphi = \lambda / L$. Здесь θ и L – соответственно угловой и линейный размеры голограммы, z_0 – расстояние от точечного источника до голограммы в процессе записи. В соответствии с этими формулами линейное разрешение голограммы зависит только от ее углового размера θ , а угловое разрешение – от линейного размера L . Следовательно, для увеличения разрешающей способности голограммы следует увеличивать ее размеры. Чтобы размер зерна фотоэмульсии не ограничивал разрешающей способности голограмм, специально для голографии разработаны мелкозернистые фотоэмульсии с разрешающей способностью несколько тысяч линий на 1 мм.

Отметим важное свойство голограмм: восстановить предметную волну можно не только с помощью всей голограммы, но и с помощью небольшого ее участка, так как в отличие от обычной фотографии, каждая точка которой соответствует некоторой точке объекта, на голограмме каждая точка соответствует всей плоскости объекта. Это связано с тем, что при записи голограммы свет, рассеянный каждой точкой предмета, падает на всю поверхность фотопластинки. Поэтому каждый участок голограммы содержит в закодированной форме информацию обо всех точках предмета, что и делает возможным с помощью части голограммы восстановить изображение всего объекта. Разрешающая способность при этом определяется размером этой части голограммы.

Следует отметить также, что, как уже отмечалось в п. 12.7, в результате облучения светочувствительный слой фотопластинки после ее химической обработки может изменить не только коэффициент пропускания T , но и коэффициент преломления n . Если происходит изменение коэффициента пропускания, голограмму называют амплитудной, а если коэффициента преломления – фазовой. Выше для простоты мы ограничивались

рассмотрением двумерных амплитудных голограмм. В амплитудной голограмме интерференционная картина фиксируется в изменениях ее коэффициента пропускания, а в прозрачной фазовой – в изменениях ее показателя преломления. Амплитудная голограмма приводит к пространственной модуляции амплитуды восстанавливающей волны, а фазовая – к пространственной модуляции фазы. Вследствие модуляции фазы, как и вследствие модуляции амплитуды, возникают дифрагировавшие волны, которые создают мнимое и действительное изображения предмета.

В качестве восстанавливающей можно использовать и волну с длиной, отличной от длины волны, использованной при записи. В этом случае линейный размер восстановленного изображения будет отличаться от соответствующего размера предмета и оно будет находиться на ином расстоянии от голограммы.

13.4. Голографическая схема Денисюка

В голографических схемах Габора и Лейта – Упатниекса, которые мы рассматривали выше, предполагается, что интерференционные полосы распределены по поверхности регистрирующего фотоматериала, поэтому на толщину фотоматериала не накладывалось никаких условий. Используемые в этих схемах голограммы, называют тонкими (или двумерными), так как в них применяются фотопластинки, толщина эмульсионного слоя которых меньше расстояния между интерференционными полосами регистрируемого оптического поля. Ю.Н. Денисюк предложил метод регистрации голограмм в толстослойных (называемых также трехмерными) эмульсиях, толщина которых больше расстояния между интерференционными полосами. Это позволяет получить трехмерную дифракционную решетку, при освещении которой можно восстановить не только изображение предмета, но и спектр излучения, в котором была записана голограмма. Последнее обстоятельство делает возможным при реконструкции

голограммы использовать некогерентный белый свет (т.е. без использования лазера; достаточно взять яркий источник света с малыми угловыми размерами).

Суть метода Денисюка можно уяснить на простом примере, когда предметной волной является плоская волна, распространяющаяся под углом θ к фотопластинке,

$$E_{\text{пр}} = E_1 \exp[ik(x \sin \theta + z \cos \theta)],$$

а опорной – плоская волна, падающая перпендикулярно к ней, $E_{\text{оп}} = E_2 \exp(ikz)$. Интерференция таких волн в объеме была рассмотрена нами в п. 3.2.1. Было показано, что интерференционная картина в этом случае представляет собой последовательность плоскостей максимальной и минимальной интенсивности. Расстояние d между плоскостями интерференционных максимумов определяется формулой (3.6)

$$2d \sin \alpha = \lambda, \quad (13.8)$$

где угол $\alpha = \theta/2$ (рис. 3.3). Если две волны распространяются в прозрачной фотоэмульсии (до фотообработки фотоэмульсия прозрачна), то в объеме фотоэмульсии в местах максимумов интенсивности после процесса проявления образуются элементарные полупрозрачные отражающие слои – зеркала из серебра, которые затем могут отражать восстанавливающую волну. Расстояние между этими слоями равно d . Места же, где интенсивность минимальна, после обработки будут полностью прозрачны. Эта система параллельных полупрозрачных отражающих слоев металлического серебра и образует пространственную дифракционную решетку с периодом d .

Если толщина L фоторегистрирующей среды много больше половины расстояния l между плоскостями, при переходе между которыми сдвиг интерференционных полос происходит на ширину полосы, то луч света, распространяющийся вдоль оси Z , много раз пересекает полупрозрачные слои, частично проходя сквозь каждый слой, а частично отражаясь от него. В этом случае

толщина фотослоя играет важную роль. Значение расстояния l между слоями, при котором происходит сдвиг интерференционных полос на ширину полосы, полученное в п. 3.2.1, равно $l = \lambda / (1 - \cos \theta)$. Поэтому условие, при котором толщина фотослоя существенна, можно записать в виде

$$L > \lambda / (1 - \cos \theta). \quad (13.9)$$

Если же толщина L фотослоя много меньше указанного расстояния l , то она никак не будет проявляться, поскольку в этом случае

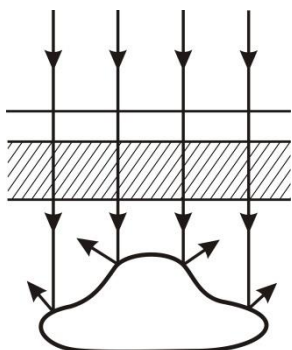


Рис. 13.13

световой луч, проходя сквозь прозрачный фотослой, не будет пересекать слои различной прозрачности.

Из условия (13.9) видно, что один и тот же фотослой можно считать толстым или тонким в зависимости от величины угла схождения волн θ . Если этот угол мал, то $l \approx 2\lambda / \theta$ велико, толщина данного фотослоя оказывается много меньше l .

Голограмма в этом случае получается тонкой. Именно это имеет место в голографической схеме Лейта – Упатниекса, где угол схождения волн порядка 0,1 радиана. При таком угле параметр $l \approx 100$ мкм, в то время как толщина голографических фотослоев обычно 10 – 20 мкм. Условие (13.8) при этом не выполняется, поэтому такую голограмму можно считать тонкой. С другой стороны, если угол схождения волн $\theta \approx 180^\circ$, т.е. 3,14 радиан, то $l \approx 0,01$ мкм. В этом случае условие (13.8) выполняется и получающаяся голограмма является объемной. Расстояние между отражающими слоями при этом

$$d = \lambda / 2\sin(\theta / 2) \approx \lambda / 2 \approx 0,25 \text{ мкм},$$

а на толщине голограммы уложится примерно 40 – 80 отражающих слоев, располагающихся почти параллельно краям

фотопластинки. Эта ситуация и реализуется в методе Ю.Н. Денисюка.

Голографическая схема Денисюка показана на рис. 13.13. Плоская опорная волна лазерного излучения падает на фотопластинку со стороны стекла и, пройдя через фотослой, освещает предмет. Рассеянные предметом волны, распространяются почти навстречу опорной волне (поэтому схему Денисюка обычно называют схемой на встречных пучках). При интерференции предметной и опорной волн в объеме фотоэмульсии образуются стоячие волны. Вследствие этого в объеме фотоэмульсии возникают системы параллельных отражающих слоев металлического серебра, образующих пространственную дифракционную решетку. Эти слои располагаются параллельно биссектрисе угла между направлениями предметной и опорной волн (рис. 13.14).

На стадии реконструкции полученную таким образом голограмму освещают плоской волной, распространяющейся в том же направлении, что и волна, служившая опорной волной на стадии формирования голограммы. Восстанавливающая волна, частично отражаясь от каждого зеркального слоя и частично проходя сквозь него,

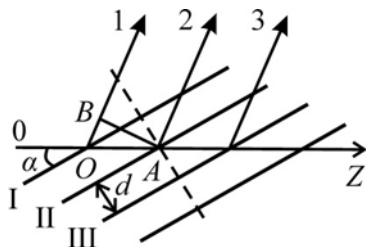


Рис. 13.14

формирует результирующую отраженную волну, которая создает мнимое изображение предмета. Отраженные от каждого слоя волны усиливают друг друга вследствие интерференции, если они удовлетворяют формуле Вульфа – Брэгга

$$2d \sin \alpha = m\lambda', \quad (13.10)$$

где α – угол скольжения падающего на отражающие поверхности восстанавливающего светового луча, λ' – длина волны, используемой при восстановлении изображения.

Рассмотрим процесс реконструкции изображения на том же примере с прямой опорной и наклонной предметной плоскими волнами. На рис. 13.14 показаны три зеркально отражающих слоя I , II , III и восстанавливающий луч света O , распространяющийся вдоль оси Z под углом α к этим слоям. В результате последовательного частичного отражения этого луча от каждого зеркального слоя и частичного прохождения сквозь него образуются отраженные волны 1, 2, 3. Видим, что восстанавливающая волна отражается в направлении, точно совпадающим с направлением распространения предметной волны на стадии записи голограммы (она падала на фотопластинку под углом $\theta = 2\alpha$ к нормали, совпадающей с направлением оси Z , как показано на рис. 3.3). Разность хода волн, отраженных от двух каких-либо плоскостей, например, волн 1 и 2, отраженных от плоскостей I , и II , как видно на рис. 13.14, $\Delta = OA - OB$. Из рис. 13.14 видно, что

$$OA = d / \sin \alpha, \angle BOA = 2\alpha, OB = OA \cos 2\alpha.$$

С учетом этого находим

$$\Delta = OA - OB = OA(1 - \cos 2\alpha) = \frac{d}{\sin \alpha} 2 \sin^2 \alpha = 2d \sin \alpha.$$

Точки A и B лежат на одной волновой поверхности AB , поэтому колебания в этих точках, созданные отраженными волнами 1 и 2, являются синфазными, а значит, полученная выше оптическая разность хода волн Δ равна длине волны λ' , используемой при восстановлении изображения. В этом случае интенсивность отраженных волн будет максимальной. Так что имеем равенство $2d \sin \alpha = \lambda'$. Сравнивая эту формулу с формулой (13.8), видим, что $\lambda' = \lambda$. Следовательно, сильное отражение плоской восстанавливающей волны возникает лишь на длине волны λ' , совпадающей с длиной опорной волны λ . Результирующая отраженная волна и представляет собой восстановленную предметную волну.

Таким образом, восстановление волнового фронта произойдет только в том случае, если оно осуществляется при той же длине волны, при которой производилась голографическая запись на фотопластинку. Это позволяет в качестве источника, восстанавливающего изображение света использовать источник сплошного спектра, например солнечный свет. Если восстанавливать голограмму белым светом, то объемная решетка «выбирает» из белого света лишь ту спектральную составляющую с длиной волны λ , которая использовалась при записи. Это свойство толстослойной голограммы используется для записи голограмм художественных произведений.

Известно, что цветовой эффект можно получить сочетаниями трех основных цветов, например красным, зеленым и синим, при соответствующим образом подобранных интенсивностях. Поэтому если объемную голограмму экспонировать в красном, зеленом и синем цветах, то каждая длина волны образует свою систему полупрозрачных отражающих плоскостей и при восстановлении белым светом волна отразится от совокупности «своих» плоскостей. В результате получится цветное, объемное изображение предмета. При этом можно добиться очень хорошей цветопередачи при восстановлении изображения белым светом.

Поскольку отраженные волны распространяются в том же направлении, что и предметная волна при записи голограммы, то толстослойная голограмма восстанавливает лишь одно (мнимое) изображение предмета. Для получения действительного изображения предмета следует использовать комплексно сопряженную плоскую восстанавливающую волну. Такая волна распространяется в противоположном направлении (так как изменение знака перед i равносильно изменению направления распространения волны), а значит, в направлении, противоположном опорной волне. Поэтому для получения действительного изображения восстанавливающий пучок должен освещать голограмму с обратной стороны. При этом восстановленная волна имеет ту же форму волновой поверхности, что и предметная волна, но с противоположным направлением

распространения. Изображение получается в том же месте, где находился предмет.

Следует отметить, что при большом числе отражающих слоев, образующих пространственную решетку, интенсивность восстановленной волны оказывается близкой к интенсивности волны, падающей на голограмму при реконструкции изображения.

13.5. Геометрия формирования голограмм и восстановления изображений

13.5.1. Точечный объект

Вернемся снова к точечному объекту и рассмотрим общий случай, когда опорная и восстанавливающая волны являются сферическими. Пусть точечный источник S_1 расположен в точке $(x_1, y_1, -a)$. Этот источник создает в плоскости фотопластинки световое поле, комплексная амплитуда которого в парааксиальном приближении представится в виде

$$E_{\text{пр}} = E_1 \exp \left\{ i \frac{k}{2a} [(x - x_1)^2 + (y - y_1)^2] \right\},$$

где $E_1 = E_{01} \exp(ika)/a$ – постоянный комплексный множитель.

Сферическая опорная волна создается точечным источником S_2 , расположенным в точке $(x_2, y_2, -b)$. Комплексная амплитуда поля, создаваемая этим источником в плоскости фотопластинки в том же парааксиальном приближении определится как

$$E_{\text{оп}} = E_2 \exp \left\{ i \frac{k}{2b} [(x - x_2)^2 + (y - y_2)^2] \right\},$$

где комплексная постоянная $E_2 = E_{02} \exp(ikb)/b$. Полная комплексная амплитуда светового поля в точке (x, y) фотопластинки будет равна

$$E(x, y) = E_1 \exp \left\{ i \frac{k}{2a} [(x-x_1)^2 + (y-y_1)^2] \right\} + E_2 \exp \left\{ i \frac{k}{2b} [(x-x_2)^2 + (y-y_2)^2] \right\},$$

а интенсивность в этой точке равна

$$I(x, y) = |E_1|^2 + |E_2|^2 + E_1 E_2 \exp \left\{ i \frac{k}{2a} [(x-x_1)^2 + (y-y_1)^2] \right\} \times \exp \left\{ i \frac{k}{2a} [(x-x_1)^2 + (y-y_1)^2] \right\} \times \exp \left\{ -i \frac{k}{2b} [(x-x_2)^2 + (y-y_2)^2] \right\} + E_1 E_2 \exp \left\{ -i \frac{k}{2a} [(x-x_1)^2 + (y-y_1)^2] \right\} \times \exp \left\{ i \frac{k}{2b} [(x-x_2)^2 + (y-y_2)^2] \right\}.$$

Амплитудное пропускание $t(x, y)$ голограммы, полученной после проявления и закрепления фотопластинки, будет определяться выражением

$$t(x, y) = t_0 + \beta E_1 E_2 \exp \left\{ i \frac{k}{2a} [(x-x_1)^2 + (y-y_1)^2] \right\} \times \exp \left\{ -i \frac{k}{2b} [(x-x_2)^2 + (y-y_2)^2] \right\} + \beta E_1 E_2 \exp \left\{ -i \frac{k}{2a} [(x-x_1)^2 + (y-y_1)^2] \right\} \times$$

$$\times \exp \left\{ i \frac{k}{2b} [(x-x_2)^2 + (y-y_2)^2] \right\}. \quad (13.11)$$

Как и прежде, нас будут интересовать два последних члена, так как именно они описывают изображения точечного источника. Назовем изображение, соответствующее первому из этих членов, основным, или истинным изображением, а второму – сопряженным. Каждое из этих изображений может быть как действительным, так и мнимым.

На этапе восстановления голограмму освещают сферической волной от точечного источника S_3 , расположенного в точке $(x_3, y_3, -c)$. Комплексную амплитуду восстанавливающей волны в плоскости голограммы в параксиальном приближении можно представить в виде

$$E_b = E_3 \exp \left\{ i \frac{k}{2c} [(x-x_3)^2 + (y-y_3)^2] \right\},$$

где E_3 – комплексный множитель, имеющий тот же вид, что и E_1 , и E_2 . Положение точек S_1 , S_2 и S_3 может быть произвольным. Они могут быть действительными или мнимыми. Эти точки могут находиться на конечных расстояниях от голограммы или в бесконечности.

Рассмотрим сначала основное изображение точечного источника S_1 . Умножив второй член выражения (13.11) на E_b , получим

$$\begin{aligned} & \beta E_1 E_2 E_3 \exp \left\{ i \frac{k}{2a} [(x-x_1)^2 + (y-y_1)^2] \right\} \times \\ & \quad \times \exp \left\{ -i \frac{k}{2b} [(x-x_2)^2 + (y-y_2)^2] \right\} \times \\ & \quad \times \exp \left\{ i \frac{k}{2c} [(x-x_3)^2 + (y-y_3)^2] \right\}. \end{aligned} \quad (13.12)$$

Это выражение в параксиальном приближении описывает сферическую волну, сходящуюся или расходящуюся к некоторому центру кривизны S' (в зависимости от соотношения между величинами a , b и c), в котором будет располагаться действительное или мнимое изображение точечного источника. Чтобы найти расстояние d_1 от этого изображения до голограммы, достаточно сравнить это выражение с общим выражением для сферической волны, испускаемой точечным источником, находящимся на расстоянии d_1 от голограммы. В том же параксиальном приближении сферическая волна запишется в виде

$$\exp\left[i\frac{k}{2d_1}(x^2 + y^2)\right]. \quad (13.13)$$

Выделим в (13.12) множитель, содержащий в показателе двучлен $(x^2 + y^2)$. Указанный множитель имеет вид

$$\exp\left\{i\frac{k}{2}\left[\frac{1}{a} - \frac{1}{b} + \frac{1}{c}\right](x^2 + y^2)\right\}. \quad (13.14)$$

Сравнивая (13.13) и (13.14), получим выражение для расстояния d_1 от основного изображения до голограммы:

$$\frac{1}{d_1} = \frac{1}{a} - \frac{1}{b} + \frac{1}{c}. \quad (13.15)$$

Сопряженное изображение описывается выражением

$$\beta E_1 E_2 E_3 \exp\left\{-i\frac{k}{2a}\left[(x-x_1)^2 + (y-y_1)^2\right]\right\} \times \\ \times \exp\left\{i\frac{k}{2b}\left[(x-x_2)^2 + (y-y_2)^2\right]\right\} \times$$

$$\times \exp \left\{ i \frac{k}{2c} [(x - x_3)^2 + (y - y_3)^2] \right\},$$

которое представляет собой амплитуду сферической волны, сходящейся или расходящейся к своему центру кривизны S'' , где так же образуется действительное или мнимое изображение точечного объекта. Для определения расстояния d_2 от этого изображения до голограммы, выделим в этом выражении сомножитель, содержащий в показателе $(x^2 + y^2)$, и сравним его с (13.13). Тогда получим

$$\frac{1}{d_2} = -\frac{1}{a} + \frac{1}{b} + \frac{1}{c}. \quad (13.16)$$

Выражения (13.15) и (13.16) являются формулами отображения, аналогичными выражениям для линзовых систем. Они связывают между собой расстояния до голограммы от изображения, предмета, источника опорной и восстанавливающих волн. Если d_1 положительно, то изображение в точке S' будет действительным; при отрицательном d_1 изображение в точке S' мнимое. То же самое относится и к изображению в точке S'' . Выражения (13.15) и (13.16) справедливы и в том случае, когда в качестве опорной или восстанавливающей волны берутся нормально падающие плоские волны. В этом случае либо b , либо c , либо b и c одновременно следует положить равными бесконечности. Например, когда опорная и восстанавливающая волны являются плоскими, $b = c = \infty$, то $d_1 = -d_2 = a$. Изображения в точках S' и S'' в этом случае будут симметричными относительно плоскости, проходящей через голограмму перпендикулярно направлению распространения восстанавливающей волны. Если одна только восстанавливающая волна является плоской, то положения главного и сопряженного изображений будут определяться формулами

$$\frac{1}{d_1} = \frac{1}{a} - \frac{1}{b}, \quad \frac{1}{d_2} = -\frac{1}{a} + \frac{1}{b}.$$

В этом случае изображение в точке S' будет действительным при $a < b$ и мнимым – при $a > b$. Для изображения в точке S'' имеют место обратные соотношения.

13.5.2. Плоский предмет

Предположим, что плоский предмет – транспарант S – освещается монохроматической плоской волной с амплитудой E_0 . Предмет находится на расстоянии a от фотопластинки (рис. 13.15).

Свяжем с плоскостью предмета координатную плоскость XU . Ось Z направим в сторону фотопластинки. Координаты в плоскости фотопластинки будем обозначать через ξ, η . Функцию амплитудного пропускания транспаранта обозначим через $t_1(x, y)$. Тогда комплексная амплитуда поля в плоскости непосредственно за плоскостью предмета будет $E_0 t_1(x, y)$. Распределение поля в плоскости фотопластинки ($z = a$), создаваемое предметом, определится преобразованием Френеля

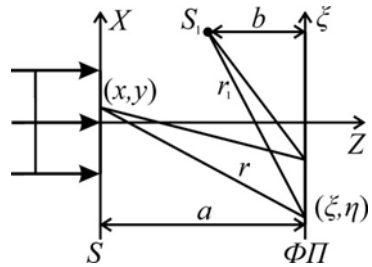


Рис. 13.15

$$E_{\text{фп}}(\xi, \eta) = A \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[ik \frac{(\xi - x)^2 + (\eta - y)^2}{2a} \right] dx dy,$$

где $A = \exp(ika)/(i\lambda a)$ – комплексная постоянная, $E(x, y) = E_0 t_1(x, y)$. Опорной волной является сферическая волна,

испускаемая точечным источником S_1 , расположенным на оси Z на расстоянии b от фотопластинки в точке $S(x_1, y_1, a - b)$. Комплексная амплитуда, создаваемая этой волной в плоскости фотопластинки

$$E_{\text{он}}(x, y) = \frac{E_{02}}{r} \exp(ikr_1),$$

где $r_1 = \sqrt{(\xi - x_1)^2 + (\eta - y_1)^2 + b^2}$ – расстояние от источника опорной волны до произвольной точки (ξ, η) плоскости фотопластинки. В параксиальном приближении

$$E_{\text{он}}(x, y) = B \exp \left[ik \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} \right], \quad (13.17)$$

где $B = E_{02} \exp(ikb) / b$ – комплексная постоянная.

Результирующее поле в плоскости фотопластинки определится суммой комплексных амплитуд объектной и опорной волн

$$E(x, y) = A \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp \left[ik \frac{(\xi - x)^2 + (\eta - y)^2}{2a} \right] dx dy +$$

$$+ B \exp \left[ik \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} \right],$$

а интенсивность

$$I(\xi, \eta) = \left| A \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp \left[ik \frac{(\xi - x)^2 + (\eta - y)^2}{2a} \right] dx dy + \right.$$

$$+ B \exp \left[ik \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} \right]^2.$$

Отсюда получаем

$$\begin{aligned} I(\xi, \eta) = & B^2 + A \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[ik \frac{(\xi - x)^2 + (\eta - y)^2}{2a} \right] dx dy \times \\ & \times A^* \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E^*(x, y) \exp \left[-ik \frac{(\xi - x)^2 + (\eta - y)^2}{2a} \right] dx dy + \\ & + AB \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left\{ ik \left[\frac{(\xi - x)^2 + (\eta - y)^2}{2a} - \right. \right. \\ & \left. \left. - \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} \right] \right\} dx dy + \\ & + AB \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E^*(x, y) \exp \left\{ -ik \left[\frac{(\xi - x)^2 + (\eta - y)^2}{2a} - \right. \right. \\ & \left. \left. - \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} \right] \right\} dx dy. \end{aligned}$$

Это распределение интенсивности будет зарегистрировано светочувствительным слоем фотопластинки. Считая, что линейная зависимость амплитудной функции записи от интенсивности выполняется, и принимая, что $I_0 = |B|^2$, получим для амплитудной функции пропускания голограммы

$$t(\xi, \eta) = t_0 + \beta A \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[ik \frac{(\xi - x)^2 + (\eta - y)^2}{2a} \right] dx dy \times$$

$$\begin{aligned}
& \times A \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E^*(x, y) \exp \left[-ik \frac{(\zeta - x)^2 + (\eta - y)^2}{2a} \right] dx dy + \\
& + \beta AB \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left\{ ik \left[\frac{(\zeta - x)^2 + (\eta - y)^2}{2a} - \right. \right. \\
& \left. \left. - \frac{(\zeta - x_1)^2 + (\eta - y_1)^2}{2b} \right] \right\} dx dy + \\
& + \beta AB \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E^*(x, y) \exp \left\{ -ik \left[\frac{(\zeta - x)^2 + (\eta - y)^2}{2a} - \right. \right. \\
& \left. \left. - \frac{(\zeta - x_1)^2 + (\eta - y_1)^2}{2b} \right] \right\} dx dy .
\end{aligned}$$

Первый член этого выражения не несет никакой информации о предмете. Второй член представляет собой автокорреляцию распределения поля на предмете. Третий и четвертый члены передают информацию о фазе прошедшей через предмет волны и о комплексной функции амплитудного пропускания предмета. По-

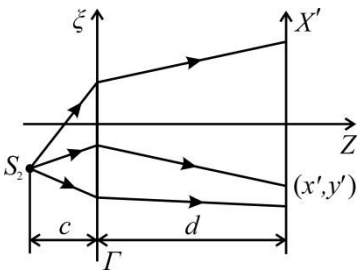


Рис. 13.16

лученная функция амплитудного пропускания голограммы будет определять результат дифракции восстанавливающей когерентной волны на голограмме на стадии реконструкции (восстановления). В качестве восстанавливающей волны возьмем сферическую волну, исходящую из точечного источника S_2 , находящегося на

расстоянии c от голограммы в точке $S_2(x_2, y_2, a - c)$ (рис. 13.16). Эта волна создает в плоскости голограммы поле с комплексной амплитудой

$$E_B = \frac{E_{03}}{r} \exp(ikr_2),$$

где $r_2 = \sqrt{(\xi - x_2)^2 + (\eta - y_2)^2 + c^2}$ – расстояние от источника восстанавливающей волны до точки (ξ, η) плоскости голограммы. В параксиальном приближении

$$E_B(\xi, \eta) = C \exp \left[ik \frac{(\xi - x_2)^2 + (\eta - y_2)^2}{2c} \right], \quad (13.18)$$

где $C = E_{03} \exp(ikc)/c$ – комплексная постоянная.

Распространение поля после дифракции на голограмме до плоскости наблюдения XY' также можно описать преобразованием Френеля. Имеем для поля в плоскости XY'

$$E'(x', y') = D \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E_2(\xi, \eta) \exp \left[ik \frac{(x' - \xi)^2 + (y' - \eta)^2}{2d} \right] d\xi d\eta,$$

где $D = \exp(ikd)/(i\lambda d)$ – комплексная постоянная, d – расстояние от плоскости наблюдения XY' до голограммы, $E_2(\xi, \eta) = E_B(\xi, \eta)t(\xi, \eta)$ – поле в плоскости непосредственно за плоскостью голограммы, или, подставив вместо $E_B(\xi, \eta)$,

$$E'(x', y') = CD \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} t(\xi, \eta) \times \\ \times \exp \left\{ ik \left[\frac{(\xi - x_2)^2 + (\eta - y_2)^2}{2c} + \frac{(x' - \xi)^2 + (y' - \eta)^2}{2d} \right] \right\} d\xi d\eta.$$

Подставив в подынтегральное выражение амплитудную функцию пропускания голограммы, получим четыре члена, из которых только два последних содержат информацию о фазе предметной волны и описывают восстановленное голографическое изображение. Поэтому только эти два последних члена представляют для нас интерес. При этом, как и в случае точечного предмета, первый из этих членов описывает основное изображение, а второй – сопряженное. Рассмотрим сначала третий член. Он имеет вид

$$E_3'(x', y') = F \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left\{ ik \left[\frac{(\zeta - x)^2 + (\eta - y)^2}{2a} - \frac{(\zeta - x_1)^2 + (\eta - y_1)^2}{2b} + \frac{(\zeta - x_2)^2 + (\eta - y_2)^2}{2c} + \frac{(x' - \zeta)^2 + (y' - \eta)^2}{2d} \right] \right\} d\zeta d\eta dx dy,$$

где

$$F = \beta ABCD = -E_{02} E_{03} \exp[ik(a + b + c + d) / (\lambda^2 abcd)]$$

– постоянная. Раскрыв скобки и сгруппировав соответствующие члены, получим

$$E_3'(x', y') = F \exp \left[-i \frac{k}{2} \left(\frac{x_1^2 + y_1^2}{b} - \frac{x_2^2 + y_2^2}{c} \right) \right] \times \\ \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[i \frac{k}{2} \left(\frac{x^2 + y^2}{a} + \frac{x'^2 + y'^2}{d} \right) \right] \times \\ \times \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left[i \frac{k}{2} \left(\frac{1}{a} - \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right) (\zeta^2 + \eta^2) \right] \right) \times$$

$$\times \exp \left\{ ik \left[\left(-\frac{x}{a} + \frac{x_1}{b} - \frac{x_2}{c} - \frac{x'}{d} \right) \xi + \left(-\frac{y}{a} + \frac{y_1}{b} - \frac{y_2}{c} - \frac{y'}{d} \right) \eta \right] \right\} d\xi d\eta \Big] dx dy. \quad (13.19)$$

Полученное выражение описывает амплитуду дифрагировавшей волны, появление которой обусловлено той частью структуры голограммы, которая приводит к возникновению основного изображения. Плоскость, на которой возникает дифракционная картина, располагается на произвольном расстоянии d от голограммы. Резкое изображение без дифракционных эффектов получается, если дифракционную картину наблюдать на расстоянии d_1 , при котором исчезает квадратичный фазовый множитель по переменным ξ и η в выражении (13.19). Для этого необходимо выполнение следующего условия:

$$\frac{1}{a} - \frac{1}{b} + \frac{1}{c} + \frac{1}{d_1} = 0.$$

Отсюда получаем выражение для обратного расстояния основного изображения от голограммы:

$$\frac{1}{d_1} = -\frac{1}{a} + \frac{1}{b} - \frac{1}{c}. \quad (13.20)$$

Оставшийся внутренний интеграл по переменным ξ и η

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left\{ \left[\left(-\frac{x}{a} + \frac{x_1}{b} - \frac{x_2}{c} - \frac{x'}{d_1} \right) \xi + \left(-\frac{y}{a} + \frac{y_1}{b} - \frac{y_2}{c} - \frac{y'}{d_1} \right) \eta \right] \right\} d\xi d\eta =$$

$$= \int_{-\infty}^{\infty} \exp \left[-ik \left(\frac{x}{a} - \frac{x_1}{b} + \frac{x_2}{c} + \frac{x'}{d_1} \right) \zeta \right] d\zeta \times \\ \times \int_{-\infty}^{\infty} \exp \left[-ik \left(\frac{y}{a} - \frac{y_1}{b} + \frac{y_2}{c} + \frac{y'}{d_1} \right) \eta \right] d\eta$$

представляет собой одну из форм двумерной дельта-функции Дирака:

$$\delta \left(\frac{x}{a} - \frac{x_1}{b} + \frac{x_2}{c} + \frac{x'}{d_1}; \frac{y}{a} - \frac{y_1}{b} + \frac{y_2}{c} + \frac{y'}{d_1} \right).$$

С учетом этого находим

$$E_3'(x', y') = F \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[i \frac{k}{2} \left(\frac{x^2 + y^2}{a} + \frac{x'^2 + y'^2}{d_1} \right) \right] \times \\ \times \delta \left(\frac{x}{a} - \frac{x_1}{b} + \frac{x_2}{c} + \frac{x'}{d_1}; \frac{y}{a} - \frac{y_1}{b} + \frac{y_2}{c} + \frac{y'}{d_1} \right) dx dy.$$

Используя фильтрующее свойство дельта-функции, т.е. подставив в подынтегральное выражение

$$x = \frac{a}{b} x_1 - \frac{a}{c} x_2 - \frac{a}{d_1} x', \quad y = \frac{a}{b} y_1 - \frac{a}{c} y_2 - \frac{a}{d_1} y',$$

будем иметь

$$E_3'(x', y') = AE \left[a \left(\frac{x_1}{b} - \frac{x_2}{c} \right) - \frac{a}{d_1} x'; a \left(\frac{y_1}{b} - \frac{y_2}{c} \right) - \frac{a}{d_1} y' \right], \quad (13.21)$$

где через A обозначено произведение всех остальных множителей.

Из выражения (13.21) видно, что основное изображение подобно самому предмету, но имеет измененный масштаб, определяемый отношением d_1 / a , а также смещено на величину $a\left(\frac{x_1}{b} - \frac{x_2}{c}\right)$ вдоль оси X' и на величину $a\left(\frac{y_1}{b} - \frac{y_2}{c}\right)$ вдоль оси Y' . Величины смещения определяются углами между осявыми лучами опорного и восстанавливающего пучков. В зависимости от знака отношения d_1 / a основное голографическое изображение будет прямым (при $d_1 / a < 0$) или перевернутым в плоскости изображения (при $d_1 / a > 0$). Кроме того, голографическое изображение будет мнимым при $d_1 < 0$ и действительным при $d_1 > 0$.

Рассмотрим, например, практически важный случай, когда источник восстанавливающей волны помещен в точку расположения источника опорной волны. Тогда, как видно из формулы (13.20) $d_1 = -a < 0$. Следовательно, изображение в этом случае будет мнимым, расположенным перед голограммой в том месте, где располагался предмет при записи голограммы (в плоскости предмета). Из выражения (13.21) следует, что при этом

$$E_3'(x', y') = AE(x', y'),$$

т.е. изображение будет прямым, с размерами, равными размерам предмета. Смещение изображения в перпендикулярном к оси Z направлении будет отсутствовать. Те же результаты получаются и в другом практически важном случае, когда опорная и восстанавливающая волны будут плоскими ($b = c = \infty$).

Изображение будет располагаться на оси Z и тогда, когда восстанавливающий источник будет находиться на этой оси между предметом и его голограммой, так как и в этом случае $x_1 / a = x_2 / b$, $y_1 / a = y_2 / b$ и выражения в скобках в (13.21) обращаются в нуль.

Рассмотрим теперь четвертый член выражения для $E'(x', y')$, описывающий сопряженное изображение. Этот член имеет вид

$$E_4'(x', y') = F \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E^*(x, y) \exp \left\{ ik \left[-\frac{(\xi - x)^2 + (\eta - y)^2}{2a} + \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} + \frac{(\xi - x_2)^2 + (\eta - y_2)^2}{2c} + \frac{(x' - \xi)^2 + (y' - \eta)^2}{2d} \right] \right\} d\xi d\eta dx dy.$$

Он отличается от третьего члена только знаком в первом и втором членах в показателе экспоненты. Проведя преобразования, подобные предыдущим, получим, что расстояние d_2 , на котором будет находиться резкое сопряженное изображение, будет определяться формулой

$$\frac{1}{d_2} = \frac{1}{a} - \frac{1}{b} - \frac{1}{c}.$$

Распределение поля в плоскости изображения в этом случае будет определяться функцией

$$E_4'(x', y') = BE^* \left[a \left(\frac{x_1}{b} + \frac{x_2}{c} \right) + \frac{a}{d_2} x'; a \left(\frac{y_1}{b} + \frac{y_2}{c} \right) + \frac{a}{d_2} y' \right],$$

где через B обозначено произведение всех остальных множителей. Как видим, сопряженное изображение является прямым, смещенным вдоль координатных осей X' и Y' на расстояния соответственно $a \left(\frac{x_1}{b} + \frac{x_2}{c} \right)$ и $a \left(\frac{y_1}{b} + \frac{y_2}{c} \right)$. Масштаб сопряженного изображения изменен в отношении d_2 / a . В случае плоских опорной и восстанавливающей волн сопряженное

изображение будет располагаться в плоскости $d_2 = a$ и описываться функцией $E_4'(x', y') = BE^*(x', y')$. Изображение будет прямым и действительным с размерами, равными размерам предмета. Оно будет располагаться за голограммой симметрично плоскости предмета. В случае, когда источник восстанавливающей волны расположен в точке, где располагался источник опорной волны при записи голограммы, сопряженное изображение также будет прямым, действительным и располагаться на том же расстоянии $d_2 = a$ от голограммы, но оно будет смещено по осям X' и Y' на величины $2\frac{a}{b}x_1$ и $2\frac{a}{b}y_1$ соответственно. Такое расположение сопряженного изображения не создает помех при наблюдении основного изображения.

13.6. Голография Фурье

13.6.1. Голография Фурье с использованием линз

Голография, которую мы до сих пор рассматривали, называется голографией Френеля. В этом случае фотопластинка располагается в ближней зоне, т.е. в области дифракции Френеля, и от каждой точки объекта распространяется сферическая волна. Рассмотрим теперь так называемую голографию Фурье, или голографию Фраунгофера, когда голограмма освещается параллельным пучком, формируемым бесконечно удаленным предметом (т.е. с помощью линзы), а фотопластинка устанавливается в дальней зоне – в области дифракции Фраунгофера (или в задней фокальной плоскости линзы).

Как уже многократно отмечалось, если за предметом, освещаемым когерентным светом, расположить собирающую линзу, то в ее задней фокальной плоскости образуется фурье-спектр распределения комплексной амплитуды поля в плоскости предмета. При этом точное, без искажающих квадратичных фазовых множителей преобразование Фурье получается, если

предмет располагается в передней фокальной плоскости линзы. Если поместить в заднюю

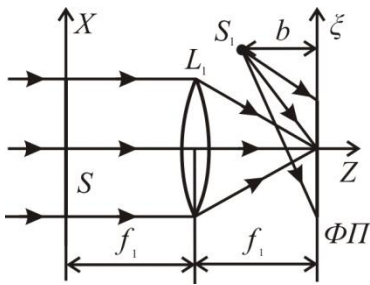


Рис. 13.17

плоскость XU . Распределение комплексной амплитуды поля в плоскости предмета обозначим через $E(x, y)$. Тогда распределение поля в задней фокальной плоскости линзы (рис. 13.17) представится преобразованием Фурье:

$$E_1(\xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[-ik \left(\frac{\xi}{f_1} x + \frac{\eta}{f_1} y \right) \right] dx dy ,$$

где ξ, η – координаты в плоскости фотопластинки, помещенной в задней фокальной плоскости линзы, f_1 – фокусное расстояние линзы. При одновременном освещении фотопластинки когерентной сферической волной от точечного источника S_1 , находящегося в точке на расстоянии b от фотопластинки, фотопластинка зафиксирует распределение интенсивности

$$I(\xi, \eta) = \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[-ik \left(\frac{\xi}{f_1} x + \frac{\eta}{f_1} y \right) \right] dx dy + B \exp \left[ik \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} \right] \right|^2 ,$$

фокальную плоскость линзы фотопластинку и одновременно осветить ее когерентной опорной волной, то после экспонирования и химической обработки фотопластинки получим голограмму, которую называют голограммой Фурье.

Совместим с плоскостью предмета S координатную

где второй член под модулем представляет собой сферическую опорную волну в параксиальном приближении, определяемую выражением (13.17); x_1 и y_1 – координаты точечного источника S_1 опорной волны; постоянная B определяется тем же выражением (13.17).

Амплитудное пропускание голограммы

$$\begin{aligned}
 t(\xi, \eta) = & t_0 + \beta \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp \left[-ik \left(\frac{\xi}{f_1} x + \frac{\eta}{f_1} y \right) \right] dx dy \times \\
 & \times \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp \left[ik \left(\frac{\xi}{f_1} x + \frac{\eta}{f_1} y \right) \right] dx dy + \\
 & + \beta B \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp \left[-\frac{x\xi + y\eta}{f_1} - \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} \right] dx dy + \\
 & + \beta B \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp \left[\frac{x\xi + y\eta}{f_1} + \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} \right] dx dy .
 \end{aligned}$$

Амплитудное пропускание голограммы для основного изображения определяется третьим членом

$$\begin{aligned}
 t_3(\xi, \eta) = & \\
 = & \beta B \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp \left[-\frac{x\xi + y\eta}{f_1} - \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} \right] dx dy .
 \end{aligned}$$

На этапе восстановления (рис. 13.18) голограмма помещается в передней фокальной плоскости другой линзы с фокусным расстоянием f_2 и освещается восстанавливающей волной от точечного источника S_2 , находящегося на расстоянии s от голограммы в точке с координатами $x = x_2, y = y_2$.

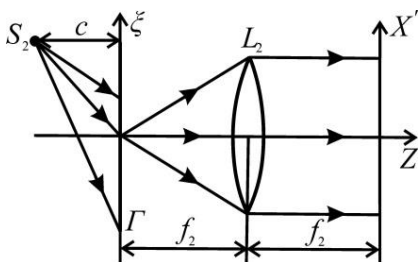


Рис. 13.18

В параксиальном приближении эта волна представляется выражением (13.18). В задней фокальной плоскости этой линзы (плоскости изображения $X'Y'$) образуется распределение поля, являющееся преобразованием Фурье

распределения поля в плоскости голограммы:

$$E_3'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E_3(\xi, \eta) \exp \left[-ik \left(\frac{x'}{f_2} \xi + \frac{y'}{f_2} \eta \right) \right] d\xi d\eta, \quad (13.22)$$

где $E_3(\xi, \eta) = E_b(\xi, \eta)t_3(\xi, \eta)$ – распределение поля в плоскости голограммы, соответствующее основному изображению, x', y' – координаты в задней фокальной плоскости линзы. Подставляя в соотношение (13.22) выражения $E_b(\xi, \eta)$ и $t_3(\xi, \eta)$, получим

$$E_3'(x', y') = \beta B \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left\{ ik \left[-\frac{x\xi + y\eta}{f_1} - \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} + \frac{(\xi - x_2)^2 + (\eta - y_2)^2}{2c} - \frac{x'\xi + y'\eta}{f_2} \right] \right\} d\xi d\eta dx dy.$$

Раскрыв скобки и сгруппировав члены, приведем это выражение к виду

$$\begin{aligned}
E_3'(x', y') = & \beta BC \exp \left[-i \frac{k}{2} \left(\frac{x_1^2 + y_1^2}{b} - \frac{x_2^2 + y_2^2}{c} \right) \right] \times \\
& \times \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \left(\int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} \exp \left[-i \frac{k}{2} \left(\frac{1}{b} - \frac{1}{c} \right) (\xi^2 + \eta^2) \right] \times \right. \\
& \times \left. \left\{ -ik \left[\left(\frac{x}{f_1} - \frac{x_1}{b} + \frac{x_2}{c} + \frac{x'}{f_2} \right) \xi + \right. \right. \right. \\
& \left. \left. \left. + \left(\frac{y}{f_1} - \frac{y_1}{b} + \frac{y_2}{c} + \frac{y'}{f_2} \right) \eta \right] \right\} d\xi d\eta \right) dx dy.
\end{aligned}$$

Условием образования изображения в плоскости $X'Y'$ является равенство единице квадратичного фазового множителя по переменным ξ и η в подынтегральной функции внутреннего интеграла. Это условие будет выполнено при $b = c$. Оставшийся внутренний интеграл представляет собой двумерную дельта-функцию

$$\delta \left(\frac{x}{f_1} - \frac{x_1}{b} + \frac{x_2}{c} + \frac{x'}{f_2}; \frac{y}{f_1} - \frac{y_1}{b} + \frac{y_2}{c} + \frac{y'}{f_2} \right).$$

Воспользовавшись фильтрующим свойством дельта-функции, окончательно получим

$$E_3'(x', y') = KE \left[f_1 \left(\frac{x_1}{b} - \frac{x_2}{c} \right) - \frac{f_1}{f_2} x'; f_1 \left(\frac{y_1}{b} - \frac{y_2}{c} \right) - \frac{f_1}{f_2} y' \right],$$

где через K обозначено произведение всех остальных множителей. Изображение располагается в задней фокальной плоскости второй линзы. Оно смещено в этой плоскости на

расстояние $f_1\left(\frac{x_1}{b} - \frac{x_2}{c}\right)$ по оси X' и на расстояние $f_1\left(\frac{y_1}{b} - \frac{y_2}{c}\right)$

по оси Y' . Изображение может быть перевернутым в плоскости изображения и увеличенным в зависимости от фокусных расстояний линз. Обычно опорная и восстанавливающая волны берутся плоскими: $b = c = \infty$. В этом случае

$$E_3'(x', y') = KE \left(-\frac{f_1}{f_2} x', -\frac{f_1}{f_2} y' \right).$$

Смещения изображения в плоскости XY' в случае плоских опорной и восстанавливающей волн не происходит.

13.6.2. Безлинзовая голография Фурье

Для осуществления преобразования Фурье, в частности, для получения голограммы Фурье наличие линз или других фокусирующих систем не является обязательным. Голограмму Фурье можно получить и без использования линз. С этой целью можно использовать обычную схему записи голограммы Френеля, в которой расстояние между источником опорной волны и голограммой равно расстоянию от предмета до голограммы ($b = a$). Такой метод называется безлинзовой голографией Фурье.

Заменив в опорной волне (13.17) b на a , для интенсивности светового поля в плоскости фотопластинки, соответствующей основному голографическому изображению, будем иметь

$$I_3(\xi, \eta) = AB \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \left\{ ik \left[\frac{(\xi - x)^2 + (\eta - y)^2}{2a} - \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2a} \right] \right\} dx dy$$

или в развернутом виде

$$I_3(\xi, \eta) = K_1 \exp\left(ik \frac{x_1 \xi + y_1 \eta}{a}\right) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \times \\ \times \exp\left(ik \frac{x^2 + y^2}{2a}\right) \exp\left[-ik \left(\frac{\xi}{a} x + \frac{\eta}{a} y\right)\right] dx dy,$$

где $K_1 = AB \exp\left(-ik \frac{x_1^2 + y_1^2}{2a}\right)$ – комплексная постоянная, A и B

определяются теми же выражениями, что и в п. 13.5.2. Если предмет достаточно мал (так что $k \frac{(x^2 + y^2)_{\max}}{2a} \ll \pi$), а

голограмма имеет размеры достаточно большие, чтобы обеспечить высокое разрешение, то квадратичный фазовый множитель в подынтегральном выражении можно положить равным единице. Тогда получим

$$I_3(\xi, \eta) = K_1 \exp\left[ik \left(\frac{\xi}{a} x_1 + \frac{\eta}{a} y_1\right)\right] \times \\ \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp\left[-ik \left(\frac{\xi}{a} x + \frac{\eta}{a} y\right)\right] dx dy. \quad (13.23)$$

Это распределение интенсивности будет зафиксировано на голограмме. Заметим, что действие опорной сферической волны в данном случае аналогично действию линзы, приводящему к компенсации искажающего квадратичного фазового множителя

$$\exp\left(ik \frac{\xi^2 + \eta^2}{2a}\right).$$

Соотношение (13.23) показывает, что распределение интенсивности на голограмме, соответствующее основному изображению, с точностью до несущественной постоянной и

линейного фазового множителя представляет собой преобразование Фурье комплексной амплитуды светового поля в плоскости предмета $E(x, y)$ на пространственных частотах $u = \xi / (\lambda a)$ и $v = \eta / (\lambda a)$. Если точечный опорный источник поместить в начале координат ($x_1 = y_1 = 0$), то линейный фазовый множитель перед интегралом исчезнет, и мы получим точное преобразование Фурье распределения амплитуды поля в плоскости предмета.

Изображение восстанавливается с помощью сферической волны (13.18). Тогда для основной дифрагировавшей волны получим

$$\begin{aligned}
 E_3'(x', y') = & F_1 \exp \left[-i \frac{k}{2} \left(\frac{x_1^2 + y_1^2}{a} - \frac{x_2^2 + y_2^2}{c} \right) \right] \times \\
 & \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E_1(x, y) \exp \left[i \frac{k}{2} \left(\frac{x^2 + y^2}{a} + \frac{x'^2 + y'^2}{d} \right) \right] \times \\
 & \times \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left[i \frac{k}{2} \left(\frac{1}{c} + \frac{1}{d} \right) (\xi^2 + \eta^2) \right] \right) \times \\
 & \times \exp \left\{ ik \left[\left(-\frac{x}{a} + \frac{x_1}{a} - \frac{x_2}{c} - \frac{x'}{d} \right) \xi + \right. \right. \\
 & \left. \left. + \left(-\frac{y}{a} + \frac{y_1}{a} - \frac{y_2}{c} - \frac{y'}{d} \right) \eta \right] \right\} d\xi d\eta \Big] dx dy,
 \end{aligned}$$

где F_1 – комплексная постоянная. Условием образования изображения является исчезновение квадратичного фазового множителя в подынтегральной функции внутреннего интеграла, что имеет место при $d_1 = -c < 0$. Следовательно, основное изображение является мнимым и возникает на расстоянии, равном расстоянию до восстанавливающего источника. Оставшийся

интеграл представляет собой двумерную дельта-функцию Дирака. Используя фильтрующее свойство δ -функции, для основного восстановленного изображения будем иметь

$$E_3'(x', y') = GE \left(x_1 - \frac{a}{c} x_2 - \frac{a}{d_1} x', y_1 - \frac{a}{c} y_2 - \frac{a}{d_1} y' \right),$$

где G – комплексная постоянная. Изображение смещено в плоскости изображения на величину $x_1 - \frac{a}{c} x_2$ по оси X' и на величину $y_1 - \frac{a}{c} y_2$ по оси Y' . Поскольку параметры a и d_1 имеют противоположные знаки, то изображение оказывается прямым.

Интенсивность, соответствующая сопряженному изображению

$$I_4(\xi, \eta) = AB \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E^*(x, y) \exp \left\{ -ik \left[\frac{(\xi - x)^2 + (\eta - y)^2}{2a} - \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} \right] \right\} dx dy,$$

при том же приближении, что и для основного изображения, приводится к виду

$$I_4(\xi, \eta) = K_2 \exp \left[-ik \left(\frac{\xi}{a} x_1 + \frac{\eta}{a} y_1 \right) \right] \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E^*(x, y) \exp \left[ik \left(\frac{\xi}{a} x + \frac{\eta}{a} y \right) \right] dx dy, \quad (13.24)$$

где $K_2 = AB \exp \left(ik \frac{x_1^2 + y_1^2}{2a} \right)$. С точностью до множителей,

стоящих перед интегралом, эта интенсивность представляет собой

преобразование Фурье комплексно сопряженной амплитуды поля в плоскости предмета. В случае, когда источник опорной волны помещен в начало координат, получим точное преобразование Фурье функции $E^*(x, y)$.

При восстановлении изображения сферической волной (13.18) получим для распределения поля в плоскости изображения

$$E_4'(x', y') = BE^* \left[x_1 + \frac{a}{c} x_2 + \frac{a}{d_2} x', y_1 + \frac{a}{c} y_2 + \frac{a}{d_2} y' \right].$$

При этом $d_2 = -c < 0$, т.е. сопряженное изображение так же, как и прямое, является мнимым и расположено на том же расстоянии от голограммы. Оно смещено вдоль координатных осей X' и Y' на расстояния $x_1 + \frac{a}{c} x_2$ и $y_1 + \frac{a}{c} y_2$ соответственно. Сравнивая смещения основного и сопряженного изображений, видим, что эти два восстановленных изображения, не перекрывая друг друга, будут располагаться симметрично относительно точки расположения опорного источника $S_1(x_1, y_1)$.

В случае, когда источник опорной волны расположен в начале координат, члены уравнения голограммы, соответствующие основному и сопряженному изображениям будут иметь вид соответственно

$$t_3(\xi, \eta) = \beta AB \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(x, y) \exp \left[-k \left(\frac{\xi}{a} x + \frac{\eta}{a} y \right) \right] dx dy$$

и

$$t_4(\xi, \eta) = \beta AB \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E^*(x, y) \exp \left[ik \left(\frac{\xi}{a} x + \frac{\eta}{a} y \right) \right] dx dy.$$

Они пропорциональны фурье-образам комплексной амплитуды $E(x, y)$, описывающей исходный предмет, и ее комплексно сопряженной функции $E^*(x, y)$. Для восстановления изображения

предметов $E(x, y)$ и $E^*(x, y)$ следует произвести фурье-преобразование функций $t_3(\xi, \eta)$ и $t_4(\xi, \eta)$. С этой целью голограмма освещается плоской волной, а за голограммой устанавливается собирающая линза. Получается схема с двойным преобразованием Фурье, описанная в п. 10.3.4. Оба изображения образуются в задней фокальной плоскости линзы.

Если источник опорной волны помещается в начале координат, то предмет следует переместить в точку (x_1, y_1) . Так как по свойству фурье-преобразований

$$F\{E(x - x_1; y - y_1)\} = F\{E(x, y)\} \exp[i2\pi(ux_1 + vy_1)],$$

то выражение (13.23) представляет собой преобразование Фурье на пространственных частотах $u = \xi/\lambda$ и $v = \eta/\lambda$ функции $E(x - x_1, y - y_1)$, описывающей распределение поля исходного предмета, смещенного в точку (x_1, y_1) , а выражение (13.24) – ее комплексно сопряженной функции $E^*(x + x_1, y + y_1)$, смещенной в точку $(-x_1, -y_1)$. Учитывая также, что второе преобразование Фурье возвращает исходную функцию, т.е. восстанавливает изображение, получим, что два восстановленных изображений также будут смещены и находиться соответственно в точках (x'_1, y'_1) и $(-x'_1, -y'_1)$, где штрихом обозначены координаты в плоскости изображения. Эти изображения будут расположены симметрично относительно оси Z .

13.7. Формирование и реконструкция голограмм изображений, создаваемых оптическими элементами

Рассмотрим теперь голографирование объектов с использованием линзы, которая создает изображение предмета в плоскости $X'Y'$. Использование линз представляет интерес, так как оно позволяет получить действительное голографическое изображение, расположенное перед голограммой. Поместим

фотопластинку справа от линзы на расстоянии a и осветим ее сферической опорной волной, испускаемой источником S_1 , находящимся на расстоянии b от плоскости фотопластинки. Поле, создаваемое опорной волной в плоскости фотопластинки, определится выражением (13.17). Объектом является плоский предмет S , расположенный на расстоянии a_1 от линзы. (Схема опыта такая же, как и на рис. 13.17, только теперь линза располагается на расстоянии a_1 от предмета и на расстоянии a от фотопластинки.) Комплексную амплитуду поля объектной волны, прошедшей через линзу, в плоскости фотопластинки можно определить, воспользовавшись выражением (11.8) и заменив в нем координаты в плоскости изображения X', Y' на координаты в плоскости фотопластинки ξ, η , координаты в плоскости линзы ξ, η на координаты x_L, y_L , расстояние b на a , где a – расстояние от линзы до фотопластинки, и положив $P(\xi, \eta)$ (в новых обозначениях $P(x_L, y_L)$, равной единице. Тогда получим

$$E_1(\xi, \eta) = -\frac{e^{ik(a+a_1)}}{\lambda^2 aa_1} \times \\ \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left\{ ik \left[\frac{(x_L - x)^2 + (y_L - y)^2}{2a_1} - \frac{x_L^2 + y_L^2}{2f} + \right. \right. \\ \left. \left. + \frac{(\xi - x_L)^2 + (\eta - y_L)^2}{2a} \right] \right\} dx_L dy_L dx dy, \quad (13.25)$$

где x, y – координаты в плоскости предмета, $E(x, y)$ – распределение комплексной амплитуды в плоскости непосредственно за плоскостью предмета, f – фокусное расстояние линзы. Первый и третий члены в показателе степени экспоненты в (13.25) обусловлены преобразованием Френеля при прохождении светом слоев свободного пространства между предметом и линзой и между линзой и фотопластинкой

соответственно, а второй – изменением фазы световой волны при прохождении линзы.

Опорная волна интерферирует с объектной волной, и возникшее в плоскости фотопластинки распределение интенсивности регистрируется фотопластинкой в виде голограммы. Амплитудная функция пропускания голограммы для основного изображения будет описываться выражением

$$\begin{aligned}
 t_3(\xi, \eta) &= \\
 &= C_1 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left\{ ik \left[\frac{(x_L - x)^2 + (y_L - y)^2}{2a_1} - \frac{x_L^2 + y_L^2}{2f} + \right. \right. \\
 &\quad \left. \left. + \frac{(\xi - x_L)^2 + (\eta - y_L)^2}{2a} - \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} \right] \right\} dx_L dy_L dx dy,
 \end{aligned}$$

где коэффициент C_1 включает в себя все амплитудные и фазовые множители, стоящие перед интегралом.

При освещении полученной голограммы сферической восстанавливающей волной от источника S_2 , расположенного на расстоянии c от голограммы (рис. 13.15), в плоскости $X'Y'$ возникнет дифрагировавшая волна с комплексной амплитудой

$$\begin{aligned}
 E'_3(x', y') &= C_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left\{ \left[\frac{(x_L - x)^2 + (y_L - y)^2}{2a_1} - \right. \right. \\
 &\quad - \frac{x_L^2 + y_L^2}{2f} + \frac{(\xi - x_L)^2 + (\eta - y_L)^2}{2a} - \\
 &\quad \left. \left. - \frac{(\xi - x_1)^2 + (\eta - y_1)^2}{2b} - \frac{(\xi - x_2)^2 + (\eta - y_2)^2}{2c} + \right. \right.
 \end{aligned}$$

$$+ \left. \frac{(x' - \xi)^2 + (y' - \eta)^2}{2d} \right\} dx_L dy_L d\xi d\eta dx dy ,$$

где последний член в показателе степени экспоненты обусловлен преобразованием Френеля при прохождении светом слоя свободного пространства между голограммой и плоскостью голографического изображения $X'Y'$; d – расстояние от плоскости голограммы до плоскости изображения $X'Y'$; C_2 – комплексная постоянная.

Представим выражение (13.25) в развернутом виде

$$\begin{aligned} E'_3(x', y') = & C_3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp\left(ik \frac{x^2 + y^2}{2a_1}\right) \times \\ & \times \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left\{-i \frac{k}{2} \left(-\frac{1}{a_1} + \frac{1}{f} - \frac{1}{a}\right) (x_L^2 + y_L^2) - \right. \right. \\ & \left. \left. - ik \left[\left(\frac{x}{a_1} + \frac{\xi}{a}\right) x_L + \left(\frac{y}{a_1} + \frac{\eta}{a}\right) y_L \right] \right\} dx_L dy_L \right) \times \\ & \times \exp\left[i \frac{k}{2} \left(\frac{1}{a} - \frac{1}{b} + \frac{1}{d}\right) (\xi^2 + \eta^2) \right] \exp\left\{ ik \left[\left(\frac{x_1}{b} - \frac{x_2}{c} - \frac{x'}{d}\right) \xi + \right. \right. \\ & \left. \left. + \left(\frac{y_1}{b} - \frac{y_2}{c} - \frac{y'}{d}\right) \eta \right] \right\} d\xi d\eta dx dy , \end{aligned}$$

(13.26)

где C_3 включает в себя все выражения перед интегралами, в том числе и фазовый множитель $\exp\left(ik \frac{x'^2 + y'^2}{2d}\right)$. Выполним интегрирование внутреннего интеграла. Используя формулу

$$\int_{-\infty}^{\infty} \exp(-\alpha x^2 \pm \beta x) dx = \sqrt{\frac{\pi}{\alpha}} \exp\left(\frac{\beta^2}{4\alpha}\right)$$

и положив сначала

$$\alpha = i \frac{k}{2} \left(-\frac{1}{a_1} + \frac{1}{f} - \frac{1}{a} \right), \quad \beta = ik \left(\frac{x}{a_1} + \frac{\xi}{a} \right),$$

а затем

$$\alpha = i \frac{k}{2} \left(-\frac{1}{a_1} + \frac{1}{f} - \frac{1}{a} \right), \quad \beta = ik \left(\frac{x}{a_1} + \frac{\xi}{a} \right),$$

придем к выражению

$$\exp \left[i \frac{k}{2} \left(\frac{(x^2 + y^2)/a_1^2}{1/a_2 - 1/a} + \frac{2(x\xi + y\eta)/a_1 a}{1/a_2 - 1/a} + \frac{(\xi^2 + \eta^2)/a^2}{1/a_2 - 1/a} \right) \right],$$

где a_2 – расстояние от линзы до плоскости $X'Y'$ изображения, даваемого линзой, и использована формула линзы $1/a_1 + 1/a_2 = 1/f$, описывающая построение изображения предмета линзой. С учетом этого выражение (13.26) принимает вид

$$E'_3(x', y') =$$

$$= C_4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[i \frac{k}{2} \left(\frac{1}{a_1^2(1/a_2 - 1/a)} + \frac{1}{a_1} \right) (x^2 + y^2) \right] \times$$

$$\times \exp \left[i \frac{k}{2} \left(\frac{1}{a^2(1/a_2 - 1/a)} + \frac{1}{a} - \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right) (\xi^2 + \eta^2) \right] \times$$

$$\times \exp \left\{ ik \left[\left(\frac{x}{aa_1(1/a_2 - 1/a)} + \frac{x_1}{b} - \frac{x_2}{c} - \frac{x'}{d} \right) \xi + \right. \right.$$

$$+ \left(\frac{y}{a_1 a (1/a_2 - 1/a)} + \frac{y_1}{b} - \frac{y_2}{c} - \frac{y'}{d} \right) \eta \left. \right\} d\xi d\eta dx dy.$$

Амплитуда дифрагировавшей волны будет соответствовать амплитуде в изображении в том случае, если исчезнет квадратичный фазовый множитель по переменным ξ и η , т.е. если

$$\frac{1}{a^2(1/a_2 - 1/a)} + \frac{1}{a} - \frac{1}{b} + \frac{1}{c} + \frac{1}{d_1} = 0.$$

Из этого соотношения получаем формулу для определения расстояния d_1 от основного голографического изображения до голограммы:

$$\frac{1}{d_1} = \frac{1}{a_2 - a} + \frac{1}{b} - \frac{1}{c}, \quad (13.27)$$

где $a_2 - a$ – расстояние от изображения, образованного линзой, т.е. от объекта голографирования, до голограммы.

Оставшийся интеграл можно записать в виде

$$\begin{aligned} E'_3(x', y') = & C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \exp \left[i \frac{k}{2} \left(\frac{1}{a_1^2(1/a_2 - 1/a)} + \frac{1}{a_1} \right) (x^2 + y^2) \right] \times \\ & \times \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left\{ ik \left[\left(\frac{x}{a_1 a (1/a_2 - 1/a)} + \frac{x_1}{b} - \frac{x_2}{c} - \frac{x'}{d_1} \right) \xi + \right. \right. \right. \\ & \left. \left. \left. + \left(\frac{y}{a_1 a (1/a_2 - 1/a)} + \frac{y_1}{b} - \frac{y_2}{c} - \frac{y'}{d_1} \right) \eta \right] \right\} d\xi d\eta \right) dx dy. \end{aligned}$$

Внутренний интеграл представляет собой двумерную дельта-функцию вида

$$\delta \left(-\frac{x}{aa_1(1/a_2 - 1/a)} - \frac{x_1}{b} + \frac{x_2}{c} + \frac{x'}{d}, \right. \\ \left. -\frac{y}{a_1a(1/a_2 - 1/a)} - \frac{y_1}{b} + \frac{y_2}{c} + \frac{y'}{d_1} \right).$$

Произведя интегрирование с дельта-функцией, окончательно получим

$$E'_3(x', y') = CE \left[a_1a \left(\frac{1}{a} - \frac{1}{a_2} \right) \left(\frac{x_1}{b} - \frac{x_2}{c} \right) - \frac{a_1a}{d_1} \left(\frac{1}{a} - \frac{1}{a_2} \right) x' - \right. \\ \left. - a_1a \left(\frac{1}{a} - \frac{1}{a_2} \right) \left(\frac{y_1}{b} - \frac{y_2}{c} \right) - \frac{a_1a}{d_1} \left(\frac{1}{a} - \frac{1}{a_2} \right) y' \right].$$

Полученное выражение показывает тождественность голографического изображения и предмета. Изображение

сдвинуто в плоскости $X'Y'$ на расстояния $a_1a \left(\frac{1}{a} - \frac{1}{a_2} \right) \left(\frac{x_1}{b} - \frac{x_2}{c} \right)$

и $a_1a \left(\frac{1}{a} - \frac{1}{a_2} \right) \left(\frac{y_1}{b} - \frac{y_2}{c} \right)$. Увеличение изображения определяется

произведением увеличений голограммы $(a_2 - a) / d_1$ и линзы (a_1 / a_2) .

Если при записи голограммы фотопластинку поместить непосредственно в плоскость изображения $X'Y'$, то можно получить голограмму, называемую голограммой сфокусированного изображения. В этом случае $a_2 = a$ и, как следует из соотношения (13.27), $d_1 = 0$, независимо от значений b и c . Голографическое изображение располагается в плоскости изображения, даваемого линзой при записи голограммы, при любых расстояниях опорного и реконструирующего источников

от голограммы. Общее увеличение изображения в этом случае будет определяться только увеличением линзы.

13.8. Применения голографии

Различные применения голографии весьма разнообразны и интересны. Их можно разделить на два направления: изобразительная голография и применение голографии в науке и технике. Здесь мы рассмотрим лишь голографические методы обработки пространственной информации, представляющие интерес для фурье-оптики.

1. Оптическая фильтрация. Рассмотрим сначала применение голографии для пространственной оптической фильтрации. Оптическая фильтрация без использования голографических методов была рассмотрена нами в п. 12.8.1. Основную трудность здесь представляет создание частотных фильтров (называемых также частотными масками) по заданному импульсному отклику $h(x, y)$, фурье-образ которого $H(u, v)$ как раз и представляет собой распределение коэффициента пропускания $K(u, v)$ этого фильтра. Голография позволяет преобразовывать как амплитуду, так и фазу светового поля в фурье-плоскости сигнала с помощью транспаранта с вещественной функцией пропускания (такой транспарант обычно изменяет только амплитуду прошедшей сквозь него волны). Поэтому голографическая пространственная фильтрация осуществляется с использованием амплитудных фильтров, изготавливаемых методами голографии. Этот метод был предложен Вандер-Люгтом.

Фильтр Вандер-Люгта изготавливается с помощью оптической системы, схема которой показана на рис. 13.19. Плоская волна от лазера частично проходит через транспарант T с функцией пропускания $t(\xi, \eta)$, соответствующей требуемой импульсной функции $h(\xi, \eta)$, где ξ и η – координаты в плоскости транспаранта, и частично попадает на преломляющую призму (лазер и преломляющая призма на рисунке не показаны). Линза

L осуществляет фурье-преобразование импульсной функции. На фото-

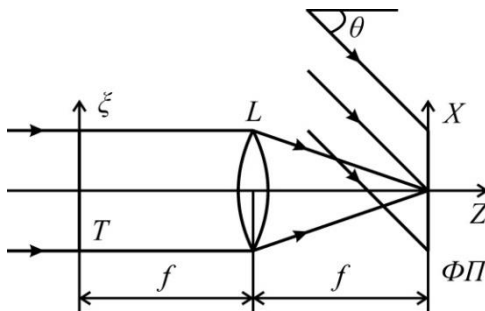


Рис. 13.19

пластинке $\Phi\Pi$, помещенной в заднюю фокальную плоскость линзы (эту плоскость будем считать плоскостью $z = 0$), фиксируется интенсивность суперпозиции двух полей: поля опорной волны, прошедшей через призму и

падающей на фотопластинку под углом $-\theta$ к нормали к ней (к

оси Z), и поля, представляющего собой фурье-образ $H\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)$

(с множителем $1 / (\lambda f)$) прошедшего через транспарант света, где x и y – координаты в плоскости фотопластинки.

Поле опорной волны в плоскости фотопластинки можно записать в виде

$$E_{\text{он}}(x) = E_1 \exp(-ikx \sin \theta) = E_1 \exp\left(-i2\pi \frac{\sin \theta}{\lambda} x\right),$$

где $\sin \theta / \lambda = u$ – пространственная частота вдоль оси X . Комплексный фурье-образ прошедшего через транспарант светового поля также можно представить в виде

$$H\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right) = |H| \exp\left[-i\varphi\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)\right].$$

Тогда суммарная интенсивность света в плоскости фотопластинки будет

$$I(x, y) = |E_1|^2 + \frac{1}{(\lambda f)^2} |H|^2 + 2 \frac{1}{\lambda f} |E_1| |H| \times \\ \times \cos \left[2\pi \frac{\sin \theta}{\lambda} x - \varphi \left(\frac{x}{\lambda f}, \frac{y}{\lambda f} \right) \right].$$

Это соотношение показывает, что можно получить транспарант с записью комплексной функции $H(u, v)$ на обычную фотопластинку, регистрирующую только интенсивность света. Амплитудная и фазовая информация о комплексной функции записана в виде амплитудной и фазовой модуляций несущей пространственной частоты $u = \sin \theta / \lambda$. После обработки фотопластинки коэффициент пропускания транспаранта (фильтра) будет пропорционален интенсивности

$$I(x, y) = |E_1|^2 + \frac{1}{(\lambda f)^2} |H|^2 + \frac{1}{\lambda f} E_1 H \left(\frac{x}{\lambda f}, \frac{y}{\lambda f} \right) \times \\ \times \exp \left(i 2\pi \frac{\sin \theta}{\lambda} x \right) + \frac{1}{\lambda f} E_1 H^* \left(\frac{x}{\lambda f}, \frac{y}{\lambda f} \right) \exp \left(-i 2\pi \frac{\sin \theta}{\lambda} x \right).$$

Следовательно,

$$t(x, y) = t_0 + \frac{1}{\lambda f} \beta E_1 H \left(\frac{x}{\lambda f}, \frac{y}{\lambda f} \right) \exp \left(i 2\pi \frac{\sin \theta}{\lambda} x \right) + \\ + \frac{1}{\lambda f} \beta E_1 H^* \left(\frac{x}{\lambda f}, \frac{y}{\lambda f} \right) \exp \left(-i 2\pi \frac{\sin \theta}{\lambda} x \right) \quad (13.28)$$

в предположении, что $I_0 = |E_1|^2 + \frac{1}{(\lambda f)^2} |H|^2$. Второй член выражения (13.28) с точностью до экспоненциального множителя

и постоянных коэффициентов представляет собой интересующую нас функцию $H(u, v)$.

Фильтр с такой амплитудной функцией пропускания используется далее в обычной оптической схеме для осуществления пространственной фильтрации (рис. 10.19). Фильтруемый сигнал $E(x, y)$ помещают в переднюю фокальную плоскость линзы L_1 и освещают плоской монохроматической волной. В задней фокальной плоскости линзы L_1 формируется

распределение поля $\frac{1}{\lambda f} E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right)$. В этой же плоскости

устанавливается и полученный по методу Вандер-Люгта голографический фильтр. Фурье-образ $E(u, v)$ фильтруемого сигнала $E(x, y)$ используется в качестве восстанавливающей волны. Тогда непосредственно за голограммой комплексная амплитуда поля будет иметь вид

$$\begin{aligned}
 E(x_1, y_1) = & \frac{1}{\lambda f} t_0 E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right) + \\
 & + \frac{1}{(\lambda f)^2} E_1 \beta E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right) H\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right) \exp\left(i2\pi \frac{\sin\theta}{\lambda} x_1\right) + \\
 & + \frac{1}{(\lambda f)^2} \beta E_1 E\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right) H^*\left(\frac{x_1}{\lambda f}, \frac{y_1}{\lambda f}\right) \exp\left(-i2\pi \frac{\sin\theta}{\lambda} x_1\right).
 \end{aligned}$$

Линза L_2 , расположенная в системе фильтрации за фильтром-голограммой, осуществляет преобразование Фурье поля $E(x_1, y_1)$ с множителем $1/(\lambda f)$. Считая, что система координат в задней фокальной плоскости линзы L_2 инверсно преобразована (координатные оси повернуты на угол π), вместо $F\{F\{E(x_1, y_1)\}\} = E'(-x_2, -y_2)$ будем писать $F\{F\{E(x_1, y_1)\}\} = E'(x'_2, y'_2)$. Кроме того, учитывая, что

$$\begin{aligned}
F\left\{\exp\left(i2\pi\frac{\sin\theta}{\lambda}x_2'\right)\right\} &= \\
&= \frac{1}{\lambda f} \int_{-\infty}^{\infty} \exp\left(i2\pi\frac{\sin\theta}{\lambda}\xi\right) \exp\left(-i2\pi\frac{\xi}{\lambda f}x_2'\right) d\xi = \\
&= \frac{1}{\lambda f} \int_{-\infty}^{\infty} \exp\left[-i2\pi\left(\frac{x_2'}{\lambda f} - \frac{\sin\theta}{\lambda}\right)\xi\right] d\xi = \\
&= \frac{1}{\lambda f} \delta\left(\frac{x_2'}{\lambda f} - \frac{\sin\theta}{\lambda}\right) = \frac{1}{\lambda f} \delta\left(\frac{x_2' - f\sin\theta}{\lambda f}\right) = \\
&= \delta\left(x_2' - f\sin\theta\right),
\end{aligned}$$

где использовано свойство δ -функции $\delta(\alpha x) = \delta(x)/|\alpha|$.
Аналогично получаем

$$F\left\{\exp\left(-i2\pi\frac{\sin\theta}{\lambda}x_2'\right)\right\} = \delta(x_2' + f\sin\theta).$$

Тогда, применяя свойство линейности преобразования Фурье и вторую теорему свертки (теорему Парсеваля), будем иметь

$$\begin{aligned}
F\{E(x_1, y_1)\} &= t_0 E'(x_2', y_2') + \\
&+ \beta \frac{E_1}{\lambda f} [E'(x_2', y_2') * H(x_2', y_2') * \delta(x_2' - f\sin\theta)] + \\
&+ \beta \frac{E_1}{\lambda f} [E(x_2', y_2') * H^*(x_2', y_2') * \delta(x_2' + f\sin\theta)]. \quad (13.29)
\end{aligned}$$

Первый член полученного выражения соответствует ослабленному по интенсивности изображению объекта,

расположенному около начала координат плоскости $X_2'Y_2'$. Второй и третий члены описывают процесс реконструкции плоской волны, использовавшейся в качестве опорной при получении фильтра. Эти два члена, соответствующие восстановленным голограммой волнам, содержат информацию о степени сходства функций $h(x, y)$ и $E(x, y)$, так как только в случае тождества обеих функций происходит восстановление плоской опорной волны, которая дает изображение точки в плоскости изображения.

Используя фильтрующее свойство дельта-функции, это выражение можно записать в интегральном виде. При этом интеграл, соответствующий второму члену

$$\beta \frac{E_1}{\lambda f} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(\xi, \eta) h(x_2' - \xi + f \sin \theta, y_2' - \eta) d\xi d\eta,$$

(13.30)

представляет собой свертку функции $E(x, y)$ с функцией $h(x, y)$, смещенной по оси X в плоскости $X_2'Y_2'$ на величину $-f \sin \theta$. Третьему члену соответствует интеграл

$$\beta \frac{E_1}{\lambda f} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} E(\xi, \eta) h^*(x_2' - \xi - f \sin \theta, y_2' - \eta) d\xi d\eta, \quad (13.31)$$

представляющий собой функцию корреляции функции $E(x, y)$ и смещенной по оси X на величину $f \sin \theta$ функции $h(x, y)$. Интеграл (13.31) отличается от свертки тем, что в него входит функция, комплексно сопряженная импульсной функции $h(x, y)$. Изображения, соответствующие корреляционному члену и члену, описываемому сверткой, расположены симметрично по обе стороны от начала координат (где расположено изображение объекта) вблизи точек $x_2' = \pm f \sin \theta$. Выражения (13.30) и

(13.31) свидетельствуют о том, что полученный фильтр является фурье-образом требуемой импульсной функции.

Фильтр Вандер-Люгта существенно повышает возможности использования когерентных оптических систем для практических целей. Получение спектра импульсного отклика методом Вандер-Люгта позволяет отказаться от использования при фильтрации фазосдвигающих элементов, так как фазовая информация кодируется в фильтре в виде модуляции несущей пространственной частоты, т.е. голографическим способом.

2. Распознавание образов. Пространственная оптическая фильтрация позволяет решить проблему распознавания образов, являющуюся одной из важнейших практических проблем теории информации. Здесь в основном рассматриваются две задачи: идентификация сигнала в серии последовательно поступающих сигналов и идентификация сигнала среди других сигналов (шума), одновременно поступающих на вход оптической фильтрующей системы. В обеих задачах основной проблемой является выбор пространственного фильтра. Рассмотренная выше когерентная оптическая система позволяет осуществить распознавание, если изготовить фильтр с комплексным коэффициентом пропускания, представляющим собой функцию, комплексно сопряженную фурье-образу распознаваемого сигнала $E(x, y)$: $H(u, v) = E^*(u, v)$. Такой фильтр называют согласованным с сигналом $E(x, y)$. В этом случае при сходстве поступающего сигнала с распознаваемым распределение амплитуды поля, прошедшего через фильтр, будет представлять собой действительную величину

$$E(u, v)H(u, v) = E(u, v)E^*(u, v) = |E(u, v)|^2.$$

Это означает, что фильтр компенсирует искривление волнового фронта сигнала в фурье-плоскости и, следовательно, на вторую линзу будет падать плоская волна, которая сфокусируется линзой в яркую точку в ее задней фокальной плоскости. Это следует и из выражений (13.29) и (13.30): при фильтре, согласованном с распознаваемым сигналом, автокорреляционной функции и

автосвертке соответствуют локализованные максимумы интенсивности наибольшей яркости

Для решения первой задачи распознавания образов, т.е. для идентификации сигналов, поступающих последовательно, можно применять оптические схемы, представленные на рис. 12.19 и 12.20, где во входной плоскости последовательно меняется изображение сигнала и каждый сигнал последовательно сравнивается с набором фильтров. При максимальном сходстве предъявляемого сигнала с распознаваемым в плоскости изображения (являющейся в «схеме 4f» задней фокальной плоскостью второй линзы) получается максимальный отклик.

Вторая задача, т.е. идентификация сигнала на фоне шума, является более сложной, поскольку в выходной плоскости, кроме автокорреляции, возникает корреляция с шумовым сигналом, которая в некоторых случаях может превышать уровень сигнала автокорреляции. Если на вход системы поступают распознаваемый сигнал $E(x, y)$ и аддитивный шум $S(x, y)$, то на входе системы будет действовать сигнал

$$E_1(x, y) = E(x, y) + S(x, y).$$

Спектр этого сигнала будет

$$E_1(u, v) = E(u, v) + S(u, v).$$

После осуществления фильтрации спектр будет

$$E_1'(u, v) = E(u, v)H(u, v) + S(u, v)H(u, v).$$

Отклик на выходе системы определится суммой сверток

$$E_1'(x, y) = E(x, y) * h(x, y) + S(x, y) * h(x, y).$$

В этом случае в плоскости изображения не весь свет будет собран в точку. Наличие распознаваемого сигнала в этом случае может быть обнаружено по относительным измерениям интенсивности в заднем фокусе второй линзы. Чтобы выделить идентифицируемый сигнал, нужно достаточно большое отношение сигнала к шуму, т.е. величина $|E_{1s}'|^2_m / |E_{1v}'|^2_m$, где

$(E_{1s}')_m$ и $(E_{1B}')_m$ – максимумы откликов сигнала и шума. Оно должно быть выше порога обнаружения измерительной системы.

Примером нахождения сигнала в шуме может служить задача обнаружения определенного слова или буквы на странице с текстом.

3. Восстановление искаженного изображения. Кроме способа восстановления искаженного изображения, описанного в п. 12.8.2, при котором голографический метод позволяет реализовать соответствующий фильтр, существует другой метод, основанный на свойствах голографической записи и восстановления волнового поля. В этом методе подлежащий восстановлению искаженный пространственный сигнал помещают в переднюю фокальную плоскость линзы (см. схему на рис. 12.19). В задней фокальной плоскости линзы регистрируется фурье-голограмма этого сигнала. При этом в качестве опорной волны при записи этой голограммы используют волну, являющуюся фурье-образом импульсного отклика процесса искажения изображения. При практическом решении этой задачи целесообразно использовать схему получения безлинзовой голограммы Фурье.

Интерференционное поле в плоскости регистрации голограммы будет описываться выражением

$$|E'(u, v) + H(u, v)|^2 = |E'(u, v)|^2 + |H(u, v)|^2 + E'^*(u, v)H(u, v) + E'(u, v)H^*(u, v).$$

Интерес для нас представляют два последних члена этого выражения, сумму которых используя соотношение

$$E'(u, v) = E(u, v)H(u, v),$$

можно переписать в виде

$$E'^*(u, v)H^*(u, v)H(u, v) + E(u, v)H(u, v)H^*(u, v) = E'^*(u, v)|H(u, v)|^2 + E(u, v)|H(u, v)|^2. \quad (13.32)$$

Полученное выражение несет информацию об объекте. Квадрат модуля фурье-образа импульсного оклика $|H(u, v)|^2 = H(u, v)H^*(u, v)$ не содержит фазового множителя, поэтому выражение (13.32) представляет собой сумму спектра первоначального неискаженного изображения и его комплексно сопряженной величины. При восстановлении голограммы с помощью сферической волны с центром в точке (x_0, y_0) в фурье-плоскости линзы образуются основное и сопряженное изображения:

$$\begin{aligned} & F\{E_B |H(u, v)|^2 E^*(u, v) + E_B |H(u, v)|^2 E(u, v)\} \sim \\ & \sim E(x, y)^* \delta(x - x_0, y - y_0) + E(x, y) \delta(x - x_0, y - y_0). \end{aligned}$$

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

37. Аблеков В.К., Зубков П.И., Фролов А.В. Оптическая и оптоэлектронная обработка информации. –М.: Машиностроение, 1976.
38. Аблеков В.К., Колядин С.А., Фролов А.В. Высокора разрешающие оптические системы. –М.: Машиностроение, 1985.
39. Айзинов М.М. Радиотехнические цепи и сигналы. –М.: Транспорт, 1966.
40. Ахиезер А.И., Ахиезер И.А. Электромагнетизм и электромагнитные волны. –М.: Высшая школа, 1985.
41. Баскаков С.И. Радиотехнические цепи и сигналы. –М.: Высшая школа, 1988.

42. Борн М., Вольф Э. Основы оптики. –М.: Мир, 1973.
43. Бутиков Е.И. Оптика. СПб, 2003.
44. Волосов Д.С. Фотографическая оптика. –М.: Искусство, 1978.
45. Гершензон Е.М., Малов Н.Н., Мансуров А.Н. Курс общей физики. Оптика и атомная физика. – М.: Просвещение, 1992.
46. Гоноровский И.С., Демин М.П. Радиотехнические цепи и сигналы. –М.: Радио и связь, 1994.
47. Гребенников О.Ф. Основы записи и воспроизведения изображения. –М.: Искусство, 1982.
48. Гребенников О.Ф., Тихомирова Г.В. Основы записи и воспроизведения информации. –СПб.: СПбГУКиТ, 2005.
49. Дитчберн Р. Физическая оптика. –М.: Наука, 1965.
50. Дьяков В.А., Тарасов Л.В., Щукин И.В. Введение в радиооптику. –М.: МИЭМ, 1975.
51. Гудмен Дж. Введение в фурье-оптику. –М.: Мир, 1970.
52. Гудмен Дж. Статистическая оптика. –М.: Мир, 1988.
53. Зверев В.А. Радиооптика. Преобразование сигналов в радиотехнике и оптике. –М.: Сов. Радио, 1975.
54. Козане А., Флере Ж. и др. Оптика и связь. –М.: Мир, 1984.
55. Ключников А.С. Радиооптика и голография. –Минск: Университетское, 1989.
56. Кольер Р., Бернхарт К., Лиин Л. Оптическая голография. М.: мир, 1973.
57. Кондратенков Г.С. Обработка информации когерентными оптическими системами. –М.: Радио, 1972.
58. Литвиненко О.Н. Основы радиооптики. –Киев: Техника, 1974.
59. Ловенталь С., Бельво И. Пространственная фильтрация и голография. –М.: Энергия, 1970.
60. Макс Ж. Методы и техника обработки сигналов. –М.: Мир, 1983.
61. Маришаль А., Франсон М. Структура оптического изображения. –М.: Мир, 1964.
62. Микаэлян А.Л. Оптические методы а информатике. –М.: Наука, 1990.
63. Миллер М. Голография. –Л.: Машиностроение, 1979.

64. Москалев В.А., Нагибина Н.М., Полушкина Н.А., Рудин В.Л. Прикладная физическая оптика. –СПб.: Политехника, 1995.
65. Оптическая голография. Т. 1. –М.: Мир, 1982.
66. Папулис А. Теория систем и преобразований в оптике. –М.: Мир, 1971.
67. Петров М.П. Оптическая обработка информации и голография. –Л.: ЛФТИ, 1981.
68. Применение методов фурье-оптики/ Под ред. Г. Старка. –М.: Радио и связь, 1988.
69. Сайкин А.С., Морозова Н.Н. Фурье-анализ в оптике. – Чебоксары: ЧГУ, 1997.
70. Сороко Л.М. Основы голографии и когерентной оптики. –М.: Наука, 1971.
71. Струок Дж. Введение в когерентную оптику и голографию. – М.: Мир, 1967.
72. Стюард И.Г. Введение в фурье-оптику. –М.: Мир, 1985.
73. Тихомирова Г.В. Физические основы получения зрительной информации. –СПб.: СПбГУКиТ, 2005.
74. Троицкий И.Н. Когерентная оптика и голография. –М.: МФТИ, 1982.
75. Франсон М. Голография. –М.: Мир, 1972.
76. Фризер Х. Фотографическая регистрация информации. –М.: Мир, 1978.
77. Черемисин А.И. Формирование оптического изображения. Учебное пособие. –СПб.: СПбГУКиТ, 1997.
78. Шульман М.Я. Измерение передаточных функций оптических систем. –Л.: Машиностроение, 1980.
79. Юу Ф.Т.С. Введение в теорию дифракции, обработку информации и голографию. –М.: Сов. Радио, 1979.
80. Якушенков Ю.Г. Теория и расчет оптико-электронных приборов. –М.: Машиностроение, 1989.