

В. А. Зверев

# ФИЗИЧЕСКИЕ ОСНОВЫ ФОРМИРОВАНИЯ ИЗОБРАЖЕНИЙ ВОЛНОВЫМИ ПОЛЯМИ



*В 1975 году автор опубликовал монографию "Радиооптика. Преобразование сигналов в радио и оптике", на ее основе более 20 лет читал курс лекций студентам радиофизического факультета, постоянно совершенствуя и перерабатывая материал. Большой педагогический опыт позволил отобрать самое существенное и употребительное, и родилась новая книга.*



Нижний  
Новгород  
1998

УДК 621.3/. 535.2  
ББК 22.34

**Зверев В. А.**

З 43

**Физические основы формирования изображений  
волновыми полями. – Нижний Новгород: ИПФ РАН,  
1998. – 252 с.**

ISBN 5-201-09309-4

С единой точки зрения – преобразования сигнала цепочкой фильтров – рассматривается дифракция волн как в свободном пространстве, так и при преобразовании волновых полей устройствами, формирующими изображения (антенны, оптические объективы и корреляторы) для когерентного и некогерентного освещения. Изложение физических основ голографии, когерентной оптики, апертурного синтеза иллюстрировано конкретными примерами, полученными математическим моделированием соотношений, выведенных в книге, что может служить практическим руководством по математическому моделированию задач формирования изображений волновыми полями.

Решением методической комиссии радиофизического факультета ННГУ книга рекомендована в качестве учебного пособия для студентов и аспирантов соответствующих специальностей.

**Рецензенты**

**доктор физико-математических наук**

**Н. С. Степанов,**

**кандидат физико-математических наук**

**Л. С. Долин**

*Монография подготовлена в рамках федеральной целевой программы  
"Интеграция" – Учебно-научный центр "Фундаментальная радиофизика"*

ISBN 5-201-09309-4

© Институт прикладной  
физики РАН, 1998 г.

## **Введение**

Во многих областях физики и техники существует необходимость получать изображения с помощью волновых полей. При решении задач электроники, ряда медицинских и технических применений требуется формировать волновое поле с заданными пространственными характеристиками, например, это будет поле, повторяющее контуры элементов микросхемы или медицинского скальпеля и т. д.

Существует и более обширный класс задач — задачи видения, в которых требуется по принятому излучению определить, где находятся его источники или какова форма и некоторые физические характеристики предмета, от которого отражаются волны или на котором волны дифрагируют. Класс задач такого рода очень широк: геология — взгляд сквозь Землю, радиофизика — определение параметров ионосферы Земли и планет, радиотехника — радиоизображения высокого качества, акустика — определение направления на источник звука, дефектоскопия и т. д. Трудно перечислить все, так как задачи такого рода множатся в связи с освоением новых диапазонов излучений и расширяющимися возможностями математической обработки принимаемых сигналов.

Если рассматривать проблему применительно к оптике, то следует обозначить еще один класс задач, которым уделяется большое внимание последние 10—15 лет, — это оптические методы обработки информации, которые стали бурно развиваться в связи с успехами когерентной оптики и голографии. Интерес к этому направлению вызван тем, что вычислительные машины, основанные на оптических принципах формирования изображений, способны обрабатывать и быстро выдавать гигантские объемы информации. Такие оптические машины существуют, и в ряде случаев (например, при обработке радиоизображений поверхности Земли) они существенно превосходят по производительности и другим параметрам возможности современных ЭВМ. Однако пока они

могут решать лишь узкий класс задач. Работа по совершенствованию этой техники ведется, и будущие результаты способны еще раз потрясти научный и технический мир. Еще раз потому, что такими потрясениями были два крупных научных открытия — это появление радиолокаторов бокового обзора с синтезируемой апертурой и голография.

Прежде чем перейти к систематическому изложению материала, хочу подчеркнуть некоторые его особенности.

Научное направление и построенная на нем учебная дисциплина имеют свой предмет исследования или изучения, который, как правило, отражен в названии. Например, электродинамика или физика плазмы, электроника, радиотехника и т. д. Однако, наряду с научными направлениями, имеющими свой физический предмет исследования, возникли интегральные научные направления, которые включают в себя части целого ряда сформировавшихся самостоятельных научных направлений. Примером такой интегральной науки является теория колебаний и волн.

Предлагаемая книга является типичным примером рассмотрения интегрального научного направления, построенного на основе оптики, акустики и радио. Интегрированию разных направлений науки в огромной мере способствовал проф. Г. С. Горелик, памяти которого посвящена эта книга. Это научное направление объединяет части, относящиеся к разным разделам физики, на основе понятия "изображение". Остановимся на этом понятии подробнее.

Под термином "изображение" обычно понимают вид или образ какого-то объекта наблюдения. Здесь мы подходим к изображению как к понятию, которое связано с образом конкретного объекта, но оторвано от него. Под изображением будем понимать распределение волнового поля, которое возникает в некоторой плоскости в результате дифракции сферической волны на конкретном объекте. При таком определении изображение одного и того же объекта оказывается различным в зависимости от параметров волны, освещающей объект, и расстояния от объекта, на котором наблюдается волновое поле. Такое определение понятия "изображение" позволяет с единых позиций рассмотреть когерентное и некогерентное освещение объектов, включая и самосветящиеся. Это удастся сделать путем корректировки понятия "изображение" для случая некогерентного освещения — достаточно понятие "изображение" связать не с самим волновым полем, а с его интенсивностью. В третьей части книги понятие "изображение" связывается с другими параметрами волнового поля.

Рассмотрение разнообразнейших ситуаций с единых позиций имеет две стороны. С одной, это упрощает рассмотрение, позволяя



вывести все соотношения, относящиеся к формированию изображений в разнообразнейших условиях, на основе одной простенькой формулы (1.1), помещенной в начале книги. С другой стороны, это же его и усложняет, связывая все вопросы в единый клубок, грозный запутаться и превратиться не в стройную систему взглядов, а в некоторый запутанный клубок, напоминающий "бороду", образующуюся при неудачном забрасывании спиннинга.

Нашей задачей будет найти ту ниточку, потянув за которую, можно вытянуть все взаимосвязанное в одну стройную логическую цепочку. Изначальная сложность взаимосвязей даже полезна, она приводит к парадоксам. Как говорил академик Л. И. Мандельштам о значении парадоксов в развитии науки и в обучении, есть две степени понимания: первая степень понимания, при которой возникает парадокс, и вторая степень понимания, когда все становится ясным во всех взаимосвязях и парадоксов уже не возникает. Когда достигнута вторая степень понимания, можно тянуть любую ниточку запутанного клубка и он при этом распутывается, а не затягивается. Без привлечения парадоксов нельзя достичь второй степени понимания, а именно к ней и нужно стремиться.

Несколько слов об истории написания настоящей книги.

В 1975 году в издательстве "Советское радио" автор опубликовал монографию под названием "Радиооптика. Преобразование сигналов в радио и оптике" [1] (а еще ранее совместно с Е. Ф. Орловым в том же издательстве была издана монография [2]). На основе [1] более двадцати лет читал курс лекций студентам радиофизического факультета, естественно, постоянно совершенствуя и перерабатывая материал, из которого большой педагогический опыт позволил отобрать самое существенное и употребительное.

Все эти годы книга [1] используется студентами радиофизического факультета Нижегородского университета, специализирующимися по направлению "Радиооптика и голография". (Позднее подобная специализация открыта в ряде других вузов, был создан Головной совет Минвуза по радиооптике). При этом, естественно, базовый лекционный курс автора, основанный на книге [1], сопровождается другими спецкурсами. В них рассматриваются конкретные задачи, схемы, современная элементная база, предельные параметры и достигнутые результаты\*.

\* См. также: Экспериментальная радиооптика / Под ред. В. А. Зверева, Н. С. Степанова. М.: Наука, 1977; Гельфер Э. И., Степанов Н. С. Методы прикладной радиооптики: Учебное пособие / ГГУ. Горький, 1986; Гельфер Э. И., Менсов С. Н. Современные оптоэлектронные устройства: Учебное пособие / ГГУ. Горький, 1987; Гельфер Э. И., Степанов Н. С. Шумы в радиооптических устройствах: Учебное пособие / ГГУ. Горький, 1988.

Насущная потребность в обновленном материале подвигла на написание новой книги. В отличие от других отечественных и зарубежных изданий, посвященных теории образования изображений волновыми полями, в настоящей книге внимание уделено не только оптическим изображениям, но и изображениям, получаемым с помощью антенн или в ходе апертурного синтеза посредством специальной математической обработки входных распределений полей.

В книге не затрагиваются технические или технологические вопросы, не уделяется никакого внимания теории сигналов и помех, в ней рассматриваются исключительно те вопросы, которые касаются физики формирования изображения.

Сейчас стало актуальным в ряде задач определять изображения, создаваемые волновыми полями, численными методами [3—5]. Здесь как нельзя более подходящей оказалась та математическая форма представления изображений, создаваемых волновыми полями, которая развита именно в "Радиооптике". Подобное представление поля встречается в книгах [6—7], но не развито столь подробно, чтобы его можно было непосредственно использовать в численных расчетах.

В настоящей монографии подведена теоретическая основа методики численных расчетов и даны примеры таких расчетов с приведением не только формул, но и математических программ. При этом использован накопленный автором опыт численных расчетов, с помощью которых выполнен ряд работ [3—5, 8—10].

## **РЕГУЛЯРНЫЕ ПОЛЯ**

### **1. ПОЛЯ В СВОБОДНОМ ПРОСТРАНСТВЕ**

Регулярным мы называем такое поле, которое можно точно предсказать, имея необходимый объем данных об этом поле, а значит, регулярное поле может быть представлено в виде математического выражения, определяющего колебания в различных точках пространства.

Здесь будем рассматривать не регулярное поле в общем случае, а только частный вид такого поля, где колебания во всех точках происходят с одной частотой, т. е. мы ограничим себя рассмотрением полей, имеющих гармоническую зависимость от времени.

Обычно, когда рассматривается только гармоническая зависимость от времени, то это делается только в качестве первого, более простого этапа рассмотрения. В дальнейшем на основе рассмотрения монохроматических сигналов с помощью преобразования Фурье можно получить возможность рассмотрения явлений и в широком спектре.

Рассматривая гармоническую зависимость, не будем преследовать цель получить исходный материал для решения общей задачи путем последующего преобразования Фурье. Выделим одну спектральную компоненту с совсем другой целью. Дело в том, что в монохроматическом поле изображение имеет иные свойства, чем в поле широкополосного источника. Изображение в монохроматическом поле значительно богаче. При расширении полосы частот никакие новые эффекты не появляются, но зато многие пропадают. Получается так: чтобы показать всю многокрасочность изображения, надо использовать всего одну краску — монохроматический сигнал. Самые интересные особенности изображения наблюдаются только при монохроматическом освещении.

Подчеркнем здесь одну, пожалуй самую важную, особенность регулярного поля. Монохроматическое поле всегда является регу-

лярным, так как оно всегда описывается известным математическим выражением и полностью предсказуемо. Однако, если поле является случайным, то ему тоже может быть придана нужная степень монохроматичности. Этого можно добиться путем фильтрации по частоте принимаемого сигнала. Из случайного поля можно выделить монохроматическую компоненту такой ширины, которая достаточна для наблюдения тех же эффектов, что и в монохроматическом поле. Единственным реальным препятствием на пути является падающая при этом интенсивность выделяемой спектральной компоненты. Это препятствие в некоторых случаях вполне преодолимо. Однако существует более принципиальное ограничение. В общем случае нельзя путем фильтрации случайного сигнала сделать его регулярным в такой степени, чтобы получались такие изображения.

### 1.1. Параметры волнового поля, существенные для получения изображений

Такие параметры введем, рассмотрев самую простую волну, а именно плоскую волну в свободном пространстве. Запишем выражение для плоской монохроматической волны в виде

$$2A \cos(\omega t - \mathbf{k}\mathbf{r} - \varphi) = Ae^{-i\varphi} e^{i\omega t - i\mathbf{k}\mathbf{r}} + Ae^{i\varphi} e^{-i\omega t + i\mathbf{k}\mathbf{r}}. \quad (1.1)$$

Первый параметр, который нам будет нужен, — это комплексная амплитуда волны. Для плоской монохроматической волны комплексная амплитуда будет

$$P_{пл} = g e^{i\mathbf{k}\mathbf{r}}, \quad (1.2)$$

где  $g = A e^{i\varphi}$ .

Комплексная амплитуда является важным параметром, так как она фактически несет всю информацию о волне. Для того чтобы по комплексной амплитуде восстановить полное поле в том виде, как оно записано в (1.1), достаточно умножить ее на  $e^{-i\omega t}$  и после этого прибавить комплексно сопряженный член. Таким образом, комплексная амплитуда не просто параметр, а эквивалент волне.

Далее всюду вместо выражения для волны будем употреблять комплексную амплитуду.

С целью большего удобства преобразуем выражение для комплексной амплитуды, введя в него в явном виде параметры, опре-

деляющие направление распространения волны, с этой целью запишем выражение (1.2) в виде

$$p(x, y, z) = g \exp(\pm iz \sqrt{k^2 - u_1^2 - u_2^2}) \exp(iu_1x + iu_2y), \quad (1.3)$$

$$k^2 = k_x^2 + k_y^2 + k_z^2, \quad (1.4)$$

где  $u_1 = k_x, u_2 = k_y$  — компоненты волнового вектора,  $g$  — комплексное число, учитывающее начальную фазу.

Выражение (1.3) записано в определенной системе координат с выделением оси  $z$ . Такая система координат и выделение в ней одного направления будут в дальнейшем обоснованы нами. Большой класс рассматриваемых задач будет таким, при котором волны распространяются фактически в одном направлении под небольшим углом к нему. В качестве такого направления будет служить ось  $z$ .

На основе выражения (1.3) напишем выражения для произвольного монохроматического волнового поля. Комплексная амплитуда произвольного волнового поля будет

$$p(x, y, z) = \frac{1}{4\pi^2} \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u_1, u_2) \exp(\pm iz \sqrt{k^2 - u_1^2 - u_2^2}) \exp(iu_1x + iu_2y) du_1 du_2. \quad (1.5)$$

Выражение (1.5) представляет собой сумму плоских волн с произвольными амплитудами, фазами и направлениями распространения. Так как мы просто сложили поля, то смысл, вложенный нами в понятие комплексной амплитуды плоской волны, при рассмотрении в (1.5) сохраняется.

Обратим внимание на две особенности выражения (1.5).

1) Это есть сумма плоских волн, направления которых определяются величинами  $u_1, u_2$ . Малые значения  $u_1, u_2$  означают, что соответствующая волна распространяется под небольшим углом к оси  $z$ .

2) Среди слагаемых (1.5) есть и неоднородные волны, которые затухают в направлении вдоль оси  $z$ . Это такие волны, для которых справедливо условие

$$u_1^2 + u_2^2 > k^2. \quad (1.6)$$

В зависимости от выбранного знака перед радикалом в (1.5) получим либо нарастающее, либо убывающее поле вдоль оси  $z$ .



Заметим, что выражение (1.5) представляет собой общее решение волнового уравнения, что означает возможность представления любого волнового поля в виде (1.5), т. е. в виде суммы плоских волн различных направлений распространения.

Введем еще одно уточнение. Будем считать, что все источники волн находятся в полупространстве  $z < 0$ , а полупространство  $z > 0$  свободно от источников. В этом случае в выражении (1.5), написанном для полупространства  $z > 0$ , не может быть нарастающих неоднородных волн. Отсюда следует, что в выражении (1.5) должен стоять знак плюс перед радикалом.

Теперь зададимся вопросом: что такое изображение?

Этот вопрос является центральным во всей теории получения изображений. Дело в том, что определения того, что такое изображение, могут быть разными, а часто и должны быть разными. Естественно, что от определения зависят все его свойства. Не углубляясь в этот вопрос, дадим одно из определений понятия "изображение", которое будет широко использоваться в дальнейшем. Назовем изображением распределение комплексной амплитуды в плоскости  $z = \text{const}$ .

Такое определение изображения имеет свои плюсы и минусы. Отметим сейчас лишь некоторые из них. Существенным плюсом такого определения является то обстоятельство, что, задав изображение на одной любой плоскости  $z = \text{const}$ , можно получить изображение в любой другой плоскости, т. е. изображения определены для всего полупространства. Это настолько важное свойство подобным способом заданных изображений, что докажем его подробно.

Комплексная амплитуда для всех  $z$  определяется выражением (1.5), а в этом выражении при условии выбора знака перед радикалом все определяется заданием функции  $g(u_1, u_2)$ . Если эта функция известна, то и все изображения для любых значений  $z$  известны. Функцию  $g(u_1, u_2)$  можно определить по одному изображению. Рассмотрим для примера изображение в плоскости  $z = 0$ . Из выражения (1.5) для этого случая получаем

$$p(x, y, 0) = \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} g(u_1, u_2) \exp[i(u_1 x + u_2 y)] du_1 du_2. \quad (1.7)$$

Это выражение представляет собой известный интеграл Фурье. Пользуясь свойствами преобразования Фурье, можно написать решение интегрального уравнения (1.7) в виде

$$g(u_1, u_2) = \iint_{-\infty}^{\infty} p(x, y, 0) e^{-i(u_1 x + u_2 y)} dx dy. \quad (1.8)$$

Выражение (1.8) позволяет по изображению в плоскости  $z = 0$ , т. е. по  $p$ , найти функцию  $g$ , а следовательно по (1.5) и изображение для любой плоскости  $z$ . Это важное свойство изображений, заданных в виде комплексной амплитуды, и мы его в дальнейшем формализуем.

Минусом такого определения изображения является его полная ненаблюдаемость. Это существенный минус, который ставит под вопрос саму целесообразность такого определения понятия "изображение", так как под изображением привычно понимать то, что можно увидеть глазами, услышать ушами или зафиксировать на фотопластинке. Комплексную амплитуду нельзя ни видеть, ни слышать и нельзя просто так фиксировать на фотопластинках. Однако, несмотря на этот очевидный минус, плюс все-таки перевешивает — это с учетом того обстоятельства, что многие излучения все равно невидимы и неслышимы, как их ни определяй, а комплексная амплитуда обладает еще одним плюсом — зная ее, можно вычислить практически любое другое возможное определение изображения. Например, модуль комплексной амплитуды, возведенный в квадрат, пропорционален энергии волны и может фиксироваться глазом, фотопластинкой или ухом. Кроме этого, появились хитрые способы фиксировать сами значения комплексной амплитуды на фотопластинках и других средах, которые реагируют на энергию волны. Такие способы получили название голографических, так как на них основана голография.

На плюсы и минусы таким образом определенного понятия изображения необходимо взглянуть и с иной точки зрения. Допустим, что, приняв такое определение, найдем некоторые общие закономерности получения изображений, выясним ряд условий, при которых хорошие изображения могут или не могут быть получены. Будут ли эти найденные закономерности справедливы и при других определениях того, что такое изображение? Ответ отрицательный. При других определениях будут совсем другие условия, которые это изображение формируют. Однако, если под изображением понимать распределение комплексных амплитуд, то такое определение позволяет не только обозначить характеристики, его формирующие, но и выснить условия формирования изображений, определенных по-другому. Обратное же несправедливо. В свое время именно это обстоятельство вызвало так называемую "революцию в оптике". Развитие когерентного света и голографии

позволило реализовать ряд возможностей, вытекающих из определения изображения как распределения комплексных амплитуд. С точки зрения распространенных тогда понятий об изображении (что видно или слышно), появившиеся возможности были абсолютно непонятны не только студентам, но и профессорам, а когда профессорам приходится переучиваться — это и есть революция в науке.

Чтобы нам не пришлось переучиваться, начнем с такого определения изображения, которое позволит понять особенности и необходимость иных определений.

Приведем пример плодотворности введенного нами понятия изображения и его представления в виде (1.5). Обратим внимание на то, что параметры  $u_1$ ,  $u_2$ , определяющие в (1.5) направления распространения плоских волн, имеют еще один физический смысл. В интеграле Фурье (1.7) эти параметры играют роль пространственных частот. Это позволяет сделать общий вывод о качестве изображения, которое может быть получено в любой изображающей системе, — о нем принято судить по числу наиболее мелких деталей. Этому соответствует широкий пространственный спектр. Ширина пространственного спектра определяется величиной интервала параметра  $u_{1,2}$  теми величинами  $u_1$ ,  $u_2$ , которые проходят через систему, формирующую изображение. На основе того, что  $u_1$ ,  $u_2$  являются проекцией волнового вектора, могут быть написаны следующие соотношения, связывающие пространственные частоты  $u_1$ ,  $u_2$  с углами, под которыми распространяются волны:

$$u_1 = k \sin \alpha; \quad U_2 = k \cos \alpha \sin \varphi; \quad k = 2 \pi / \lambda. \quad (1.9)$$

Эта зависимость замечательна тем, что пространственный спектр определяется очень малым числом параметров. Всего два параметра определяют ширину пространственного спектра — это длина волны и тот максимальный угол, под которым волны входят в изображающее устройство. Эти два параметра определяют максимально возможное качество получаемого изображения, причем связь получается очень простая. Например, если углы  $\alpha$ ,  $\varphi$  достаточно малы (позже станет ясно, что это вполне реальная ситуация), то ширина пространственного спектра определяется произведением  $k\alpha$ . В этих условиях улучшить качество изображения можно только за счет увеличения угла  $\alpha$  или уменьшения длины волны. Естественно, что из-за различных неточностей работы изображающей системы качество изображения, которое она дает, может быть и хуже и сколько угодно хуже, чем то, которое определяется по соотношению (1.9), но важно заметить, что лучше оно не может

быть. Соотношение (1.9) позволяет обозначить тот предел качества изображения, которое может быть достигнуто в системе при идеальной ее работе. Связь пространственных частот, определяющих качество изображения и спектра (набора) углов, под которыми распространяются волны, является одной из самых фундаментальных закономерностей той науки, о которой здесь идет речь. Если бы требовалось выделить всего одно самое важное соотношение, самый важный закон из рассматриваемого в данной монографии, то я бы выделил именно эту связь пространственных частот изображения и углового спектра волн, его образующих или формирующих. Все остальные соотношения, закономерности, формулы являются простым следствием этого соотношения, поэтому-то его и можно рассматривать как фундаментальное.

Отсюда выведем следующее правило. *Для улучшения возможного качества получаемого изображения существуют два пути — это либо укорочение длины волны, либо увеличение угла, под которым волны могут попасть в изображающую систему.*

Других путей нет.

Казалось бы, при поиске пределов возможного качества получаемых изображений совершенно ни к чему знать, как на самом деле такая система устроена. Докажем теперь, что это не так. Система может иметь такое устройство, которое позволяет ей получать изображения любого качества (т. е. в очень широком диапазоне пространственных частот, в принципе ничем не ограничиваемом) при небольшом, даже сколь угодно малом, угле, под которым волны могут попадать в систему. Радиофизик, знакомый с основами радиотехники, знает, как такая система должна быть устроена. Покажем это.

В системе должно быть устройство, которое может запоминать пространственный спектр сигналов, проходящих в него, и которое способно сдвигать этот спектр по оси пространственных частот на заданное расстояние. Такое устройство должно работать следующим образом. Каждый раз при освещении через систему проходит часть пространственного спектра изображения, которая запоминается. Далее производится смещение пространственных частот в спектре путем умножения изображения на множитель вида  $e^{inx + imy}$ . Такой процесс называется в радиотехнике гетеродинированием и очень широко применяется. Для изображений процесс гетеродинирования также освоен и может применяться с помощью ультразвуковых ячеек. Процесс гетеродинирования переводит спектр изображения из той области, которая не могла пройти через систему, в ту область частот, которая через систему проходит. На выходе системы спектры запоминаются, сдвигаются в нужный

диапазон, и по ним может быть восстановлено изображение, имеющее очень широкий пространственный спектр. Все перечисленные операции, в принципе, возможны: в ряде случаев осуществимы, а в ряде случаев могут быть даже упрощены за счет получения нужной информации без операции восстановления изображения по его пространственному спектру, т. е. получая необходимую информацию непосредственно из спектра изображения или даже определяя некоторые параметры спектра изображения. Все это верно и сулит большие перспективы, однако на практике ничего такого применить не удастся. Почему? Запомним этот вопрос. Ответ на него будет дан ниже и не совсем уж отрицательный. Пока сделаем для себя следующий важный вывод. В рассматриваемой области науки категоричные выводы о возможности или невозможности чего-либо надо делать очень осторожно.

Для дальнейшего нам надо немного формализовать полученные нами результаты и сделать их частью более широких представлений о преобразовании сигналов.

## **1.2. Частотная характеристика и отклик свободного пространства**

В заголовке стоят понятия, взятые из теории линейных систем, или, как иногда называют, из радиотехники линейных фильтров. Будем придерживаться этой терминологии.

Для того чтобы система была линейной, необходимо и достаточно, чтобы входные и выходные сигналы этой системы удовлетворяли принципам суперпозиции и транспозиции. Принцип суперпозиции состоит в следующем: если на вход системы подан сигнал в виде суммы сигналов, то и выходной сигнал тоже может быть представлен в виде суммы сигналов, причем каждое из слагаемых как входного, так и выходного сигналов являются также входными и выходными сигналами той же системы. Принцип транспозиции: если сдвигается время начала входного воздействия, то на такое же время без изменений своей формы сдвигается и выходной сигнал.

Само название "линейная система" обязано своим происхождением тому, что такие системы описываются линейными дифференциальными уравнениями с постоянными коэффициентами. Подробнее об этом сказано в разд. 2.1 "Радиооптики" [1].



Из принципов суперпозиции и транспозиции следует, что вход и выход любой линейной системы связаны следующими интегральными соотношениями:

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c_x(\omega) k(i\omega) e^{i\omega t} d\omega, \quad (1.10)$$

$$y(t) = \int_{-\infty}^{\infty} x(t') h(t-t') dt'. \quad (1.11)$$

В соотношении (1.10) выход линейной системы выражен через спектр входного воздействия  $c_x(\omega)$  и характеристику самой системы, которая однозначно определяет все ее свойства. Эта характеристика является частотной характеристикой линейной системы и формально может быть определена как отклик системы на воздействие вида  $e^{i\omega t}$ .

Соотношение (1.11) определяет выход линейной системы непосредственно по ее входному сигналу  $x(t)$ . Функция  $h(t)$  называется импульсной характеристикой системы и формально является откликом системы на сигнал вида  $\delta(t)$ .

Каждая из перечисленных характеристик однозначно определяет систему, и значит, они обязательно должны быть связаны между собой некоторым универсальным соотношением, не зависящим от свойств самой системы. Эти характеристики сопряжены между собой по Фурье:

$$k(i\omega) = \int_{-\infty}^{\infty} h(t) e^{-i\omega t} dt; \quad h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} k(i\omega) e^{i\omega t} d\omega. \quad (1.12)$$

Входной сигнал и его спектр связаны следующими соотношениями:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c_x(\omega) e^{i\omega t} d\omega; \quad c_x(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt. \quad (1.13)$$

Посмотрим теперь на соотношение (1.5) и сравним его с (1.10). Между ними есть аналогия, разница заключается только в обозначениях и в числе измерений, но оба отличия несущественны. Отсюда можно утверждать, что свободное пространство представляет собой некоторую линейную систему. Входным сигналом является распределение комплексных амплитуд на плоскости  $z = 0$ , а выходным сигналом — распределение комплексных амплитуд на плос-

кости  $z = \text{const}$ . Частотной характеристикой свободного пространства является функция вида

$$\zeta(u_1, u_2) = e^{iz \sqrt{k^2 - u_1^2 - u_2^2}}. \quad (1.14)$$

Найдем теперь отклик свободного пространства. Соответствующие вычисления приведены в разд. 2.2 [1]. Точное выражение для отклика получается путем преобразования Фурье частотной характеристики (1.14) и выглядит следующим образом:

$$H(x, y) = -\frac{1}{2\pi} ik \frac{e^{ikR}}{R} \frac{z}{R} + \frac{1}{2\pi} \frac{1}{R^2} e^{ikR} \frac{z}{R},$$

$$R = \sqrt{x^2 + y^2 + z^2}. \quad (1.15)$$

Однако это выражение почти не встречается в литературе (причины выясним позже), а всюду используется приближенное (приближенность не всегда указывается) выражение для отклика свободного пространства, справедливое при следующих условиях:

$$k \gg \frac{1}{R} \text{ или } R \gg \lambda; \quad z \approx R \text{ или } z^2 \gg x^2 + y^2. \quad (1.16)$$

При этих условиях выражение (1.15) можно записать в виде

$$H(x, y) = \frac{1}{i\lambda} \frac{e^{ikR}}{R}. \quad (1.17)$$

Входной сигнал свободного пространства можно представить в виде, аналогичном (1.11). Это выражение имеет вид

$$p(x, y, z) = \frac{1}{i\lambda} \iint_{-\infty}^{\infty} p(x', y', 0) \frac{e^{ikR}}{R} dx' dy';$$

$$R = \sqrt{(x - x')^2 + (y - y')^2 + z^2}. \quad (1.18)$$

Полученное соотношение хорошо известно. Оно называется иногда математическим выражением принципа Гюйгенса — Френеля. Неожиданным является то, что здесь это соотношение выступает как приближенное, справедливое лишь при условиях (1.16), в то время как во всех учебниках по общей физике его приближенный характер не отмечается.

Отметим, что приближенность соотношения (1.18) прекрасно снимает известный парадокс со светлым пятном в центре тени от непрозрачного круглого диска. Результат этого известного опыта, поставленного в свое время Араго по предложению Пуассона, может

быть вычислен на основе формулы (1.5). На рис. 1П (см. приложение в конце книги) приведен результат такого вычисления: слева — форма непрозрачного круглого диска, который использовался в расчетах в качестве входной функции  $P(x, y, 0)$ ; справа показан результат вычисления интенсивности поля на расстоянии, равном одной зоне Френеля от плоскости диска, это расстояние равно  $R^2/\lambda$ . Подробности вычислений и соответствующая математическая программа приведены ниже в разд. 1.8.1. Эксперимент и расчет, основанный на приведенных здесь формулах, подтверждают существование только одного светлого пятна в центре тени, находящегося на расстоянии одной зоны Френеля. Расчет же по приближенной формуле (1.18) дает целую серию таких светлых пятен, располагающихся все ближе и ближе к поверхности непрозрачного диска. Наличие такого результата говорит о том, что в этих областях уже не справедливы условия приближения (1.16). Физический смысл этих условий определим после того, как напишем в том же приближении частотную характеристику свободного пространства. Для этого применим преобразование Фурье к (1.17) и получим

$$\zeta(u_1, u_2) = k (\sqrt{k^2 - u_1^2 - u_2^2})^{-1} \exp(iz \sqrt{k^2 - u_1^2 - u_2^2}). \quad (1.19)$$

Выражения (1.17) и (1.19) являются характеристиками свободного пространства, полученными в одинаковом приближении (1.16). Однако, сравнивая (1.19) с точной формулой (1.14), можно придать условиям (1.16) иной вид, а именно:

$$u_1^2 + u_2^2 \ll k^2. \quad (1.20)$$

Чтобы было справедливо и (1.17), и (1.19), достаточно одного условия (1.20), которое имеет простой физический смысл: все плоские волны, составляющие волновое поле, должны распространяться под небольшим углом к оси  $z$ . Это условие называется условием малоуглового приближения. Подобный физический смысл и название можно отнести и к двум условиям (1.16).

Малоугловое приближение не сулит изображений хорошего качества. Малые значения  $u_1, u_2$  — это малые пространственные частоты и отсутствие достаточно мелких деталей в изображении по сравнению с изображениями, полученными в условиях больших значений  $u_1, u_2$ .

### 1.3. В каких случаях и почему необходимо малоугловое приближение?

Откуда же взялось и почему так употребительно малоугловое приближение? Уж не является ли это следствием стремления к простоте выражения (1.17)?

Малоугловое приближение необходимо при решении задач видения с помощью волновых полей. Этот краткий ответ является почти исчерпывающим. Однако, хотя добавить к нему практически ничего нельзя, он нуждается в разъяснении. Этим разъяснениям и посвящен данный раздел.

Прежде всего поясним более подробно, что имеется в виду, когда речь идет о задачах видения. Типичная постановка такой задачи выглядит следующим образом.

В плоскости  $z = 0$  существует некоторая неоднородность, имеющая ограниченный размер. Эта неоднородность может представлять собою отверстие в непрозрачном для волн экране. Форма отверстия может быть сложной. Осветив весь экран на участке, включающем отверстие, плоской или сферической волной и находясь при этом на некотором расстоянии  $z$  от плоскости  $z = 0$ , по распределению комплексных амплитуд в этой плоскости можем определить форму отверстия. Аналогичным образом ставится задача по определению формы непрозрачного для волн экрана, который в плоскости  $z = 0$  помещен в поле плоской или сферической волны: надо по распределению комплексных амплитуд в плоскости  $z$  определить форму этого непрозрачного экрана. То же самое и для прозрачных неоднородностей — речь идет о распределении показателя преломления (скорости распространения волн) внутри тела. В соответствующей литературе их иногда называют дифракционными задачами.

Решение такого рода задач имеет два этапа — это так называемые прямые и обратные задачи. Прямая задача решается для заданной формы и материала неоднородности, когда требуется найти распределение комплексной амплитуды в плоскости  $z$  при освещении неоднородности заданной формы плоской или сферической волной. Прямые задачи сводятся к нахождению волнового поля, дифрагирующего на неоднородности, и поэтому называются дифракционными. Задачи видения являются по отношению к дифракционным обратными задачами, которые, как правило, решаются на втором этапе, когда решена дифракционная задача. Таким образом, задачи видения и дифракционные задачи тесно связаны друг с другом. Решение дифракционной задачи является первым этапом решения задачи видения.

Очень важно отметить, что, если дифракционная задача успешно решена, это вселяет надежду на то, что может быть решена и обратная задача — задача видения, хотя это и не всегда так. Однако, если дифракционная задача не может быть решена, то и задача видения заведомо не решается. Оказывается, что дифракционную задачу можно решить только в малоугловом приближении. При отказе от малоуглового приближения и расширении области пространственных частот  $u_1, u_2$  за пределами неравенства (1.20) решение дифракционной задачи не найдено до сих пор, хотя такая задача стоит перед наукой уже более ста лет. Это утверждение требует некоторого разъяснения. Поясним более подробно, в чем здесь состоят трудности.

Дифракционную задачу тоже можно разделить на два этапа. Первый этап состоит в том, чтобы найти волновое поле (распределение комплексных амплитуд) в плоскости  $z = 0$ , а второй этап состоит в нахождении поля в плоскости  $z$ . На втором этапе дифракционная задача решается без каких-либо затруднений, вполне точно и не требует для упрощения решения никаких приближений. Это решение дается соотношением (1.5), которое особенно просто выглядит, если его записать для пространственных спектров комплексных амплитуд:

$$g_z(u_1, u_2) = \zeta(u_1, u_2) g_0(u_1, u_2). \quad (1.21)$$

Здесь  $g_0$  — пространственный спектр поля в плоскости  $z = 0$ ,  $\zeta(u_1, u_2)$  — пространственный спектр поля в плоскости  $z$ .

Соотношение (1.21) позволяет весьма просто, без всяких приближений находить решение как дифракционной задачи, так и обратной задачи — задачи видения на соответствующем этапе.

Однако есть еще и первый этап задачи, на котором требуется найти поле в плоскости  $z = 0$ . Именно этот этап и представляет трудность. В малоугловом приближении задача первого этапа решается очень просто. Правила для решения таких задач хорошо известны. Они были предложены Кирхгофом сто лет тому назад и состоят в следующем.

Поле возмущено только там, где находится сама неоднородность. В тех местах плоскости  $z = 0$ , где нет неоднородности, поле остается таким, каким оно было до внесения неоднородности. Таким образом, если имеется отверстие сложной формы в непрозрачном экране, освещаемом плоской волной, то всюду внутри отверстия колебания будут синфазными. Этого достаточно для того, чтобы найти пространственный спектр такой неоднородности, по



которому с помощью (1.21) можно найти поле на любом расстоянии от неоднородности.

Для прозрачных экранов тоже существует соответствующее очень простое правило нахождения поля в плоскости расположения самой неоднородности. Внутри тела проводятся "лучи", совпадающие с направлением волновых векторов в свободном поле (без неоднородностей). Вдоль этих "лучей" путем интегрирования находятся фазы колебаний на некоторой плоскости, располагающейся непосредственно за неоднородностью. Полученные значения фаз колебаний определяют значения комплексных амплитуд поля в этой плоскости. Такого распределения фаз тоже достаточно для того, чтобы найти пространственный спектр поля.

Эти правила распространяются и на решение обратных задач — задач видения. Получив значения полей в плоскости  $z = 0$ , можно отождествить их с самим предметом на основе тех же правил.

Возникает вопрос, насколько такие простые правила Кирхгофа на самом деле справедливы. Ответ следующий: они справедливы при малоугловом приближении. Это подтверждается в тех случаях, когда дифракционную задачу удастся решить точно, — для отверстий и экранов простой формы в виде круглого диска или эллипсоида. Здесь точное решение и решение, полученное согласно правил Кирхгофа, совпадают в пределах малоуглового приближения. Для тел сложной формы точных решений нет. В этом случае справедливость малоуглового приближения может быть установлена чисто опытным путем или доказана исходя из некоторых общих соображений. На этом подробно останавливаться не будем. Рассмотрим условия, при которых правила Кирхгофа не работают за пределами приближения (1.20). То, что эти правила не должны работать всюду, ясно хотя бы из того, что комплексная амплитуда поля не может быть произвольной функцией координат, а обязана удовлетворять волновому уравнению. Волновому уравнению удовлетворяет далеко не любая функция, а форма отверстия в экране может быть любой, и экран тоже может быть любым и совсем не обязан иметь такую форму, которая удовлетворяла бы волновому уравнению. Отличия в распределении комплексной амплитуды от тех значений, которые получаются в силу применения правил Кирхгофа, будут на высоких пространственных частотах и будут тем сильнее, чем пространственная частота ближе к  $k$ .

Из рассмотрения этого, казалось бы, элементарного вопроса, можно извлечь ряд следствий, которые, или по крайней мере некоторые из них, уже не будут столь очевидными.

1. Сначала сформулируем самое существенное для нас следствие. Так как задачи видения могут быть решены только в условиях

малоуглового приближения, то не имеет смысла каким-то путем стремиться выйти за его пределы, решая подобные задачи, не имеет смысла различать выражения для поля вида (1.5), справедливые всегда, и выражение (1.18), справедливое только при малоугловом приближении. Лишена смысла и всякая изобретательская деятельность, направленная на создание изображающих систем с очень широким охватом углов, позволяющих выйти за пределы малоуглового приближения.

2. Малоугловое приближение на практике может выступать не как некоторый теоретический предел наших возможностей, а как чисто техническое ограничение. В ряде случаев дело так и обстоит. Например, при наблюдении удаленных звезд и даже планет мы никогда не дойдем до принципиальных ограничений и будем все время оставаться в рамках малоуглового приближения, стараясь изо всех сил увеличить углы охвата принимаемого излучения. Эта точка зрения на малоугловое приближение как на некоторый технический предел допустима только до поры, пока совершенствование не перейдет определенной черты. К сожалению, этой черты провести однозначно нельзя, так как она зависит от формы того предмета, который мы хотим увидеть. К чему это может привести? К тому, что некоторые предметы видны правильно и такими, каковы они есть в действительности, а другие видны неправильно и совсем не такими, каковы они есть на самом деле. Поэтому надо с большой осторожностью относиться к тому, что видит исследователь в микроскоп, объектив которого почти уперся в объект исследования для увеличения угла охвата излучения и возможности наблюдения мелких деталей изображения. При этом мелкие детали изображения на самом деле появляются, но они могут не иметь ничего общего с тем предметом, на который направлен объектив микроскопа. Дело осложняется тем, что исследователя, смотрящего в такой сверххороший микроскоп, вышеприведенными рассуждениями убедить практически невозможно, а поймать за руку нельзя, так как ничем другим, кроме микроскопа, посмотреть на объект нельзя. Другое положение в акустике. Там тоже есть микроскопы, и они также строятся для получения хороших изображений с учетом большого охвата углов. Однако там можно на тот же объект взглянуть иначе и убедиться, что увеличение угла охвата никогда не ведет к получению изображения, совпадающего с объектом рассмотрения. Акустические изображения всегда плохие, так как соответствуют объекту только в пределах малоуглового приближения, независимо от того, насколько совершенна изображающая система.

3. Так обстоит дело только с задачами видения, которые были сформулированы выше и решение которых связано с решением дифракционных задач. Однако есть другой класс задач видения, который не связан с решением дифракционных задач. Речь идет о наблюдении источников излучения, когда задача видения может быть сформулирована как задача определения координат источников излучения. В этом случае задача разделяется тоже на прямую и обратную, но решение прямой задачи уже не связано с решением дифракционной задачи. Соотношение (1.18) уже является точным, а соотношение (1.5) будет справедливо только в малоугловом приближении. Такого рода задачи не содержат принципиальных ограничений на величины  $u_1$ ,  $u_2$  и могут успешно ставиться и решаться и в условиях существенного нарушения условия (1.20). Однако сейчас мы не будем заниматься решением таких задач, так как задача о видении источников в регулярных полях может быть поставлена не как дифракционная только в том случае, если имеется всего один источник. Если источников несколько, то это либо дифракционная задача, несколько иначе сформулированная, но имеющая те же ограничения, либо задача, не относящаяся к области регулярных полей. Для того чтобы поле было регулярным необходимо, чтобы все источники излучали поля с неизменным во времени соотношением фаз. В случае независимых источников это не получается, а в случае, когда источник один, а остальные создаются с помощью его излучения, всегда возникает дифракционная задача. Поэтому задачу видения источников следует отложить до появления источников случайных колебаний, когда задача становится типичной, а дифракционная задача отодвигается на второй план.

В связи с этим можно вернуться несколько назад к вопросу, а что же наблюдает исследователь в сверххороший микроскоп. Если объект исследования удастся осветить так, чтобы отдельные его точки светились самостоятельно и поле, излучаемое объектом, было бы нерегулярным, то исследователь в такой микроскоп имеет шанс увидеть то, что ему на самом деле надо.

4. В этом пункте хотелось бы коротко обобщить сказанное выше. Очень важно — что понимать под изображением. Если это распределение комплексной амплитуды, то прямая и обратная задачи решаются точно с помощью (1.21) для любых  $u_1$ ,  $u_2$ . Если под изображением понимать форму самого предмета, то все становится сложнее и задача может быть сведена к предыдущей только в условиях малоуглового приближения. Если под изображением понимать координаты источников, то задача решается совсем иначе, но в регулярных полях такая задача не возникает. Малоуг-

ловое приближение может выступать как принципиальное ограничение и как чисто техническое ограничение. При решении задач видения в регулярных полях малоугловое приближение выступает как принципиальное ограничение до тех пор, пока мы не научимся решать соответствующие электродинамические или механические задачи.

#### 1.4. Упрощения при малоугловом приближении

Малоугловое приближение допускает и ряд упрощений в формулах, а если оно приобрело принципиальный характер и от него нельзя так просто отказаться, то целесообразно использовать, по крайней мере, те упрощения, которые оно допускает.

При выполнении условия (1.20) радикал в выражении для частотной характеристики свободного пространства (1.14) можно разложить в степенной ряд вида

$$\begin{aligned} \sqrt{k^2 - (u_1^2 - u_2^2)} &= \\ &= k \left[ 1 - \frac{1}{2} \frac{(u_1^2 - u_2^2)}{k^2} + \frac{1}{2^3} \frac{(u_1^2 - u_2^2)^2}{k^4} - \dots \right], \end{aligned} \quad (1.22)$$

который быстро сходится и в нем не должно быть более одного члена. Однако в выражении (1.14) этот ряд стоит в показателе экспоненты, и поэтому быстрая сходимость ряда еще не является достаточным условием того, что в этом ряду можно оставить всего один член. В данном случае должен быть применен другой критерий, который строится с учетом того, что экспонента в выражении (1.14) является периодической функцией с периодом  $2\pi$ . Поэтому при укорочении ряда (1.22) следует сравнивать отбрасываемый член не с суммой всех предыдущих членов ряда, а с величиной  $2\pi$ .

Таким образом, сохранить один первый член ряда (1.22) можно при условии

$$z \frac{(u_1^2 - u_2^2)}{2k} \ll 2\pi. \quad (1.23)$$

Это есть условие применимости геометрической оптики. В разд. 3.2 [1] показано, как из общего условия (1.23) может быть получен обычный частный случай этого условия.

В приближении геометрической оптики изображение не преобразуется свободным пространством, и этот случай не особенно интересен.

Для сохранения первых двух членов ряда (1.22) необходимо выполнение следующего условия:

$$z \frac{(u_1^2 + u_2^2)^2}{2^3 k^3} \ll 2\pi. \quad (1.24)$$

При его выполнении характеристика свободного пространства приобретает вид

$$\zeta(u_1, u_2) = \exp \left[ ikz - i \frac{z}{2k} (u_1^2 + u_2^2) \right]. \quad (1.25)$$

Это приближение известно как приближение дифракции Френеля и справедливо в пределах  $z$ , удовлетворяющих условию (1.24).

Покажем, что выражение (1.25) может быть использовано не только при условии (1.24), т. е. при ограничениях  $z$ , но и при всех значениях  $z$ . На первый взгляд кажется, что этого не может быть, так как на основании разложения радикала в ряд (1.22) и условия (1.24) ничего другого получить нельзя. Здесь как раз очень удобно дальнейшее изложение построить в виде некоторого парадокса. Для того чтобы получился парадокс, продолжим наше рассмотрение следующим образом.

Рассмотрим отклик свободного пространства в том же малоугловом приближении. В силу условий (1.16) расстояние  $R$  можно представить в виде степенного ряда:

$$\begin{aligned} \sqrt{z^2 + (x^2 + y^2)} &= \\ &= z \left[ 1 + \frac{1}{2} \frac{(x^2 - y^2)}{z^2} + \frac{1}{2^3} \frac{(x^2 + y^2)^2}{z^4} + \dots \right]. \end{aligned} \quad (1.26)$$

В выражении для самого  $R$  ряд (1.26) можно укоротить вплоть до одного члена, а при подстановке в (1.18) следует сравнивать отбрасываемый член с  $2\pi$ .

На этом основании при условии

$$\frac{k(x^2 + y^2)^2}{2^3 z^3} \ll 2\pi \quad (1.27)$$

укоротим ряд (1.26) до двух членов и получим отклик в виде

$$H(x, y) = \frac{1}{ix} \frac{\exp \left[ ikz + \frac{ik}{2z} (x^2 + y^2) \right]}{z}. \quad (1.28)$$



Полученное выражение в силу условия (1.27) справедливо для больших значений  $z$ . Найдем теперь выражение для частотной характеристики свободного пространства на основании отклика вида (1.28). Пользуясь соотношением (1.12), получим следующее выражение:

$$\zeta(u_1, u_2) = \exp \left[ ikz - \frac{z}{2k} (u_1^2 + u_2^2) \right]. \quad (1.29)$$

Теперь сравним то, что мы получили, с выражением (1.25). Легко убедиться, что оба выражения полностью совпадают. Оба получены в малоугловом приближении.

Парадокс состоит в том, что выражения (1.25) и (1.29) получены во взаимно исключающих друг друга условиях. Выражение (1.25) справедливо только при малых  $z$  по условию (1.24), а выражение (1.29) справедливо только при больших значениях  $z$  по условию (1.27). Но при этом оба выражения полностью совпадают. Следовательно, (1.29) справедливо и при малых  $z$ , а (1.25) справедливо и при больших  $z$ , что мы и взялись доказывать. Однако эти утверждения противоречат условиям вывода как выражения (1.29), так и выражения (1.25). Вот и получилось то, что принято называть парадоксом. Кажется, нарушены все незыблемые законы физики и математики, хотя мы их всеми силами старались соблюдать\*.

В чем дело?

Наличие парадокса, по-видимому, объясняет тот факт, что во всех учебниках по общей физике и оптике употребляется только одна форма представления распространения волны в свободном пространстве, а именно (1.18), и никогда (1.5), хотя они совершенно эквивалентны друг другу. Дело в том, что предельный переход к большим значениям  $z$  осуществляется от выражения (1.18), как только что было показано, вполне естественно; столь же естественно осуществить переход от выражения (1.5) нельзя, не наткнувшись на тот парадокс, который только что был получен.

Парадокс легко разрешить, если принять во внимание, что предельный переход к большим значениям  $z$ , исходя из (1.18) и (1.5), осуществлен нами по-разному, и поэтому нет ничего удивительного

\* Условие (1.27) оказывается недостаточным при наклонном падении освещающей волны на транспарант. В разложении (1.26) необходимо учитывать и члены с нечетными степенями по  $x$ ,  $y$ . Выражение (1.29) не имеет места. Появляются интересные особенности дифракции волн как в ближнем, так и в дальнем поле. Подробнее см.: Зверев В. А., Кособурд Т. П., Степанов Н. С. О структуре дифракционного поля во френгоферовой зоне // ДАН СССР. 1982. Т. 217, № 3. С. 604-606; Кособурд Т. П., Степанов Н. С. Особенности дифракции при косом падении света на транспарант // Оптика и спектроскопия. 1986. Т. 60, вып. 3. С. 588-592; Кособурд Т. П., Степанов Н. С. О трансформации когерентных полей плоскими структурами // Оптика и спектроскопия. 1986. Т. 61, вып. 4. С. 852-855.

тельного в том, что получились совсем разные несопоставимые результаты.

В ряде (1.26) переход к большим значениям  $z$ , как видно из этого ряда, совершается при условии, что  $x^2 + y^2 = \text{const}$ , а в ряде (1.22) переход к большим значениям  $z$  совершается при условии, что  $u_1^2 + u_2^2 = \text{const}$ . Эти условия существенно разные. Предельный переход всегда совершается при условиях, когда  $x^2 + y^2 = \text{const}$ , т. е. когда сохраняются неизменными геометрические параметры той области, где производятся измерения. Условия  $u_1^2 + u_2^2 = \text{const}$ , при которых осуществлялся предельный переход в ряде (1.22), соответствуют сохранению углов, под которыми наблюдается излучение, и практически никогда не реализуются на практике. Поэтому всегда и используется выражение (1.18) и именно на его основе осуществляется предельный переход к большим значениям  $z$ . Покажем, что при этих условиях парадокс на самом деле снимается. Для этого следует перейти в ряде (1.22) к большим значениям  $z$  при условии сохранения не углов, а геометрических размеров той области, где наблюдается излучение. При этом учтем, что справедливо малоугловое приближение, на основе которого из (1.9) можно получить

$$u_1^2 + u_2^2 = k^2 \theta^2, \quad (1.30)$$

где  $\theta$  — угол между направлением волнового вектора и осью  $z$ . Угол  $\theta$  может быть выбран через геометрические размеры  $x$  и  $y$ , расстояние  $z$  определяется следующим образом:

$$\theta^2 = \frac{x^2 + y^2}{z^2}. \quad (1.31)$$

Считая геометрические размеры неизменными и выразив  $u_1^2 + u_2^2$  через  $x^2 + y^2$  посредством (1.31) и (1.30), получим полное совпадение условий (1.24) и (1.27), откуда следует, что если из (1.5) сделать предельный переход к большим значениям  $z$  при тех же условиях, при которых делается такой переход в (1.18) (что только и является корректным), то получается одно и то же — и никаких парадоксов не возникает.

Отметим, что парадокс, связанный с тем же обстоятельством, может возникать и по другому поводу, на него можно наткнуться практически всегда, когда одно и то же явление рассматривается то на языке координат, то на языке пространственных частот. Для того чтобы подобный парадокс не возникал, необходимо с разных языков переходить на один.

Из вышесказанного сделаем для себя следующий вывод. Для всех значений  $z$  будем пользоваться только одним приближенным выражением для частотной характеристики свободного пространства, а именно выражением (1.25).

Еще одно замечание. При решении дифракционных задач и задачи видения на основе решения дифракционных задач более правильным будет исходить из выражения (1.5), а не из выражения (1.18), как делают авторы всех учебников. Выражение (1.5) справедливо и при малых  $z$ , и при больших, а выражение (1.18) при малых  $z$  несправедливо, так как при этом неминуемо нарушаются условия малоуглового приближения. Поэтому, исходя из выражения (1.5) при условии соблюдения нужных или требуемых условий предельного перехода к большим  $z$ , можно смело подходить и к области малых значений  $z$ , чего в выражении (1.18) делать нельзя.

Итак, мы рассмотрели то упрощение, которое получается в выражении для частотной характеристики свободного пространства в малоугловом приближении. Теперь рассмотрим, какое упрощение получается в этих условиях для самого значения поля на больших расстояниях  $z$ . Так как речь пойдет об очень больших, предельно больших расстояниях, то в основу рассмотрения положим выражение (1.18). С учетом (1.28) оно примет вид

$$p(x, y, z) = \frac{e^{ikz}}{i\lambda z} \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x', y', 0) \exp \left\{ i \frac{k}{2z} [(x - x')^2 + (y - y')^2] \right\} dx' dy'. \quad (1.32)$$

Упрощение (1.32) возможно при больших значениях  $z$  и ограниченной по линейным размерам области, которую занимает  $p(x', y', 0)$ . Пусть максимальный размер этой области будет  $D$ .

В этом случае при соблюдении условий

$$\frac{k}{2z} D^2 \ll 2\pi \quad (1.33)$$

выражение (1.32) можно записать в виде

$$p(x, y, z) = \exp \left[ ikz + i \frac{k}{2z} (x^2 + y^2) \right] \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x', y', 0) \exp \left( i \frac{kx}{z} x' + i \frac{ky}{z} y' \right) dx' dy', \quad (1.34)$$

где интеграл имеет вид интеграла Фурье. На основании (1.8) его можно представить в виде

$$p(x, y, z) = \frac{\exp \left[ ikz + i \frac{k}{2z} (x^2 + y^2) \right]}{i \lambda z} g \left( -\frac{kx}{z}; -\frac{ky}{z} \right). \quad (1.35)$$

Полученное упрощение представления поля на расстояниях  $z$ , достаточно больших, удовлетворяющих (1.33), называется дифракцией Фраунгофера. В этом случае поле имеет вид собственного пространственного спектра. В дальнейшем нам придется очень часто встречаться с дифракцией Фраунгофера, когда вместо поля появляется его спектр, и с условиями (1.33).

Другой вывод выражений для дифракции Фраунгофера и условий (1.33) на основе соотношения (1.5) приведен в разд. 3.6 [1].

## 1.5. Голография

Физической основой голографии является отмеченное выше свойство комплексной амплитуды волны, состоящее в том, что распределение комплексной амплитуды в какой-либо одной плоскости однозначно определяет распределение комплексных амплитуд во всех других плоскостях (при значениях  $z$  в области  $z > 0$ ).

Таким образом, если бы удалось зафиксировать в какой-либо плоскости распределение комплексных амплитуд волны, то это была бы такая запись, которая при ее воспроизведении в форме волны дала бы возможность восстановить распределение комплексных амплитуд во всех плоскостях.

Во многих случаях никакого специального процесса обработки по переносу изображения из одной плоскости в другую даже и не требуется.

Представим себе, что существует возможность фиксировать распределение комплексных амплитуд  $p(x, y, z)$  в плоскости  $z$  на площади, имеющей минимальный линейный размер  $D$ . Процесс такой фиксации пусть будет пока чисто фантастическим. На пластинке фиксируется некоторый рельеф и распределение прозрачности таким образом, что при освещении плоской монохроматической волной этой пластинки выходящая из нее волна на расстоянии, удовлетворяющем условию (1.23), повторяет значения модуля и фазы комплексной амплитуды  $p(x, y, z)$ .

Такая пластинка представляет собой наиболее полную запись изображения предмета или объекта, на котором дифрагировала

волна, образовав зафиксированное нами поле  $p(x, y, z)$ . Следуя Габору, назовем такую пластинку с записью голограммой, т. е. полной записью. Такую пластинку можно хранить, как и фотографию, а при освещении ее плоской волной, она воспроизводит не контуры предмета, а распределение комплексной амплитуды  $p(x, y, z)$ , которое было при освещении этого предмета плоской или сферической волной.

Рассмотрим, что мы зафиксировали, изготовив голограмму, и как ею можно пользоваться. Представим себе, что непосредственно перед глазом наблюдателя на расстоянии от глаза, удовлетворяющем условию (1.24), помещена голограмма, освещенная плоской волной. В глаз попадает такое же волновое поле, какое было в момент, когда освещенная волна дифрагировала на объекте. Глаз воспринимает это поле точно так же, как и поле, которое, освещая объект, попадает затем в глаз. Если голограмма сделана совершенно точно и на самом деле фаза и амплитуда выходящей из нее волны не отличаются от тех значений, которые были в момент освещения волной объекта, то у наблюдателя нет возможности с помощью только зрительных впечатлений (не ощупывая объекта) узнать, смотрит ли он на объект в момент его освещения волной или он смотрит сквозь голограмму на источник плоской волны. Если размер  $D$  голограммы достаточен для того, чтобы можно было смотреть сквозь голограмму сразу двумя глазами, то можно селективировать предметы по дальности, пользуясь бинокулярным зрением. Смещая голограмму относительно глаза, можно добиться того же эффекта, что и при смещении точки рассматривания самого объекта, т. е. можно оглядывать объект с разных сторон, просто перемещая голограмму относительно глаза. Для того чтобы зафиксировать объект со всеми его деталями, доступными для рассматривания, достаточно иметь голограмму с размером  $D$ , равным диаметру зрачка глаза.

Вместо глаза наблюдателя может использоваться любая другая система, формирующая изображение. Независимо от того, как конкретно она устроена, и при условии, что размер  $D'$  голограммы равен или больше диаметра входного зрачка системы, такая система не сможет отличить изображение, считанное с голограммы, от воспринятого непосредственно от освещенного волной объекта. Разницы между изображением, записанным с помощью голограммы, и натурально увиденным не должно быть никакой независимо от того, каким образом определено понятие "изображение" и как и что фиксируется изображающей системой на ее выходе.

Голограмма может быть изготовлена таким образом, что может формировать изображение предмета без посредства какой-либо изображающей системы в виде глаза, фотоаппарата или чего-то еще.

В основе этого утверждения лежат два положения.

1. Тожество распределения волнового поля и самого объекта, т. е. справедливость правил Кирхгофа и связанного с ними малоуглового приближения.

2. Принципиальная возможность решения обратных волновых задач, которая следует из (1.21).

В малоугловом приближении соотношение (1.21) выглядит следующим образом:

$$g_z = \exp(ikz) \exp\left[i \frac{z}{2k} (u_1^2 + u_2^2)\right] g_0. \quad (1.36)$$

На основании (1.36) решим обратную задачу, т. е. найдем  $g_0$  по известному  $g_z$ . Получим

$$g_0 = \exp(-ikz) \exp\left[-i \frac{z}{2k} (u_1^2 + u_2^2)\right] g_z. \quad (1.37)$$

Соотношение (1.37) справедливо и для комплексно сопряженных функций, т. е. на основании (1.37) можно написать, что

$$g_0^* = \exp(ikz) \exp\left[i \frac{z}{2k} (u_1^2 + u_2^2)\right] g_z^*. \quad (1.38)$$

Сравнивая (1.36) с (1.38), видим, что если изготовить голограмму комплексно сопряженного поля, то волна, пройдя через эту голограмму, на расстоянии  $z$  от нее сформирует поле  $p^*(x, y, 0)$ , которое согласно правилу Кирхгофа будет повторять контуры объекта, расположенного в плоскости  $z = 0$ . Это изображение будет сформировано без помощи какой-либо изображающей системы самим свободным пространством.

Все сказанное выше относительно голограмм было очень хорошо и достаточно широко известно более чем за сто лет до официального дня рождения голографии, предложенной Габором в 1949 году.

Почему же год рождения голографии именно 1949-й, что в это время произошло и что именно было изобретено?

Изготовить голограмму по тому фантастическому рецепту, который нами рассматривался выше, т. е. придать некоторой пластинке свойство изменять фазу и амплитуду волны в соответствии с требуемым распределением комплексной амплитуды  $p(x, y, z)$ , до

изобретения Габора можно было только одним способом. Надо было требуемый закон изменения фазы получить путем изменения толщины пластинки, а требуемый закон изменения амплитуды — путем затемнения пластинки. Затемнение еще можно было сделать с помощью фотографии, а изменение фазы сделать трудно по двум причинам. Во-первых, нет подходящего способа для измерения фазового фронта волны, а во-вторых, нет технологической возможности выполнить с требуемой точностью гравировку пластинки. Поэтому вышеприведенные рассуждения о голографии до изобретения Габора были всего-навсего схоластической мечтой, относительно которой можно было со всей определенностью сказать, что она никогда воплощена не будет. И это верно. До сих пор никто не сделал и не пытался сделать голограмму таким прямым способом — гравировкой стекла по соответствующему закону.

Габор предложил способ изготовления голограмм, который целиком основывался на обычной фотографии и никакой гравировки не требовал. Оказывается, можно изготовить не только фотографию, но и голограмму, которая дает возможность при ее освещении получать в точности то распределение комплексных амплитуд, которое заключает в себе информацию об объекте.

Идея, на которой основывается запись голограмм, необычайно проста. В ее основе лежит интерференция волн. Суть этой идеи можно пояснить следующим образом.

Пусть имеется распределение комплексных амплитуд вида  $p(x, y, z)$ , которое необходимо зафиксировать, а затем воспроизвести. Назовем это поле объектным пучком, так как оно несет информацию об объекте. Добавим теперь к этому полю так называемую опорную волну, в качестве которой будет плоская волна, ее волновой вектор составляет некоторый угол  $\theta$  с осью  $z$ . (Для определенности пусть волновой вектор опорной волны лежит в плоскости  $y = 0$ .) В соответствии с (1.3) и в малоугловом приближении выражение для опорной волны будет иметь вид  $A e^{iu_1 x}$ .

Итак, в плоскости  $z$  интерферируют объектный пучок и опорная волна. Сфотографируем получившуюся интерференционную картину с помощью фотопластинки, помещенной в плоскость  $z$ . Пусть фотопластинка реагирует на интенсивность света так, что ее почернение пропорционально этой интенсивности, т. е. квадрату модуля комплексной амплитуды. Получим

$$I = \left[ p(x, y, z) + A \exp(iu_1 x) \right] \left[ p^*(x, y, z) + A \exp(-iu_1 x) \right]. \quad (1.39)$$

Раскроем скобки в (1.39):

$$I = pp^* + A^2 + A p^* \exp(-iu_1 x) + A p \exp(iu_1 x). \quad (1.40)$$

Такое же поле будет, если пластинку, на которой записана интерференционная картина, осветить плоской волной. Поэтому далее мы будем с помощью (1.40) описывать то поле, которое получается при освещении голограммы плоской волной, т. е. при воспроизведении голограммы.

Первоначальная идея записи голограммы, предложенная и осуществленная впервые Габором, отличалась тем, что  $u_1$  было равно нулю (спорная и объектная волна распространялись вместе — схема голографии Габора). Как видно из (1.40), запись интенсивности результата интерференции объектной и опорной волн содержит запись комплексной амплитуды поля, которая дает возможность увидеть объект глазами или сфотографировать (мнимое изображение), и комплексно сопряженный член, который дает возможность получить изображение объекта на экране или фотопластинке без какой-либо изображающей системы, просто за счет распространения волны в свободном пространстве (действительное изображение). Однако схема голографии, предложенная Габором, обладает существенным недостатком (из-за того, что  $u_1 = 0$ ), так как все четыре слагаемых в (1.40) в этой схеме при воспроизведении не разделяются, и поэтому мнимое изображение видно на фоне поля, которое способно формировать действительное изображение, действительное изображение видно на фоне поля, которое способно формировать мнимое изображение, а оба эти поля видны на фоне постоянной составляющей и того поля, которое образует квадрат модуля комплексной амплитуды.

Этот недостаток устраняет схема голографии, предложенная Лейтом и Упатниексом в 1964 году, которая отличалась тем, что в (1.40)  $u_1 \neq 0$ , т. е. объектный и опорный пучки пересекались под углом друг к другу при интерференции. В результате этого появилась возможность разделения сигналов, несущих информацию о действительном и мнимом изображениях, и отделения их от других полей, получающихся при воспроизведении голограммы.

Разделение составляющих в (1.40) основывается на том, что можно сделать пространственные спектры тех слагаемых, которые хотелось бы выделить из общей суммы, не перекрывающимися ни с какими другими. Действительно, умножение на  $\exp(iu_1x)$  означает сдвиг спектра на величину  $u_1$  (гетеродинамирование), а умножение на  $\exp(-iu_1x)$  — сдвиг спектра в другую сторону на ту же величину  $u_1$ . В результате такого гетеродинамирования пространственные спектры сигналов последних двух слагаемых в (1.40) раздвигаются относительно друг друга на величину  $2u_1$ , по отношению



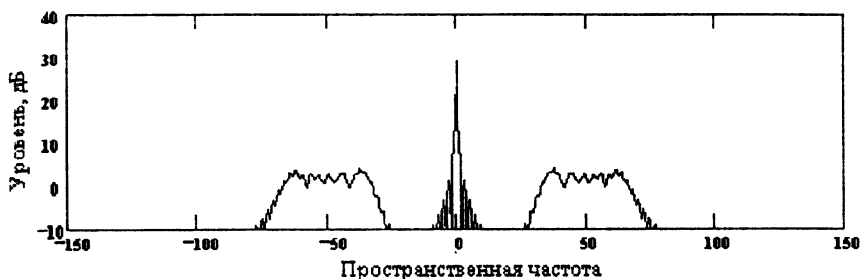


Рис. 1. Спектр функции (1.40). Сигналом  $p$  является поле сферической волны. Программа расчета приведена на рис. 111П

к первым двум слагаемым сдвигаются одно на величину  $u_1$ , а другое слагаемое — на  $u_1$  в другую сторону, как это изображено на рис. 1.

Пространственные спектры объектного и опорного пучков определяются угловой расходимостью этих пучков, которая в схеме голографирования может быть сделана такой, что смещение спектров на  $u_1$  будет достаточным для полного разделения пространственных спектров последних двух членов в (1.40). При этом появляется возможность и пространственного разделения пучков, так как по мере распространения волны все ближе становится выполнение условий для дифракции Фраунгофера (1.33) и превращения полей в их пространственные спектры (такое разделение наступает значительно раньше, чем будут выполнены условия (1.33)).

Таким образом получается принципиальная возможность записи и воспроизведения голограммы, обладающей свойствами той фантастической записи, о которой шла речь выше. В чистом виде может быть воспроизведен сигнал, несущий информацию о мнимом или о действительном изображении. Здесь говорится о принципиальной возможности записи голограммы, так как для того, чтобы практически можно было осуществить саму запись и воспроизведение голограммы рассмотренным способом, требуется соблюдение ряда условий.

Определенные искажения могут возникнуть из-за нелинейности фотоматериала, на котором ведется регистрация голограммы. Однако такого рода искажения не имеют принципиального характера. Голограмма, пусть с искажениями, но все же будет зарегистрирована и может быть воспроизведена.

Есть одно принципиальное условие, соблюдение которого необходимо для возможности регистрации голограммы. Для того чтобы иметь возможность пояснить это важное условие, рассмотрим два последних члена в (1.40). Они в сумме образуют действительный

сигнал, имеющий пространственную несущую частоту  $u_1$ . Нужная информация заключается в модуляции несущей частоты по амплитуде и фазе, а чтобы такой сигнал можно было зарегистрировать, разрешающая способность пластинки должна быть такой, чтобы несущая пространственная частота  $u_1$  могла бы быть на ней зарегистрирована. Так как несущая пространственная частота существенно выше тех частот, с которыми сигнал модулируется, то разрешающая способность пластинки для регистрации голограммы должна быть существенно выше той, при которой может быть достаточно хорошо зарегистрировано изображение самого объекта. Поэтому, если для регистрации объекта, т. е. для его фотографирования, нужна мелкозернистая пластинка, то для регистрации голограммы требуется особо или сверх мелкозернистая пластинка. С этим условием связаны две проблемы. Одна из них состоит в том, что для голографии требуются особые, очень мощные, источники монохроматического света, так как сверхмелкозернистые пластинки имеют низкую светочувствительность. Вторая проблема заключается в том, что все-таки высокие пространственные частоты записываются, как правило, с малым контрастом и поэтому при восстановлении сигналы, несущие полезную информацию, становятся существенно меньше тех, которые этой информации не несут, что затрудняет чистое воспроизведение сигналов. Эта проблема известна как проблема низкой дифракционной эффективности голографических материалов. Рассматривать эти проблемы здесь тоже не будем.

Наличие высокочастотной пространственной несущей частоты позволяет ввести в процесс записи по крайней мере два существенных отличия от записи обычных сигналов:

1. Голограмму можно после записи и проявления полностью отбелить. После того как она превратится в совсем прозрачную структуру, ее не только можно успешно воспроизвести, но и получить более яркое изображение, чем было до процесса отбеливания. Это объясняется тем, что на отбеленной голограмме остаются мельчайшие канавки, глубина которых достаточна для того, чтобы модулировать свет по фазе на большую глубину (глубина канавки для этого должна быть сравнимой с длиной волны). Информация об изображении заключается в изменении расстояния между канавками и их глубины, эта информация сохраняется и на отбеленной голограмме, а ее дифракционная эффективность может быть значительно выше. Такие голограммы называют фазовыми голограммами.

2. На одном кусочке пленки или пластинке можно записать сразу несколько голограмм, которые могут быть воспроизведены

отдельно друг от друга. С целью получения такой записи с многократным использованием пленки достаточно при записи каждой новой голограммы изменять наклон или ориентацию опорной волны. При считывании информации с такой голограммы все световые пучки, относящиеся к разным голограммам, могут быть разделены в пространстве.

Что обещала нам голография и что дала?

Появление голографии сразу же породило длинный ряд надежд, но, к сожалению, многим из них не суждено было сбыться, и они забылись. Но тогда, в начале 60-х годов, все выглядело вполне реальным и казалось вполне научно обоснованным. Почему так быстро нарастали идеи практического использования голографии, настолько быстро, что это было подобно научному взрыву?

До этого в оптике на протяжении более чем ста лет не было никаких крупных, заметных достижений, в то время как другие науки интенсивно развивались, что привело к очень заметным успехам. Появились электричество, радио, телевидение, атомная энергия и атомная бомба, а в оптике как были микроскопы, телескопы да фотоаппараты, так и остались. Как теперь принято говорить, в оптике царил эпоха застоя. Физическая оптика преподавалась только в рамках курсов общей физики, где ее теснили другие, бурно развивавшиеся области знаний, так что не везде физическая оптика преподавалась на должном уровне. Тем не менее специалисты, пусть в небольшом числе, но были, преподавание сохраняло и даже умножало старые вековые традиции, хотя чуда никто не ждал. Чисто фантастические возможности обдумывались как учебные задачи, способствующие лучшему уяснению того, что возможно, а что совсем не возможно в оптике.

И вдруг на фоне застоя появилась голография, и не одна, как раз в то время совершенно независимо от голографии появились лазеры. В принципе голографию можно осуществлять и без помощи лазеров, но при этом ее возможности становятся несравненно меньше. С появлением лазеров в оптике появились регулярные поля. До этого регулярные поля были только в радио для радиоволн. Отличие в технике и методах формирования изображений в оптике и радио было не только в длинах волн, а в том, что в оптике были только случайные поля, а в радио — только регулярные. С изобретением лазеров в оптике появились регулярные поля, а в радио с изобретением высокочувствительных радиометров стало возможным работать и со случайными полями.

Все это и обусловило тот поистине взрывной характер появления новых идей, основанных на использовании голографии в науке,

быту и даже в искусстве. Все должно было претерпеть существенные изменения.

Самой плодотворной и полностью оправдавшей все возлагавшиеся на нее надежды оказалась идея о голографической интерферометрии.

Идея заключалась в следующем. Восстановленное с голограммы поле можно просто наблюдать, можно сложить его с полем, которое получается при освещении того же объекта в момент воспроизведения голограммы. Эти поля могут быть получены с помощью одного источника освещения и могут интерферировать. Таким образом удастся получить интерференцию полей, освещающих данный объект в разное время (во время записи голограммы и при ее воспроизведении). С помощью такой голографической интерферометрии можно замечать и регистрировать очень малые изменения в форме и положении предметов во времени. Без голографии ничего подобного сделать было нельзя.

В настоящее время голографическая интерферометрия имеет многочисленные применения и развита чрезвычайно. Были попытки применить голографию и в быту. К сожалению, кроме музеев голографии пока ничего из этого не вышло.

Были попытки создать голографические экраны для стереокино. Такой экран позволяет смотреть стереокино без каких-либо очков, а съемка такого кинофильма ведется также и для обычного очкового стереокино. Теоретически такая идея проходит, диссертации на эту тему защищены успешно, а соответствующего стереозала до сих пор нет.

Были попытки создания голографического кино и телевидения. Опыты проходили успешно, восторженные статьи корреспондентов появлялись, но пока видимого ничего нет.

Между Москвой и Ленинградом была установлена голографическая телевизионная линия. По ней из Ленинграда в Москву передали голограмму, которая после фотографирования с экрана телевизора в Москве была поездом отправлена снова в Ленинград для восстановления. К удивлению всех участников эксперимента, на восстановленной голограмме у всех изображенных красавиц отсутствовали головы, а все остальное хорошо получилось. Помавшись с восстановлением голов и ничего не добившись, авторы эксперимента поняли, в чем было дело. При записи голограммы использовалась сферическая волна, вследствие чего разные участки объекта получились на разных пространственных несущих частотах. Телевизионная линия сверхвысокой четкости, специально организованная, все же срезала те пространственные частоты, на которые пришлось головы красавиц. Телевизионщики привыкли к

тому, что ограничение полосы пространственных частот ухудшает качество изображения, но доносит его до зрителя всегда целиком. Голографическое телевидение реагирует на ограничение полосы передаваемых пространственных частот совсем иначе.

Трудно сказать, насколько нужен голографический стереоскоп, голографическое кино и телевидение, но была одна идея, которая, на мой взгляд, заслуживает внимания, хотя она тоже не внедрена в наш быт. Речь идет о создании голографического видеопроектирующего устройства. Такой видеопроектировщик может работать по воспроизведению программ, записанных на очень дешевом простом материале. К тому же, сами записи можно легко изготавливать большими тиражами с применением техники для изготовления звуковых записей в виде дисков (пластинок).

Голограмма видеофильма записывается в виде фазового рельефа на лавсановой ленте и может быть воспроизведена с помощью лазера, спроектирована на телевизионную трубку и наблюдаться на экране телевизора как и обычный, не голографический видеофильм. Особенностью голографического видеофильма является еще то, что его при воспроизведении можно протягивать непрерывно, а не покадрово, как это приходится делать при проектировании кинофильма на экран. Записи для такого видеопроектирующего устройства могут быть очень дешевыми, они практически не изнашиваются, допускают сколь угодно продолжительный стоп-кадр. Такие устройства были бы очень эффективны в учебном процессе. Однако дальше образцов, демонстрировавшихся с большим успехом на выставках, дело почему-то не пошло.

На голографию возлагались большие надежды в плане получения изображений очень высокого качества, которое должно было потрясти мир. Ожидалось, что голография позволит существенно превзойти фотографию по качеству получаемых изображений. В пользу такого ожидания было по крайней мере два аргумента. Во-первых, голографическое изображение может быть получено без посредства какой-либо оптической системы, формирующей изображение. Такое изображение должно быть свободно от аберраций, присущих любой оптической системе и неизбежных в ней. Качество изображения должно определяться так называемым дифракционным пределом, т. е. шириной волнового пучка, который формирует изображение. Качество изображения, формируемого любой оптической системой, имеющей достаточно большой угол захвата лучей, намного ниже того, что позволяет дифракционный предел. Во-вторых, голографическое изображение должно быть свободно от ограничений, вызванных зернистостью фотоматериала, так как это

изображение записано на высокой пространственной частоте и на него не влияют даже явные царапины и дефекты на голограмме.

Однако эта самая большая и наиболее аргументированная надежда не оправдалась. Изображение, восстановленное с голограммы, оказалось намного хуже обычного фотографического изображения. В чем дело — вначале было неясно. Габор — изобретатель голографии, которому за это изобретение была присуждена Нобелевская премия, — обратился к научным работникам, которых он знал лично, с просьбой разобраться и устранить этот, казалось бы, неприципиальный недостаток голографического изображения. За это Габор обещал автору спасения голографии добиться и для него Нобелевской премии.

Однако, несмотря на необычайную актуальность задачи, никто не сумел устранить этот недостаток голографии. Работали в течение двух пятилеток, но никто не смог ничего сделать. Результаты проделанной работы изложены в специальном выпуске журнала Американского оптического общества (*Journal of the Optical Society of America*, 1976. V. 66, № 11). В этом журнале, являющемся ведущим мировым журналом по оптике, приведено очень много результатов, которых добились исследователи, борясь за улучшение качества голографических изображений. Сделано много полезного, так как тот враг, который портит изображение, был, наконец, обнаружен и поставлен на службу, однако он своих позиций при ухудшении голограмм не сдал, и стало ясно, что не сдаст никогда. Выявилось новое для оптики принципиальное ограничение разрешающей способности, о которой раньше никто не знал. Этому вопросу, поскольку он практически и принципиально очень важен, посвятим специальный раздел.

## 1.6. Зернистость, или спекл-шум

Приведенные здесь рис. 2 и 3 взяты из только что упоминавшегося американского журнала (*Kozma A., Cristensen Ch. R. Effects of speckle on resolution // JOSA*, 1976. V. 66, № 11. P. 1257). Фотография оптической миры решетки с разным числом штрихов, которые используются оптиками для измерения разрешающей способности и частотных характеристик оптических систем (рис. 2, а), — это обычная фотография миры, выполненная с помощью традиционной оптической системы.

Измерив контраст полученной картины, можно определить, какова частотная характеристика оптической системы на той про-

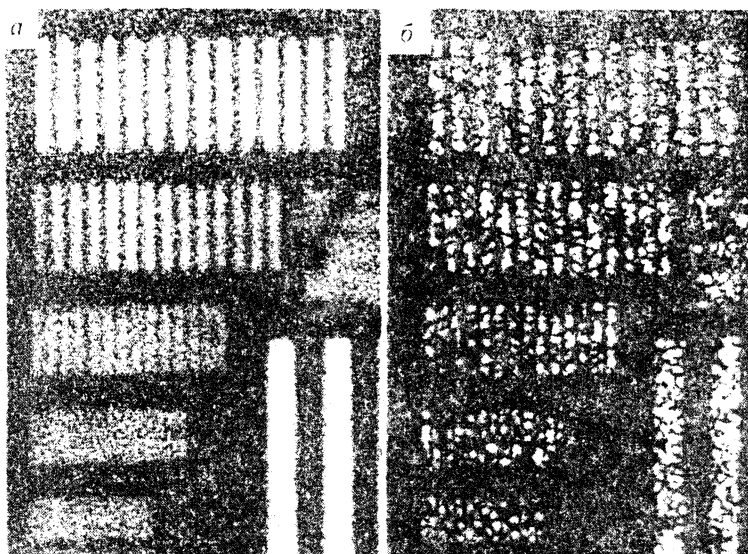
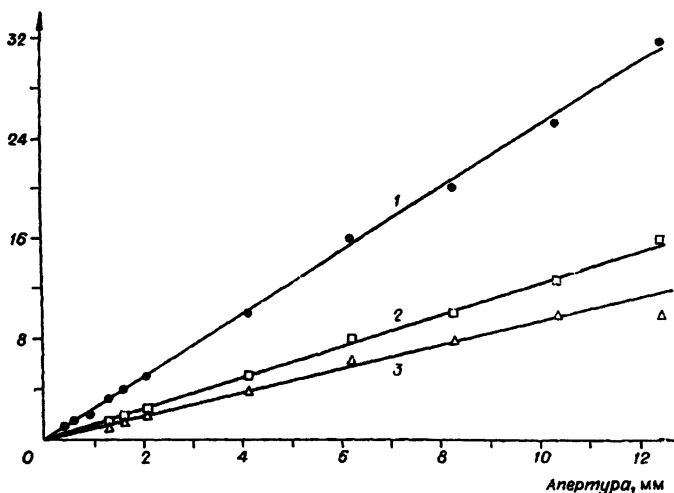


Рис. 2. Оптическая мира в некогерентном (а) и когерентном (б) свете

пространственной частоте, которой соответствует период штрихов на данной мире. (Следует иметь в виду, что на фотографиях приведены сильно увеличенные изображения).

На рис. 2, б — фотография той же мира, но восстановленная с голограммы: отчетливо виден шум, проявляющийся в виде случайно расположенных пятен. Наличие именно этого шума и является тем фактором, который ограничивает разрешающую способность голографии. Насколько существенно такое ограничение, показано на рис. 3, где приведены результаты фотографирования мир с различной пространственной частотой, примеры которых приведены на рис. 2.

Как видно из рис. 3, разрешающая способность обычных оптических систем из-за влияния шумов на голографическую систему существенно выше, чем у голографической. Шум в голографической системе может быть устранен только одним путем — пространственным усреднением полученных изображений. При таком усреднении вместе с шумом сглаживается и контраст на высоких пространственных частотах и снижается разрешающая способность. В обычных оптических системах такой шум, который наблюдается в голографических системах, не возникает, поэтому для исследователей, имеющих до этого дело только с обычными, не



**Рис. 3.** Разрешающая способность как функция апертуры оптической системы для некогерентного изображения (1), когерентного изображения с разрешающей способностью, оцененной по наблюдению за всей картиной полос (2), и когерентного изображения с разрешающей способностью, оцененной по наблюдению за картиной полос через щель, ширина которой равна периоду решетки (3)

голографическими системами формирования изображений, появление такого шума было крайне неожиданно.

Что же это за шум, который мешает голографическим системам проявить свои лучшие качества? В иностранной литературе этот шум принято называть спекл-шумом или просто спеклом. В отечественной литературе в глухие застойные времена была выработана и прочно внедрена традиция обязательно переводить на русский язык все термины. Поэтому в отечественной литературе этот шум принято называть зернистостью. В данном случае с педагогической целью лучше бы сохранить за этим шумом название спекл. Дело в том, что зернистость возникает и при обычном фотопроцессе и тоже ограничивает в целом ряде случаев разрешающую способность оптических систем. Эта зернистость связана с физико-химией фотографического процесса и не имеет никакого отношения к оптике. Физическая природа спекл-шума совсем иная, хотя внешне он выглядит точно так же, как и обычный шум зернистости, характерный для фотографического процесса. Такое внешнее сходство, соединяющееся с общим для обоих совершенно разных видов шумов названием, может приводить к трудностям в обучении.



С целью избежать подобных трудностей отметим сразу же существенное отличие между этими двумя видами шумов. Зернистость при обычном фотопроцессе видна не всегда, но если условия таковы, что она проявляется, то это не зависит от того объекта, который фотографируется в данный момент. Зернистость при голографическом процессе (спекл) наблюдается тоже не всегда, но его появление зависит от того, какой объект "голографируется". Если голограмма регистрируется и восстанавливается с объекта, который не модулирует или слабо модулирует волновой фронт волны (фотография на оптически плоском стекле), то спекл не возникает совсем. Фотографии, изображенные на рис. 2, делались с мир, которые располагались на матовом стекле, обеспечивающем глубокую фазовую модуляцию волны, проходящей через мир с широким набором различных пространственных частот. На рис. 2, а наличие такого матового стекла совсем не сказывается, а на изображении, восстановленном с голограммы, спекл-шум наблюдается только при наличии матового стекла. Спрашивается, зачем нужно ставить специально матовое стекло и получать спекл, с которым приходится либо бороться, либо мириться, не лучше ли — не ставить никакого матового стекла и с самого начала исключить спекл из практики и из рассмотрения? Дело в том, что все обычные предметы, которые являются объектами фотографирования, обладают шероховатой поверхностью, модулирующей световую волну по фазе подобно матовому стеклу. Если поверхность не такая, то она выглядит блестящей и дает блики. Но даже и блестящая поверхность может не быть гладкой в области больших пространственных периодов. Фотографии мир делались на фоне матового стекла с целью имитации поверхности реальных предметов. Тот факт, что на наличие спекла оказывает влияние не характеристика фотопроцесса, не применяемое увеличение изображения, а просто характеристика его поверхности, говорит о том, что у спекла и зернистости совершенно разная физическая природа, разные причины возникновения и имеются наряду с совершенно тождественными способами борьбы и совсем разные.

Следует учесть, что рассмотрение тождественности терминов, обозначающих внешне похожие, но в принципе совершенно разные вещи, имеет важное педагогическое значение, поэтому здесь голографический шум будем называть спеклом, и ответ на вопрос о физической природе спекла попробуем изложить в форме диалога преподавателя (П) и студента (С).

П. Физическую природу спекла Габор во вступительной статье, помещенной в упомянутом выше журнале Американского оптического общества, изложил следующим образом. Спекл возникает по

двум причинам: первая — монохроматическое поле полностью регулярно, и вторая — поверхности объектов являются шероховатыми. Вследствие этих двух причин полностью белая поверхность (например, лист белой бумаги) будет испещрена темными и светлыми точками в результате интерференции волн. Эти интерференционные пятна и представляют собой спекл.

С. Меня такое объяснение не удовлетворяет. В некоторой плоскости, удаленной на расстояние  $z$  от плоскости освещенной белой бумаги, действительно будет наблюдаться интерференция, о которой ведет речь Габор. В произвольной плоскости интенсивность света будет модулирована случайным образом, так как поверхность бумаги обладает сложным рельефом. Эта модуляция будет изменяться при перемещении плоскости, в которой мы ведем наблюдение. Но это явление не имеет ничего общего с изображением того же листа бумаги, восстановленного с голограммы. Вы же сказали, что в условиях малоуглового приближения форма рельефа белого листа бумаги будет полностью соответствовать распределению фаз в волне. Голографическая изображающая система в плоскости, в которой комплексная амплитуда поля будет полностью соответствовать входному распределению поля, которое было на бумаге, должна быть именно такой, как и на самой поверхности бумаги, а на ней — волны, не искаженные никакой интерференцией. Так что объяснение Габора не проходит, да и никакое другое не проходит. Если голографическая изображающая система действительно способна воспроизвести без искажений входное распределение поля, то никакого спекла вообще не должно наблюдаться.

П. Ваши возражения вполне понятны и вы правы. Действительно, дело обстоит не так просто, и объяснение Габора нуждается в уточнении. Оно приведено в том же журнале, но только в другой статье. Если бы голографическая система могла работать без каких-либо искажений и входное распределение комплексных амплитуд было бы в точности равно выходному, которое уже восстановлено с голограммы, то никакого спекла быть не должно. Спекл появляется в результате того, что выходное распределение комплексных амплитуд не может быть в точности таким же, как и входное. Существуют такие искажения, которые появляются обязательно, они неизбежны в любой системе, формирующей изображение. С такими искажениями мы уже встречались. Самым принципиальным из них является так называемый дифракционный предел, который обойти никак нельзя. Тех искажений, которые неизбежны в силу наличия дифракционного предела, достаточно для возникновения спекла. Дифракционный предел состоит в том, что ограничена полоса воспроизводимых пространственных частот. Эта

полоса определяется углом, под которым предметный пучок может попасть на голограмму. Этот угол конечен и не слишком велик, так как иначе различные члены уравнения (1.40) не разделятся при воспроизведении голограммы. Если допустить, что все искажения воспроизведенной голограммы связаны только с ограничением полосы воспроизведенных в ней пространственных частот, а во всем остальном восстановленная голограмма является полной копией оригинала, то уже возникает спекл. Дело в том, что обычная шероховатая поверхность, которая кажется матовой и не содержит явных бликов, очень глубоко модулирует падающую на нее волну по фазе. Эта модуляция является глубокой, т. е. амплитуда этой модуляции много превышает  $2\lambda$  не только интегрально, но и на многих пространственных частотах, так как шероховатая поверхность имеет очень широкий пространственный спектр. Механизм образования спекла в этом случае можно пояснить следующим образом. Возьмем две пространственные частоты, одна из которых настолько высока, что близка к пределу, который может воспроизвести система без заметных искажений; другая частота намного ниже. Пусть и та, и другая пространственные частоты имеют глубокую фазовую модуляцию (амплитудной модуляции нет, так как мы имеем белый лист бумаги, отражающий все излучение без заметных потерь). Известно из радиотехники, что для воспроизведения глубокой фазовой модуляции сигнала необходимо без существенных искажений передать по крайней мере десять первых гармоник модулирующей частоты. В нашем случае модулирующей частотой является та пространственная частота, которую мы хотим передать и воспроизвести с помощью нашей голографической системы. Для того чтобы выполнить условие неискаженной передачи глубокой фазовой модуляции, необходимо, чтобы основная пространственная частота этой модуляции была бы на порядок ниже дифракционного предела. Для нижней рассматриваемой нами частоты это условие выполняется или может быть выполнено, а для той пространственной частоты, которая находится вблизи дифракционного предела, это условие не выполняется. Что будет в этом случае? Какие возникнут искажения? В этом случае, как это хорошо известно в радиотехнике, искажения сводятся к тому, что фазовая модуляция не остается фазовой модуляцией, переходит и в амплитудную модуляцию. Эта амплитудная модуляция сигнала и наблюдается в виде спекла, что объясняет, почему спекл существенно снижает разрешающую способность голографических систем. Фазовая модуляция переходит в амплитудную на пространственных частотах, которые менее чем на порядок отличаются от той частоты (ниже ее), которая определяет дифракционный предел. На

тех пространственных частотах, на которых ниже дифракционного предела уместаются не только основная частота, но и ее десять высших гармоник, спекл не наблюдается. Это соответствует тому, что изображено на рис. 3.

С. Это объяснение для меня не очень понятно и слишком длинно. Нельзя ли объяснить то же самое, но покороче и более понятно?

П. Попробую. Приведенное объяснение содержится в том самом журнале, который целиком посвящен проблемам спекла. Это объяснение должно быть полностью понятно тому, кто хорошо разбирается в радиотехнике. Специалисты в области радиотехники много сделали для создания и развития голографии. Лейт и Упатниекс, о которых я уже упоминал, тоже специалисты в области радиотехники, и они изобрели или, вернее, существенно усовершенствовали оптическую голографию после своих работ по антенной технике. (Об этих работах, о том, как и почему они переросли естественным путем в голографию, речь будет идти дальше в разделе, где рассматриваются антенны с синтезируемой апертурой). Первые свои работы по голографии Лейт и Упатниекс изложили на чистом радиотехническом языке, абсолютно не доступном для оптиков (чистых оптиков). Теперь попробую дать более короткое объяснение для спекла, пользуясь при этом исключительно оптическим языком. Каждая пространственная частота на белой шероховатой бумаге представляет собой фазовую дифракционную решетку с глубокой фазовой модуляцией. Известно, что фазовая дифракционная решетка дает много порядков дифракции, и один порядок дифракции в случае уж очень неглубокой фазовой модуляции, когда дифрагированный сигнал намного меньше невозмущенного поля. Во всех других случаях первый порядок дифракции обязательно дает изменение амплитуды волны. Голографическая система не захватывает высших порядков дифракции и поэтому возникает амплитудная модуляция, т. е. спекл. Для достаточно низких пространственных частот голографическая система в состоянии захватить и высшие порядки дифракции, что приводит к истинной передаче всей сложной структуры сигнала.

Приведенное выше рассуждение поясняется рисунками, которые получены с помощью компьютерных программ, иллюстрирующих формулы, описание их — в разд. 1.8, а сами рисунки вынесены в приложение.

На рис. 4, а изображен вид модуля некоторой сложной, чисто фазовой структуры. Эта структура имитирует экран, изготовленный из ровной белой бумаги, освещенной монохроматическим (лазерным) светом. Наблюдается лишь фазовая модуляция, а амплитудная — отсутствует.

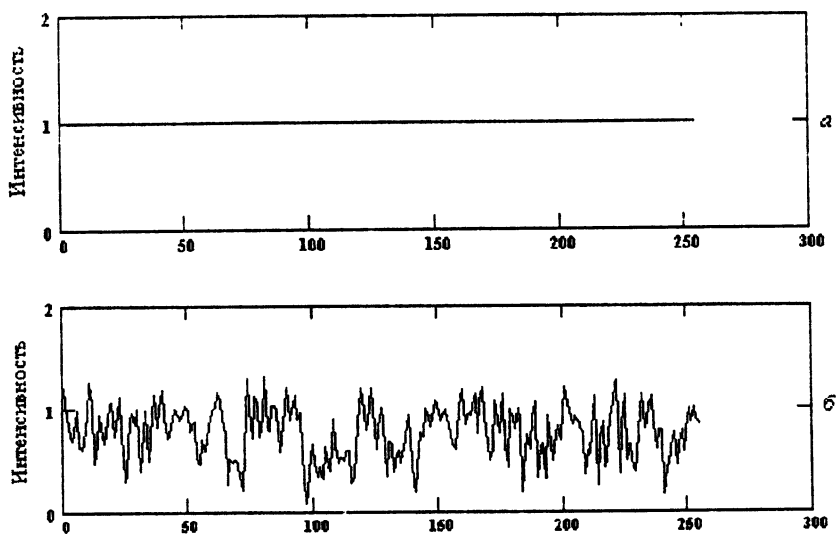


Рис. 4. Интенсивность сигнала, модулированного только по фазе (а) и после прохождения им фильтра, срезающего высокие частоты (б)

литудной нет. Модуль равен единице, или спекла на этой картине не наблюдается. На рис. 4, б изображен модуль той же самой структуры после того, как сигнал прошел через фильтр, который срезал высокочастотную часть спектра. После фильтрации тоже определяется только модуль (см. рис. 4, б). Фильтр имитирует область пропускания свободного пространства, а результат фильтрации показывает, как будет выглядеть волновое поле на некотором расстоянии от экрана.

На следующих рисунках промоделирован описанный выше механизм образования спекл-шума. Фаза некоторой структуры промоделирована двумя частотами в 10 и 70 единиц (рис. 5). На этих же рисунках изображена частотная характеристика свободного пространства в виде прямоугольного фильтра (жирная линия). Низкую частоту и все ее гармоники фильтр пропускает, а более высокую частоту частично срезает.

В результате на рис. 6 виден спекл-шум для высокочастотной модуляции, спектр которой задерживается фильтром, а для низкочастотной модуляции спекла нет, так как все гармоники этой частоты через фильтр проходят свободно.

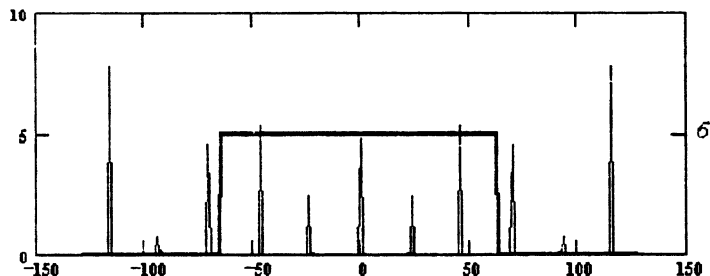
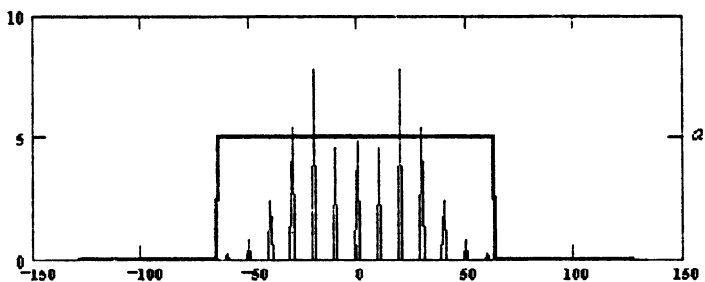


Рис. 5. Спектр сигнала, промодулированного по фазе с частотой, равной 10 (а) и 70 (б). Жирной линией показана полоса пропускания фильтра

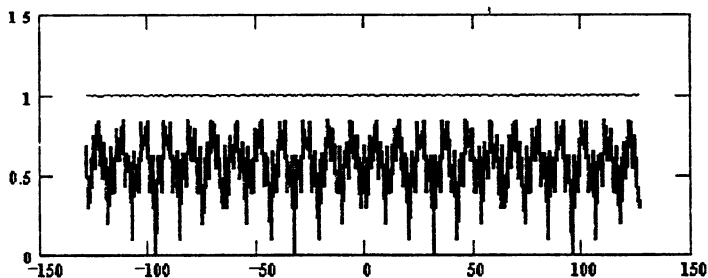


Рис. 6. Интенсивность сигнала, промодулированного по фазе с частотой 10 (верхний) и 70 (нижний) и пропущенного через фильтр, срезающий верхние частоты

Соответствующие компьютерные программы, позволяющие по лучать те же или иные иллюстрации с иным набором параметров, — см. разд. 1.8 и приложение.

С. А можно объяснить еще короче?

П. Можно. Шероховатая поверхность всегда рассеивает падающее на нее излучение на большие углы. Ведь ту же белую бумагу видно под тем углом, под каким мы на нее не смотрим. Голографическая изображающая система имеет ограничение по углам попадающих в нее излучений. Если в систему попали не все рассеянные объектом волны, то система не в состоянии из-за недостатка информации восстановить этот объект без искажений. Среди этих искажений наблюдается и спекл.

С. Аргументы понятны, но есть два серьезных возражения. Они независимы, и поэтому я начну с одного, а потом выскажу и другое, если на первое смогу получить ответ.

Ограничение полосы пространственных частот, воспроизводимых с голограммы, связано с размером самой голограммы. Голограмма большего размера обладает полосой пропускания большей, а голограмма меньшего размера обладает полосой меньшей. Вы утверждаете, что спекл связан именно с ограничением полосы пропускания пространственных частот голографической системой. Представим себе, что мы имеем в своем распоряжении очень большую голограмму, которая во много раз больше, чем зрачок наблюдателя. Вы упомянули, что голограмма может быть такого размера, что через нее можно смотреть одновременно двумя глазами. В этом случае полосу пространственных частот, воспроизводимых с голограммы, ограничивает уже не голографическая система, а глаз наблюдателя или объектив фотоаппарата, если размер голограммы существенно превышает размер объектива фотоаппарата. Таким образом, как вы уже и говорили в свое время, изображение, наблюдаемое глазом или фотографируемое фотоаппаратом, не будет ничем отличаться от того, которое мы видим или фотографируем непосредственно, т. е. без какой-либо изображающей голографической системы. В то же время, когда мы непосредственно смотрим или фотографируем белую бумагу, никакого спекла мы не видим. Почему же на восстановленном с голограммы изображении наблюдается спекл? Может, в данном случае спекла не должно наблюдаться? Тогда почему из вашего объяснения этого не следует и вы ничего об этом не сказали?

П. На эти вопросы очень легко дать ответ. Вы правы только в одном. В рассмотренном случае действительно наблюдаемое и восстановленное изображения не должны отличаться. Они, тем не менее, отличаются. На восстановленном с голограммы изображении спекл есть (он неизбежен в случае изображения любой шероховатой поверхности), а при рассматривании объекта непосредственно спекла не видно. В чем же дело? Почему мы считаем эти

изображения тождественными и куда девается неизбежный спекл при непосредственном рассматривании объекта? Вопрос настолько серьезный, что следует очень внимательно посмотреть на сам объект и окончательно убедиться в том, что при непосредственном рассматривании спекл действительно напрочь отсутствует. Если бы это оказалось так, то мы попали бы в сильное затруднение, потому что никакими разумными доводами этого объяснить нельзя. А значит, прежде чем выносить приговор всей нашей науке, необходимо очень внимательно посмотреть непосредственно на объект, причем при тех же условиях его освещения, при которых с него снимается голограмма. Только тогда получается полное тождество рассматриваемого непосредственно объекта и его голографического изображения. Если мы именно так и поступим, то придем к следующему выводу. Спекла такого же, как на голограмме, при непосредственном рассматривании объекта не видно, но и нельзя утверждать, что объект виден чисто, без всяких посторонних шумов. На объекте при его непосредственном рассматривании видны какие-то пятна, но они постоянно перемещаются по объекту, видны то тут, то там, и поэтому есть возможность не обращать на эти пятна внимания, а видеть в них явно посторонний шум, который не мешает рассматривать объект. Эти бегающие по объекту пятна и представляют собой самый настоящий спекл, который возникает при рассматривании объекта в силу тех же причин, что и при его голографировании. Рассматриваем объект также при ограниченной полосе пространственных частот. Разница состоит только в том, что при рассматривании эти пятна бегают, а при голографировании они остаются неподвижными и нет возможности от них отвлечься. Пятна бегают потому, что освещение является регулярным только в течение небольшого промежутка времени, которого должно быть достаточно, чтобы иметь возможность зафиксировать голограмму. Если фиксировать голограмму в течение большего времени и за это время наблюдаемые на изображении пятна сместятся, то голограмма не получится. В этом случае восстановленное изображение вообще не появится. Поэтому на изображении, восстановленном с голограммы, всегда виден спекл. Когда мы просто рассматриваем объект непосредственно глазом или фотографируем его, то какие-либо ограничения, связанные с регулярностью поля, отсутствуют и изображение получается в любом случае и при любой экспозиции. Кроме того, глаз требует для рассматривания значительного времени и не может удовлетвориться кратковременной экспозицией. Частота света и его пространственный спектр не остаются во времени строго постоянными и поэтому мы рассматриваем объект не в условиях строго регулярного поля, а скорее в условиях,



которые приближаются к условиям случайного поля. Поэтому и появляется разница между тем, что мы ожидаем увидеть (такой же спекл, что и на изображении, восстановленном с голограммы), и тем, что мы фактически видим, рассматривая объект непосредственно глазом (легко можем отстроиться от пятнистой структуры изображения и не замечать ее).

С. Насколько я понял, мы увидим спекл, если будем смотреть на объект непосредственно ровно столько же времени, сколько экспонируется голограмма.

П. Совершенно верно. Если уравнивать длительности экспонирования и рассматривания, то никакой разницы между непосредственным восприятием объекта и изображением, восстановленным с голограммы, не будет. Оба будут оставлять желать лучшего в смысле достигаемой разрешающей способности.

С. Я хочу задать еще и второй вопрос, хотя понимаю, что ответ на него я уже почти получил. Вопрос состоит в следующем: почему в обычной фотографии, которая также связана с дифракционным пределом, мы не наблюдаем спекла?

П. Как уже говорилось, в случае обычной фотографии спеклшум тоже можно наблюдать, но это очень трудно сделать — экспозиция должна быть столь короткой, что поле за это время несколько бы не изменилось, т. е. его можно было бы считать регулярным. На самом деле экспозиция получается настолько больше, что спекл очень хорошо усредняется и его на фотографиях обнаружить невозможно. Хочу добавить, что из этих рассуждений отчетливо видно, насколько все вопросы, рассматриваемые здесь, взаимно связаны. Для того чтобы просто объяснить, почему возникают шумы в системе, работающей в регулярных полях, нам пришлось затронуть и вопросы, касающиеся работы систем в случайных полях и которые будут специально рассматриваться много позже.

С. Вредное влияние спекла на изображение я более или менее понял. В чем же может проявляться положительное влияние спекла, о котором вы упоминали?

П. Действительно, со спеклом связано не только отрицательное, но и нечто положительное. Остановлюсь на одном явлении. Выберем такой промежуток времени, в течение которого поле можно считать регулярным. (Это определяется свойствами и характеристиками применяемого лазера). При таком промежутке времени спекл наблюдается также хорошо и уверенно, как и на изображении, восстановленном с голограммы. Теперь представим себе, что мы начали перемещать матовое стекло, которое стоит за наблюдаемым изображением. Структура спекла (она должна быть непо-

движна, так как наблюдение происходит в достаточно короткое время) при движении матового стекла придет в движение. Для того чтобы вся картина спекла изменилась кардинальным образом, шероховатую поверхность надо сдвинуть совсем немного (на величину порядка периода пространственной частоты, ответственной за появление спекла, т. е. находящейся близко к дифракционному пределу). Изменение структуры спекла можно регистрировать, хотя бы измеряя интенсивность света в какой-либо конкретной точке. Чем быстрее двигается шероховатая поверхность, тем быстрее изменяется структура спекла и тем шире получается спектр интенсивности света в какой-либо точке наблюдения. По ширине спектра можно измерять скорость перемещения шероховатой поверхности. Такой способ весьма чувствителен, точен и удобен для реализации. Представление о возможной чувствительности такого метода измерения скорости можно составить на основе следующих рассуждений. Частота изменения интенсивности света получается согласно эффекту Доплера в следующем виде:

$$\Delta f = f \frac{v}{c} \sin \theta = f \frac{v}{c} \frac{D}{z} = \frac{vD}{\lambda z}. \quad (1.41)$$

Наименьшая регистрируемая частота определяется временем, в течение которого наблюдается спекл. Если это время  $T = 1/\Delta f$ , для  $v$  получается следующая оценка:  $v = (\lambda z)/(DT)$ . При  $\lambda = 5 \cdot 10^{-5}$  см,  $z/D \approx 10$ ,  $T \approx 0,1$  с может быть зарегистрирована скорость перемещения шероховатой поверхности, имеющая порядок  $v = 5 \cdot 10^{-3}$  см/с. Это очень малая скорость, и ее измерение представляет собой интересную прикладную задачу. Таким образом, благодаря спеклу появляется возможность измерения малых скоростей перемещений шероховатых поверхностей. Подобные измерения можно делать не только в оптическом диапазоне, но и в радиодиапазоне. Статьи с довольно подробным анализом имеются в том же журнале Американского оптического общества. При таких измерениях роль света может с таким же успехом играть и ультразвук, что позволяет производить измерения в непрозрачных средах.

## 1.7. Замечания о численном нахождении поля

Прежде чем вести речь о программировании, особенностях программ, их языке и разных иных "хитростях", познакомимся с некоторыми принципиальными преимуществами и ограничен-

ниями численных расчетов, которые тесно связаны с физической стороной проблемы.

В последнее время все большее распространение получают исследования, выполняемые с помощью численных методов. Такие методы особенно эффективны в акустике благодаря тому, что в акустике используются сравнительно низкие частоты и возможны такие многоэлементные антенные решетки, которые позволяют оперировать непосредственно с амплитудой и фазой сигналов, принимаемых отдельными приемниками антенны. Это позволяет широко применять численные методы как для выделения нужных сигналов на фоне помех, так и для определения параметров волнового поля, принимаемого такими решетками.

Кроме того, есть целый ряд актуальных акустических задач, решение которых возможно только благодаря использованию численных методов. Например, расчет волнового поля, рассеянного взволнованной поверхностью моря или неровным дном, возможен, если задана форма рассеивающей поверхности. Задать эту поверхность математической формулой крайне затруднительно, а численное ее задание осуществляется весьма просто, так как для компьютера почти нет никакой разницы, задана ли входная функция в виде формулы или же численно. Таким образом, появилась возможность, оперируя численными методами, в известные теоретические формулы подставлять сигналы в виде взволнованной морской поверхности, неровного дна и даже в виде реальной человеческой речи.

Такие сигналы обрабатываются по известным формулам, но делать это приходится исключительно численными методами, что приводит к ряду особенностей. Компьютер не знает высшей математики, поэтому при численных расчетах трудно переходить к пределу, оперировать бесконечными пределами интегрирования и непрерывными функциями. Это приводит к тому, что известные и эквивалентные друг другу формулы и пути расчетов по ним при численных расчетах выглядят иначе.

Покажем это на примере конкретной задачи о дифракции волн, математическое решение которой было приведено выше в разд. 1.2. Как уже было отмечено выше, исходное волновое поле на плоскости можно задать численно. Дальнейшее вычисление поля, при возможности делать его пером на бумаге, можно было бы провести двумя эквивалентными методами [1]. Один из них, рассмотренный в разд. 1.2 принцип Гюйгенса – Френеля (ПГФ), оперирует с полями, а второй – с их пространственными спектрами.



Рис. 7. Реальная часть частотной характеристики (1.14) и она же (квадратики), полученная преобразованием Фурье (1.15)

Пусть в плоскости  $z = 0$  поле задано численно. Тогда с помощью ПГФ поле в плоскости  $z = \text{const}$  определяется формулой (1.18). Точное решение волновой задачи дается в спектральном виде формулой (1.5). Формула (1.5) не ограничена малоугловым приближением, а справедлива для всех углов, включая и мнимые. При численных расчетах с использованием ПГФ необходимо вычислять выражение (1.14), которое получается как промежуточный результат преобразования Фурье (1.15).

Расчет с использованием формулы (1.5) позволяет исключить одно преобразование Фурье путем подстановки в (1.5) выражения для частотной характеристики в виде (1.14) непосредственно без получения его с помощью фурье-преобразования функции (1.15). Исключение одного преобразования Фурье является существенным плюсом спектрального подхода к решению задачи о распространении волны в свободном пространстве. Существует и еще один плюс спектральной процедуры, на который мы хотим обратить внимание.

Эта особенность иллюстрируется рис. 7, полученным с помощью программы, приведенной и подробно описанной ниже в разд. 1.8. На нем показан разрез реальной части функции  $\zeta(u_1, u_2)$  (1.14) вдоль пространственной частоты  $u_1$  при  $u_2 = 0$ . На этот график, показанный сплошной линией, нанесены в виде квадратиков результаты промежуточного вычисления по классической процедуре ПГФ, теоретически приводящие к той же самой функции (1.14). Из графика видно, что в отдельных точках результат промежуточных вычислений отличается от точной формулы (1.14). Эти отличия, по-видимому, связаны с влиянием краевых эффектов в преобразо-

вании Фурье в особых точках функции (1.14), где действительные волновые числа становятся мнимыми. На задачи, в которых пространственные спектры волн лежат как раз в этом диапазоне, этот краевой эффект преобразования Фурье может оказывать влияние.

Процедура вычисления полей по выражению (1.5), основанная на пространственной фильтрации, вдвойне полезна. Во-первых, она позволяет исключить одно преобразование Фурье при численных расчетах, и, во-вторых, заменяет это преобразование заведомо точной формулой (1.14), исключая возможное влияние математических особенностей преобразования Фурье в некоторых точках. Поэтому спектральная процедура вычисления полей является предпочтительной.

Однако численные методы преподносят сюрпризы. Оказывается, рано делать вывод, что в численных расчетах следует всегда использовать точную спектральную процедуру (1.5). Казалось бы, это логично делать всегда, так как для компьютера не составляет проблемы вычислять, пользуясь точными (пусть и более громоздкими) формулами, самому при этом думать не надо, так как точное решение не подведет.

А вот и не так! Это лишний пример того, что самому думать приходится всегда!

Точное решение задачи, если пользоваться пером и бумагой, работает безупречно. Однако при численном счете дело обстоит не так. Чтобы убедиться в этом, никаких расчетов делать не надо, достаточно взглянуть на формулу (1.14), которая описывает частотную характеристику пространственного фильтра. В этой формуле пространственные частоты представлены дискретным набором значений. Существенно, что шаг дискретизации зависит от используемой апертуры, но абсолютно никак не связан со значением расстояния  $z$ , входящего в эту формулу. Сколь бы ни был мал шаг дискретизации по пространственной частоте, при достаточно большом значении  $z$  этот шаг обязательно окажется больше периода комплексной экспоненты. Начиная с таких (и больших) значений  $z$  частотная характеристика будет вычисляться неправильно. Это явление сбоя компьютера можно описать иначе. Увеличение  $z$  в формуле (1.14) можно трактовать как увеличение индекса фазовой модуляции частотной характеристики, которое расширяет спектр (1.14). При достижении им ширины, при которой половина периода дискретизации станет больше обратной величины ширины спектра (1.14), эта функция не может быть представлена в компьютере с необходимой точностью согласно известной теореме отсчетов [1, 16]. При таких условиях (это обязательно будет при достаточно

большом значении  $z$ ) аптекарски точная спектральная процедура может дать фантастически не те результаты!

На основе высказанных соображений получаются следующие количественные оценки: Неточности только начинают "разъесть" формулу (1.14) на верхних пространственных частотах, когда

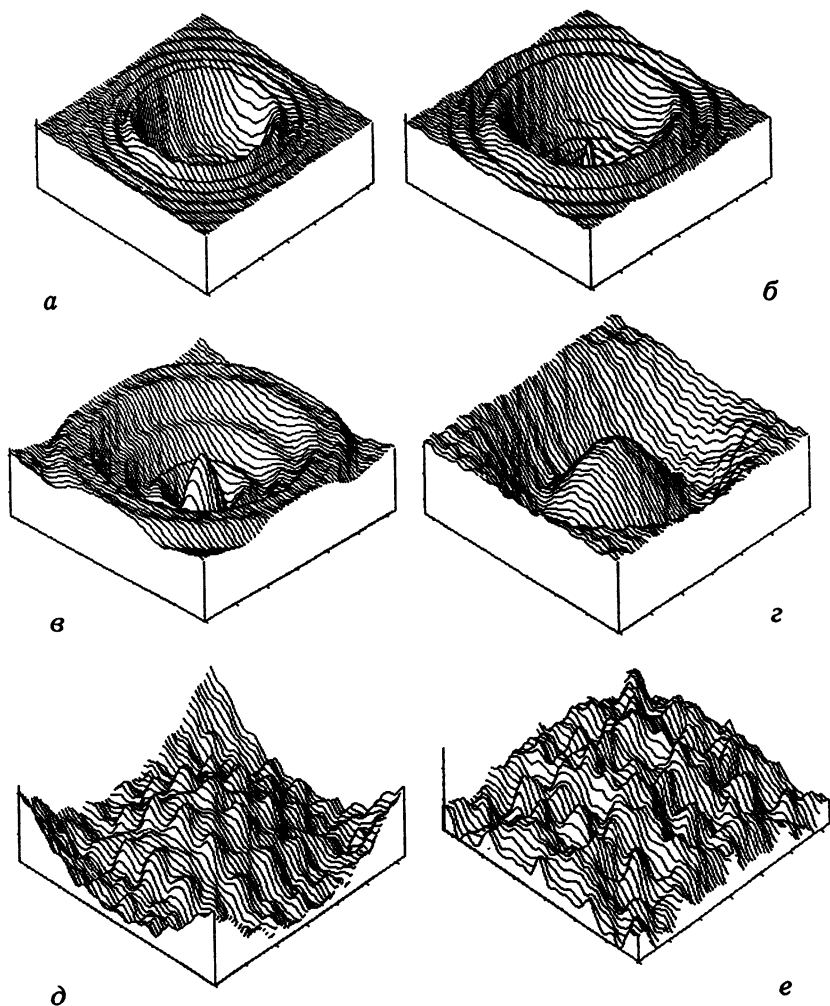
$$z/\lambda = (L/2\lambda)^{1/2}. \quad (1.42)$$

При  $z$ , определяемом (1.42), существует единственный сбой в формуле (1.14). С ростом  $z$  "разъедание" проникает вглубь, и в какой-то момент номер дискретной частоты, до которой все уже "разъедено", определяется следующим выражением:

$$N = (L^2\lambda)/(\lambda^2z). \quad (1.43)$$

Максимальное значение  $N$  составляет  $N_{\max} = 0,5 (LK/\lambda)$ , где  $K$  – число дискретных точек, укладываемых на  $\lambda$ . Формула (1.43) справедлива при условии, что  $N \ll N_{\max}$ . Величина  $K$  для вычисления поля в диапазоне действительных углов, исключая лишь скользкие, должна быть не менее 2. Поведение численного решения дифракционной задачи с использованием точной спектральной процедуры иллюстрирует рис. 8. На нем с помощью приведенной ниже программы численного расчета представлены его результаты. В качестве тестовой задачи рассмотрена дифракция плоской волны на непрозрачном круглом диске. С решения именно этой задачи начинается раздел, посвященный особенностям программирования различных задач. Здесь нас интересует не программа расчета, а только результат ее применения. Эта задача взята в качестве тестовой, так как она решается не только численно. Ее решение хорошо известно. С момента получения экспериментального подтверждения решения именно этой задачи, казавшегося парадоксальным, получила полное признание волновая природа света. Яркой парадоксальной особенностью дифракции плоской волны на диске является появление на расстоянии одной зоны Френеля от диска в центре его геометрической тени светлого пятна, амплитуда поля в котором такая же, как и в отсутствии диска.

При таком значении  $z$ , когда должно быть светлое пятно, (1.42) уже не выполняется, но на рис. 8 явных ошибок при этом не видно. Это еще одна особенность поведения данного решения. Ошибки в (1.14) приводят к ошибкам в расчетах не всегда, а только в том случае, если спектр входного сигнала попадает на ту область (1.14), в которой есть "сбой". В нашем конкретном случае спадающий к высоким частотам спектр диска существенно уменьшает величину ошибок. С ростом  $z$  они становятся все более заметными.



**Рис. 8.** Модуль волнового поля на различных расстояниях от освещенного плоской волной непрозрачного диска. Показан участок со стороной в  $16\lambda$ . Полная апертура – квадрат со стороной в  $64\lambda$ . Радиус диска  $4\lambda$ . Приведены картины распределения амплитуды поля на расстояниях:  $a - z = \lambda$ ;  $b - z = 4\lambda$ ;  $c - z = 16\lambda$ ;  $d - z = 32\lambda$  (зона Френеля для диска);  $e - z = 256\lambda$

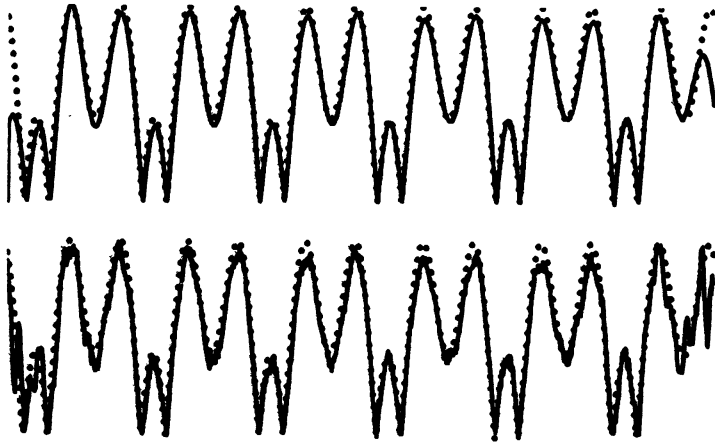


Рис. 9. Модули амплитуд волнового поля на входе и на выходе оптической системы, вычисляемые приближенно (вверху) и точно (внизу)

В случае, когда апертура мала сравнительно с расстоянием  $z$ , хорошо работает приближенная формула (1.18).

Один из парадоксов численного счета заключается в том, что приближенная формула для ПГФ с использованием для  $H(R, z)$  выражения (1.17) с вычислением по формуле (1.18) (без перехода к пространственным спектрам) при некоторых параметрах задачи дает более точные значения, чем точная спектральная процедура. В качестве примера приведем результат решения соответствующей этому случаю тестовой задачи, в которой использовалось одномерное задание полей в соответствии с [1]. Чтобы задача получилась тестовой и было видно, что вычисления соответствуют действительности, была численно смоделирована целая оптическая система, преобразующая изображение некоторого транспаранта в перевернутое изображение того же транспаранта. Для этого численно по формулам и программам, приведенным далее в разд. 4, между двумя участками свободного пространства равной длины был смоделирован объектив, проектирующий изображение транспаранта в выходную плоскость системы. Если вычисления сделаны правильно, то перевернутое изображение входного транспаранта должно совпасть с изображением, вычисленным в выходной плоскости всей системы. На рис. 9 приведены результаты вычисления полей на выходе оптической системы, выполненного двумя методами: с использованием ПГФ в виде (1.18) (без спектральной процедуры) и



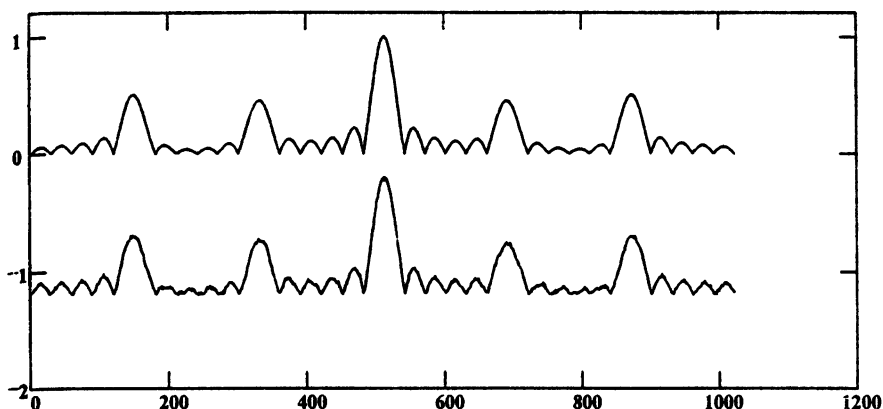


Рис. 10. Результат оптической обработки транспаранта, вычисленный двумя способами: приближенно (вверху) и точно (внизу)

с использованием точной спектральной процедуры по формуле (1.5). Из рисунка видно, что приближенный метод работает лучше и поэтому предпочтительнее точной процедуры.

На рис. 10 показан результат вычисления модуля волнового поля в плоскости расположения объектива. Это результат оптической обработки входного транспаранта [1, 16]. Чтобы сфокусировать в этой плоскости пространственный спектр транспаранта, придав ему вид, показанный на рисунке, он "освещался" сферической волной, сходящейся в центр объектива (см. разд. 4). Здесь расчет по приближенной формуле тоже точнее.

Причину, по которой приближение оказывается точнее, иллюстрируют рис. 11–12. На рис. 11 приведен график действительной части функции (1.17), использованной для приближенных вычислений. Этот график не имеет никаких особенностей, что гарантирует точность вычислений. На рис. 12 показан небольшой низкочастотный участок частотной характеристики (1.14), на котором видны многочисленные "сбои". Место положения первого сбоя хорошо описывается (1.43). Несмотря на то что частотная характеристика в данном случае "изъедена" практически вся, результат вычисления выглядит прилично благодаря тому, что пространственный спектр преобразуемого сигнала достаточно узкий.

Таким образом, применение численных методов заставляет из двух эквивалентных путей решения задачи отдавать предпочтение только одному из них и не всегда заведомо точному. Кроме того, существуют такие сочетания параметров задачи, при которых чрезвычайно трудно (возможно, даже нельзя) выбрать адекватный

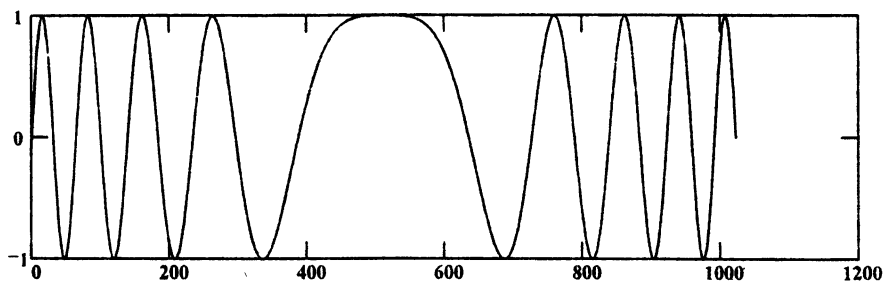


Рис. 11. Реальная часть функции (1.17) с параметрами, использованными при получении рис. 10

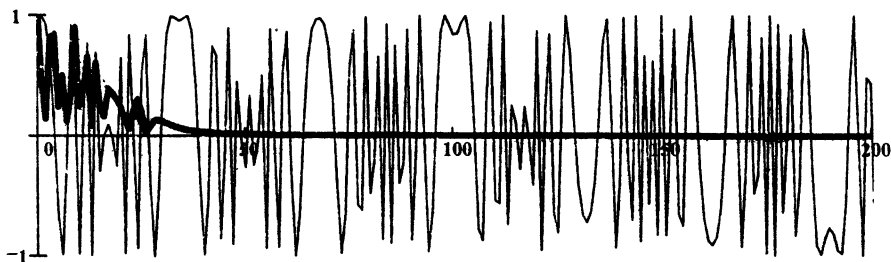


Рис. 12. Начальный участок частотной характеристики (1.14) с параметрами, использованными в рис. 10 (тонкая линия), и модуль спектра входного сигнала (жирная линия)

численный метод расчета. К такому случаю мы придем непременно, если будем, вычисляя поле, удаляться от исходной плоскости, увеличивая расстояние  $z$ . На достаточно малых расстояниях, включая ноль, надежно будет работать формула (1.5) и только она. При расстояниях, намного превышающих апертуру, будет хорошо работать ПГФ (1.18). Для некоторых промежуточных значений  $z$  выбор подходящего алгоритма поля может представить собой такую задачу, над которой придется думать по-настоящему, несмотря на то что все формулы для ее решения давно и хорошо известны.

После такого предисловия перейдем к рассмотрению конкретных программ, позволяющих моделировать формирование изображений волновыми полями. Во всех случаях, где это возможно,

будем стараться использовать точную спектральную процедуру (1.5). С учетом сказанного выше читателю должно быть понятно, что если, пользуясь возможностями, произвольно изменять в приведенных программах параметры, то можно получить "абракадабру" вместо ожидаемого точного решения. В этом случае может быть виновата не программа и не элементарная ошибка, а сам принцип численных расчетов с применением дискретизации. Итак, при вычислении полей мы будем стараться пользоваться спектральным представлением полей с использованием частотной характеристики, определяемой (1.14), и выражением (1.5).

Для большего удобства вычислений преобразуем выражения (1.5) и (1.14) к безразмерным переменным. Для этого введем один важный параметр, который необходим для численных расчетов, — величину того массива данных, на котором будут делаться вычисления. Пусть этот массив состоит из  $L$  элементов, а величина  $L$  равна  $2^N$ , где  $N$  — целое число. Это необходимо для использования на выбранном массиве быстрого преобразования Фурье. Если пространственные частоты условиться относить к  $2\pi/L$ , то эти частоты получим в единицах разрешающей способности по этому параметру. Из-под знака радикала в (1.14) извлечем  $1/\lambda$ . При этом переменная  $z$  будет выражена в длинах волн, а пространственные частоты, стоящие под знаком радикала, умножатся на  $\lambda$ . Примем эти пространственные частоты, умноженные на  $\lambda$ , за новые безразмерные пространственные частоты. В выражения (1.5), (1.7), (1.8) введем новые безразмерные пространственные частоты и получим возможность ввести новые безразмерные пространственные координаты в виде отношений координат  $x$  и  $y$  к  $\lambda$ . Однако, если ограничиться только этим, то мы будем измерять длины в длинах волн и все расстояния окажутся равными целым длинам волн. Поля в ряде случаев должны быть заданы в более широком пространственном спектре, чтобы избежать неоднозначности. Для получения такой возможности следует ввести специальный параметр  $E_\lambda$ , который будет обозначать число точек дискретизации, приходящихся на интервал в  $\lambda$ . Чтобы задавать длины  $x$  и  $y$  в долях длины волны, следует единицы дискретизации, соответствующие этим координатам, делить на  $E_\lambda$ , а пространственные частоты, чтобы при этом все было правильно, следует умножать на  $E_\lambda$ . Кроме того, пространственные частоты, как уже было сказано выше, целесообразно определять в единицах  $2\pi/L$ .

В результате в программе в качестве пространственных частот будем иметь

$$u_k = \frac{k2\pi E_\lambda}{L}. \quad (1.44)$$

В качестве координат получим

$$x_k = \frac{k}{E_\lambda}. \quad (1.45)$$

### 1.8. Примеры математических программ, иллюстрирующих формирование полей при их распространении в свободном пространстве

Проиллюстрируем полученные выше соотношения с помощью численных методов. Причем кроме иллюстраций математических соотношений, дадим средство, позволяющее получать эти результаты не только в качестве иллюстрации. Особенности программирования полученных соотношений удобнее всего иллюстрировать с помощью программы "Mathcad". Особенностью этой программы является наглядная запись математических команд, совпадающая с записью формул, принятых в статьях. Благодаря этому программу, записанную в "Mathcad", легко перевести на любой иной язык программирования. Эта программа позволяет производить все математические операции, используемые в этой книге.

По этой программе имеются подробные руководства [12], что позволяет не вдаваться в детали использования этой программы.

**1.8.1. Знаменитое светлое пятно в центре тени непрозрачного диска.** Прежде всего поясним задачу.

Имеется экран, расположенный в плоскости  $z = 0$ . В центр экрана помещен непрозрачный круглый диск, а вокруг него свободное пространство. Рассматривается дифракция плоской волны, падающей на плоскость  $z = 0$  так, что волновой вектор этой волны перпендикулярен этой плоскости, т. е. плоскости фронтов волны параллельны плоскости  $z = 0$ . Требуется вычислить поле за экраном на расстоянии от диска, равном величине первой зоны Френеля относительно диска. Это расстояние, измеряемое в длинах волн  $\lambda$ , равно квадрату отношения радиуса темного диска к  $\lambda$ .

Решение задачи дается формулой (1.5), вычисления проводятся в два этапа. Сначала спектр входного распределения поля, определяемый по формуле (1.8), умножается на частотную характеристику свободного пространства вида (1.14). Второй этап вычислений

состоит в обратном преобразовании Фурье от полученного произведения. Значение расстояния  $z$  входит в частотную характеристику (ЧХ) свободного пространства в качестве параметра.

Рассмотрим, как эта математическая схема реализуется в вычислительной программе "Mathcad". Программа, по которой персональный компьютер, вооруженный программой "Mathcad 6.0 plus", решает эту задачу, полностью приведена на рис. 1П (рисунки, описывающие программы, вынесены в приложение). Как любая математическая программа, "Mathcad" производит вычисления слева направо и сверху вниз. Поясним программу, изображенную на рис. 1П (в приложении).

На первой строчке вначале  $L$  определяет общий массив данных, в которых производится численный расчет. Последующая программа написана так, что величину  $L$  можно в случае необходимости изменять, сохраняя прежним весь текст программы. Необходимость изменения величины  $L$  возникает в случае нехватки памяти для вычислений ( $L$  в этом случае приходится уменьшать) или в том случае, когда необходимо обеспечить нужную точность дискретизации ( $L$  надо увеличить). В данном случае достаточно величины  $L = 128$ . Далее в этой же строчке приведены так называемые диапазонные переменные [12], принимающие значения в указанном диапазоне с интервалом в единицу.

Строчкой ниже приведены определения пространственных переменных через диапазонные переменные так, чтобы пространственные переменные имели начало координат в центре массива данных. Для того чтобы какая-то функция или переменная пробегала при вычислениях диапазон изменений, в программе введен индекс. Запись  $x_k$  означает, что переменная  $x$  принимает в ходе вычислений все значения, определяемые тем выражением, которое связывает  $x_k$  с диапазонной переменной  $k$ .

Далее приведены численные значения параметров задачи, без введения которых вычисление не начнется. Здесь  $R$  — радиус черного диска, на котором происходит дифракция. Через  $p$  обозначена длина волны. Поле должно быть определено в пределах всех действительных углов и части мнимых углов. Для этого период дискретизации должен быть обязательно меньше длины волны. Так как диапазонные переменные изменяются через единицу, то она и является интервалом дискретизации функций, определяемых этими диапазонными переменными. Чтобы интервал дискретизации был меньше  $\lambda$ , ее величина  $p$  должна составлять несколько единиц. В данном случае  $p = 4$ .

Следующим параметром идет расстояние до плоскости, в которой мы хотим вычислить поле. Это расстояние должно быть равно одной зоне Френеля, что и записано в программе. В дальнейшем значение  $z$  используется в длинах волн.

Далее в программе идет запись поля в исходной плоскости экрана. Это делается в виде программ, имеющихся в "Mathcad 6.0 plus", подробно описанных в [12]. Функции  $P_{k,n}$  представляют собою приближенные по Кирхгофу выражения для поля плоской волны, возмущенной наложенным на нее непрозрачным экраном.

Следующей переменной являются пространственные частоты. Они принимают как положительные, так и отрицательные значения. Диапазон изменений пространственных частот включает значения, приводящие к уменьшению модуля ЧХ.

Следующей в программе стоит ЧХ свободного пространства, записанная на основе (1.14). В этом выражении из-под знака радикала вынесена  $\lambda$ . Параметр  $z$  – расстояние – измеряется уже в длинах волн  $\lambda$ , что уже отражено в сделанной выше записи значения этого параметра  $z$ . Вид пространственных частот  $U_{1,k}$  и  $U_{2,n}$  приведен выше. Он получен на основе выражения для ЧХ. Пространственные частоты, как  $U_{1,k}$ , так и  $U_{2,n}$ , должны пробегать значения от нуля до  $2\pi$ , когда диапазонные переменные  $k$  и  $n$  пробегают  $1/p$  часть значений интервала  $0,5L$ .

Следующая операция преобразует ЧХ, полученную в соответствии с выражением (1.14), к виду, в котором в программе "Mathcad" записываются комплексные спектры функций. Это преобразование заключается в том, что половина пространственных частот, помещаемая в начало изменения диапазонной переменной, должна быть взята из второй половины изменения диапазонной переменной в ЧХ, а вторая половина формируемого спектра берется из первой половины значений диапазонной переменной ЧХ, но в инверсном порядке. Для этого преобразования вводятся специальные диапазонные переменные  $u$  и  $v$ , пробегающие половину значений переменных  $k$  и  $n$ .

Далее следует математическая программа вычисления полей, состоящая из двух этапов. Сначала находятся спектры полей при интересующем нас значении расстояния  $z$ , полученном путем умножения пространственных спектров полей на ЧХ. Для нахождения самих значений полей совершается обратное преобразование Фурье.

Результаты вычислений см. на том же рис. 1П. Приведен вид модуля поля на участке плоскости, содержащей темный экран и вид модуля поля на таком же участке плоскости, удаленной от

исходной ровно на одну зону Френеля. На рисунке отчетливо видно знаменитое светлое пятно в центре тени темного диска, послужившее в свое время убедительным экспериментальным подтверждением справедливости волновой теории света, предложенной Френелем в противовес корпускулярным представлениям Ньютона.

В данном случае это же светлое пятно служит подтверждением того, что мы правильно программируем соотношения волновой теории.

**1.8.2. Дифракция волны на щели.** Эта задача отличается от предыдущей тем, что прослеживается распространение волны в зависимости от расстояния  $z$ . Сделать это, оставаясь в программе предыдущей задачи, сложно, потому что надо вычислять при каждом значении  $z$  поле во всей плоскости. Это большая вычислительная работа, требующая большого объема памяти. Уменьшение объема памяти путем сокращения площади, по которой вычисляется поле, неприемлемо, так как ограничивается расстояние  $z$ , до которого это поле может быть вычислено.

Пример поясняет, как можно производить вычисление полей для больших значений  $z$  при всех промежуточных значениях  $z$ , при этом не увеличивая, а даже сокращая объем вычислений.

За сокращение вычислительной работы приходится чем-то жертвовать. Из приведенного примера следует, что в жертву целесообразно принести число измерений. Как показано в [1], задав поле не зависящим от координаты  $y$ , получим поле, не зависящее от этой координаты во всем пространстве. При этом вычисление поля в зависимости от  $z$  может производиться по формуле (1.5) с исключением из нее одной пространственной частоты. В результате перехода от площади к линии объем вычислений резко сокращается и появляется возможность детально проследивать изменение полей с удалением от исходной плоскости  $z = 0$  и увеличивать величину рассматриваемого участка поля, обеспечивая тем самым возможность проследивать поле до достаточно больших значений  $z$ .

На рис. 2П (см. приложение) приведена программа, по которой производятся вычисления в программе "Mathcad".

В первой строке указан массив данных  $L$ , который в данном примере может быть существенно увеличен без риска выхода за ограничение памяти. Все дальнейшие действия комментариев не требуют, поскольку они подробно пояснялись при рассмотрении рис. 1П.

Остановимся на одном результате, приведенном на рис. 2П. На рисунке показана форма реальной части частотной характеристики; видно, что большое место занимает область запредельных

частот. Величина этой области сравнительно с областью частот, связанных с действительными углами, определяется значением  $p$  – числом точек на длине волны. Увеличивая это число (длину волны), сократим область действительных углов, сжимая тем самым весь график и заставляя его укладываться в отведенную ему часть массива данных. Это становится тем труднее, чем больше осцилляций имеет ЧХ в области действительных углов. Число этих осцилляций увеличивается с ростом расстояния от исходной плоскости  $z$  и обязательно наступит такое положение, при котором осцилляции перестанут воспроизводиться из-за недостаточного числа точек на графике. Это может произойти, если увеличим значение  $p$  или  $z$ .

Картина спектров, приведенная на рис. 2П, в состоянии подсказывает значения величин параметров  $p$  и  $L$ , обеспечивающих точное вычисление полей.

На рисунке приведена картина дифракции волны на щели, соответствующая физическим представлениям об этом явлении.

**1.8.3. Фокусировка.** На рис. 3П приведен пример еще одной программы. Она отличается от предыдущей – дифракции на щели – формой волны, падающей на щель. В этом примере на щель падает сходящаяся цилиндрическая волна. Эта задача решается тоже в цилиндрических координатах с исключением зависимости от координаты  $y$ , как и предыдущая, так как необходимо проследить, как происходит сжатие волны или ее фокусировка в точку ее схождения. В остальном схема расчета и его программа не отличаются от приведенной на рис. 2П. На этом примере еще раз убеждаемся в том, что расчет по приведенным формулам и соответствующая программа приводят нас к тем результатам, которые должны быть. Эта задача может играть роль тестовой при отладке программ.

**1.8.4. Рассеяние волны неровной границей.** Эта задача формулируется следующим образом. На границу двух сред под скользящим углом падает плоская волна. Требуется определить, как эта волна рассеивается, если граница имеет неровности.

Чтобы результат решения и этой задачи можно было сопоставить с элементарными физическими соображениями о том, как должна рассеиваться волна, рассмотрим рассеяние волны на площадке, размеры которой ограничены щелью. Изучим зависимость поля от расстояния  $z$ , а значит, снова рассмотрим одномерный случай.

Задание поля на исходной поверхности имеет в данном случае две особенности, одна из которых состоит в том, что плоская волна падает на границу раздела под предельно скользящим углом. Это



позволяет подойти к математическому моделированию распространения волны в волноводе на низких слабо затухающих модах. Второй особенностью является случайный характер неровностей границы. Для их задания используются случайные числа, генерируемые специальной математической программой (в данном случае использован "Mathcad" [12]). Поверхность, образованная набором случайных чисел, представляет собою сверхмелкозернистое матовое стекло. Такая структура рассеивает падающее на нее поле равномерно во все стороны, независимо от направления падающей волны. В данном случае использован набор случайных чисел, равномерно распределенных в интервале от нуля до единицы. Такой набор чисел имеет среднее значение, близкое к 0.5, и поэтому наряду с рассеянным полем должно быть видно и регулярное поле невозмущенной волны, обязанное среднему значению.

Математическая программа и полученные с ее помощью некоторые результаты приведены на рис. 4П. Программа отличается от приведенной на рис. 3П использованием набора случайных чисел [12] и записью выражения плоской волны, падающей под углом к исходной плоскости  $z = 0$ . Такая волна записана на основе формулы (1.3) (при  $z = 0$  и отсутствии зависимости от  $y$ ).

Остановимся на полученных результатах. Картина спектра поля и реальной части ЧХ дает возможность судить о том, что волна действительно падает под очень скользким углом. Ровно половина спектра этой волны приходится на интервал действительных углов, а другая половина лежит в области запредельных пространственных частот или мнимых углов. На спектре отчетливо видна и случайная часть спектра. Плавные линии спектра обусловлены тем, что поле ограничено достаточно узкой щелью. Случайная часть спектра – сплошная, т. е. простирается на всю частотную ось. Благодаря этому случайная часть спектра простирается через область запредельных частот и заходит во вторую половину ЧХ, формирующую поле обратного направления по отношению к падающему регулярному полю.

На этом примере наглядно видна роль параметра  $p$ . Уменьшение этого параметра до предельно малого значения  $p = 2$  исключит область запредельных частот. В результате волна, падающая под скользким углом, будет распространяться и в обратном направлении, что практически не может иметь место. Такое "математическое" отражение является следствием дискретности координаты  $x$ , которая на практике всегда непрерывна. Чтобы уменьшить влияние дискретизации координат, приходится увеличивать значение параметра  $p$ .

Получившаяся картина рассеяния волны полностью соответствует физическим соображениям. Мы видим поле, рассеянное под всеми углами, в том числе и обратное рассеяние, которое имеет место и на самом деле, так как спектр возмущения через область запредельных частот достигает области, формирующей распространение волн в обратном направлении. Такая картина равномерного рассеяния волн во все стороны наблюдается при достаточно равномерном пространственном спектре неровной поверхности. В том случае, когда этот спектр к высоким частотам убывает, интенсивность обратного рассеяния упадет, так как область соответствующих частот отстоит дальше от области частот прямого рассеяния на интервал запредельных частот.

**1.8.5. Особенности программирования связи импульсной характеристики свободного пространства с его частотной характеристикой.** Особенность программирования указанной связи заключается в необходимости преобразования двумерных матриц к виду, в котором записываются результаты комплексного преобразования Фурье в программе "Mathcad". Программа расчета приведена на рис. 5П. Вначале дается массив данных  $L$  и диапазонные переменные. Далее записана функция  $H(x, y)$ , определяемая выражением (1.15), приведенная к безразмерному виду путем определения расстояния  $z$  в  $\lambda$ .

Далее следует преобразование матрицы  $H(x, y)$  к виду, в котором "Mathcad" записывает двумерные спектры функций с помощью комплексного преобразования Фурье. Эта операция является особенностью производимого преобразования, на которую мы хотим обратить внимание. Далее совершается двумерное преобразование Фурье. В результате этого действия должно получиться выражение (1.14), использованное нами в предыдущих примерах.

На графике полученный результат сравнивается с выражением для ЧХ (1.14). С этой целью строится график реальной части ЧХ (1.14) в виде функции  $u_1$  при  $u_2 = 0$  (сплошная линия). Результат преобразования Фурье в виде функции только  $u_1$  при  $u_2 = 0$  показан точками. Эти точки всюду, за исключением отдельных точек, ложатся на сплошную линию, что убеждает в правильности программирования задачи. Несовпадение отдельных точек можно объяснить тем, что в этих точках функция (1.14) имеет особенности, связанные с переходом от действительных волновых чисел к мнимым, в которых преобразование Фурье может иметь краевые эффекты. Получающиеся отличия очень похожи на них тем, что

приходится ровно на половину истинного значения функции в особой точке, что характерно для краевых эффектов.

**1.8.6. Моделирование голограмм.** Проще и нагляднее начать моделирование голограмм с задачи, известной как "обращение волнового фронта" (ОВФ). Соответствующая программа и результаты приведены на рис. 6П. Особенностью этой программы является задание сферической волны программными средствами. Комплексную амплитуду такой волны можно записать сразу на основе формулы (1.17), однако мы здесь поступили иначе и вычислили это поле. Функция  $p_k = O(k)$  задает сам источник сферической волны. Он определен в виде функции, которая на линии из  $L$  точек равна  $L$  в одной точке, а во всех остальных точках она равна нулю. Далее решается не сферическая, а цилиндрическая задача, в которой поля зависят только от одной координаты  $x$ , а от  $y$  не зависят. В этом случае получим не сферическую волну, а цилиндрическую. Цилиндрическая волна получается с помощью преобразования Фурье функции  $p$ . Умножая получившийся спектр на ЧХ пространства (записано в виде:  $gr_k = g_k z 1_k$ ) и совершая обратное преобразование Фурье, получим исходное поле  $a$  цилиндрической волны. Это поле ограничиваем щелью ширины  $M$  и получаем исходное поле в плоскости  $z = 0$ . Реальная часть полученного таким образом поля приведена на графике (это  $b1_k$ ) с надписью "исходное поле" (см. рис. 7П). Далее предполагается, что это поле распространяется в пространстве. Здесь вычислено поле, образующееся на расстоянии  $r$ , равном тому расстоянию, на котором сформировано поле от источника. Оба поля формируются по соотношению (1.5). Полученный результат показан на графике с надписью "без ОВФ" (это  $br1_k$ ). Далее совершается операция обращения волнового фронта. Эта операция заключается в том, что исходное поле берется в виде комплексно-сопряженной функции (это  $gr2_k$ ). С этим полем совершаются те же операции, что и с исходным полем, т. е. оно вычисляется на том же расстоянии  $r$  от исходной плоскости. Однако вследствие комплексного сопряжения, эквивалентного изменению знака расстояния  $r$ , получаемая при обращении волнового поля функция должна быть такой, какой она была в источнике. В данном случае она не получается в точности такой, как в источнике, из-за влияния щели шириной  $M$ , которая никаким обращением фронта убрана быть не может. Полученный результат показан на рис. 7П с надписью "результат ОВФ" (это функция  $B_k$ ). Все три кривые показаны на одном общем графике и не перекрываются, так как имеют добавочное смещение вдоль оси ординат.

На рис. 7П показана программа, позволяющая продемонстрировать особенности голографии Габора. Исходное поле получено, как и на предыдущем рисунке, оно записано в виде функции  $A_k$ , представляющей голограмму Габора. Вид этой голограммы показан на графике. Поле с этой голограммы восстанавливается, как и на предыдущем рисунке. Предполагается, что при освещении голограммы  $A_k$  плоской волной поле будет совпадать с видом функции  $A_k$ , поэтому она и взята в качестве исходного поля. Результат восстановления приведен на графике. Он представляет собою сумму сигналов, показанных на рис. 7П по отдельности. В голографии Габора эти сигналы неразделимы. Другой особенностью голограммы Габора является необходимость придания ей малого контраста. В данной программе это достигается с помощью множителя  $10^{-4}$ , на который умножается поле перед его суммированием с единицей.

На рис. 8П приведена программа, показывающая особенности голографии Лейта. Для записи этой голограммы формируется специальная плоская опорная волна в виде функции  $A2_k$ . Угол, под которым эта опорная волна падает на исходную плоскость, определяется числом  $f$ . Голограмма имеет вид функции  $A_k$ , по сравнению с голограммой Габора ей придан больший контраст. Если эту голограмму восстановить плоской волной, как и голограмму Габора, то получится угловой спектр, показанный на рис. 1. Особенностью этого спектра являются протяженные пустые промежутки, не содержащие сигнала. Это для математического моделирования практически не годится, так в пустые промежутки уходит слишком много точек. В этой программе все пустые промежутки устранены посредством выделения только той области спектра, которая занята интересующим нас сигналом. Физическим аналогом использованного в этой программе математического приема является считывание опорной голограммы опорной волной или комплексно-сопряженной опорной волной. Вид голограммы и результаты ее считывания как опорной волной, так и комплексно-сопряженной опорной волной приведены на рис. 8П. В голографии Лейта разделяются оба сигнала, показанные на рис. 6П при моделировании обращения волнового фронта.

**1.8.7. Моделирование спекл-шума.** На рис. 9П приведена программа математического моделирования спекл-шума. Функция  $U$  является встроенной функцией "Mathcad", позволяющей получать массив  $L$  чисел, обладающих нормальным распределением с заданными значениями среднего и дисперсии. Эта функция стоит в показателе экспоненты  $W$  в виде чисто мнимого числа с некоторым

числовым множителем. Это выражение имитирует модуляцию волны, падающей нормально на неровную поверхность и идеально отражаемой ею. Модуль такой функции или интенсивность волны будет единица, что видно на левой картинке, где изображен модуль функции  $W$ . Так обстоит дело при рассеянии света листком белой шероховатой бумаги. На правой картинке изображен тоже модуль той же функции  $W$ , но после того, как эта функция прошла фильтр, задерживающий высокие частоты. Фильтр формируется функцией  $F(k)$ . Получающийся на правой картинке спекл-шум происходит вследствие частотной фильтрации широкополосного сигнала, модулированного только лишь по фазе.

Физический механизм этой модуляции и механизм снижения разрешения за счет появления спекл-шума показаны на рис. 10П.

Математическая программа целиком повторяет действия, показанные на рис. 9П, но с другими входными функциями. Здесь в качестве сигналов, осуществляющих фазовую модуляцию, использованы монохроматические сигналы. Соответствующие функции  $W1_k$  и  $W2_k$  модулированы по фазе косинусоидами вида  $U1_k$  и  $U2_k$ , которые отличаются значениями частот в 7 раз. Эти сигналы параллельно обработаны по тем же формулам и программам, что и сигнал, показанный на рис. 9П. Приведены тоже две картинки — левая без фильтрации, правая с фильтрацией. Спекл-шум наблюдается только для функции  $W2_k$ , модулируемой более высокочастотным сигналом.

Чтобы понять, что при этом происходит и почему при модуляции низкой частотой спекла нет, посмотрим на две нижние картинки спектров (см. рис. 10П), на которых изображены модули спектров функций  $W1_k$  и  $W2_k$ , а также частотная характеристика фильтра  $F(k)$ . Видно, что весь спектр функции  $W1_k$  целиком лежит в области пропускания фильтра  $F(k)$  и поэтому спекла в этом случае нет, так как фильтрации не происходит. Спектр  $W2_k$  выходит за пределы области пропускания фильтра  $F(k)$  и поэтому наблюдается спекл точно так же, как и на рис. 9П.

На основе показанных спектров можно пояснить механизм снижения разрешения в голографии. Для того чтобы не было искажений, вызванных спекл-шумом, надо, чтобы внутри фильтра целиком укладывались все гармоники основной частоты модуляции (что мы и наблюдаем для функции  $W1_k$ ). Значит, основная пространственная частота должна быть достаточно низкой, что и воспринимается в виде потери пространственного разрешения.

## 2. АНТЕННЫ

В первом разделе мы рассмотрели общие принципы формирования изображений в свободном пространстве. Системы, формирующие изображения, при этом были только упомянуты, теперь же мы перейдем к рассмотрению особенностей самих систем. Классической системой, формирующей изображение, является оптическая система. Однако существенно более поздняя система, а именно антенна, оказалась намного проще для рассмотрения. Поэтому начнем с рассмотрения антенн, а не с оптических систем.

В отличие от оптических систем антенна может быть не только рассмотрена математически, но и реализована практически в виде некоторого вычислительного устройства. Оптические системы пока существуют только в аналоговом виде и не могут быть заменены вычислительными устройствами. Рассмотрение систем, формирующих изображение, в виде некоторых вычислительных устройств для наших целей привлекательно, так как это позволяет не вдаваться в технические подробности устройства, непосредственно формирующего изображение, а ограничиться рассмотрением физического алгоритма его работы. Все принципиальные особенности работы устройства можно иллюстрировать на этом примере.

Введем основные параметры антенн и выясним, как они влияют на формируемое антенной изображение.

### 2.1. Параметры антенны как системы, формирующей изображение

Начнем вводить параметры с так называемой *плоскости раскрыва антенны*. Плоскостью раскрыва называется такая плоскость, в которой совершаются все необходимые математические операции, или такая плоскость, к которой принято их относить, если на самом деле они совершаются в другом месте.

Для того чтобы задать антенну, необходимо в плоскости раскрыва задать комплексные множители, которые обозначим как  $A(x, y, z)$ . Эти множители записаны в виде функции от трех координат в том случае, если система координат никак не связана с положением плоскости раскрыва антенны. Если систему координат связать с плоскостью раскрыва антенны и сделать так, чтобы

эта плоскость совпадала бы с одной из координатных плоскостей, то множители  $A(x, y, z)$  будут зависеть только от двух переменных  $A(x, y)$ . Воспользуемся таким представлением этих множителей и введем следующий параметр – пространственный спектр этих множителей посредством следующих соотношений:

$$\Gamma(u_1, u_2) = \iint_{-\infty}^{\infty} A(x, y) \exp(-iu_1x - iu_2y) dx dy;$$

$$A(x, y) = \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \Gamma(u_1, u_2) \exp(iu_1x + iu_2y) du_1 du_2. \quad (2.1)$$

Сами множители  $A(x, y)$  специального названия не имеют, однако их пространственный спектр  $\Gamma(u_1, u_2)$  называется *диаграммой направленности антенны* и является определяющей характеристикой антенны наряду с множителями  $A(x, y)$ .

Антенна устроена так, что совершает следующие операции с полями:

- умножение значений комплексной амплитуды поля на множители  $A(x, y)$  в плоскости раскрыва антенны;
- интегрирование результата такого произведения по всей плоскости раскрыва антенны.

Получившееся в результате этих операций комплексное число назовем выходом антенны. Связь выхода антенны с полем в плоскости ее раскрыва можно выразить следующим образом:

$$P = \int_s p(x, y, z) A(x, y, x) dS. \quad (2.2)$$

Сделаем ряд уточнений к этому алгоритму работы антенны.

1. Плоскость раскрыва антенны никогда не бывает бесконечной, а имеет четкие границы. Размеры плоскости раскрыва антенны играют очень большую роль в ее работе, так как они определяют качество получаемых с ее помощью изображений. В связи с этим установилась специальная терминология. Линейный размер плоскости раскрыва антенны называется *апертурой* антенны. Обозначим его буквой  $D$ . До этого мы уже фактически имели дело с апертурой, когда вводили такие понятия, как линейный размер голограммы, который тоже обозначался нами как  $D$ . Все эти понятия очень тесно связаны, но где-то надо основное понятие ввести фундаментальным образом. Понятие апертуры удобнее ввести именно в связи с рассмотрением антенн. Апертура может быть разной в различных направлениях, в таком случае принято указы-

вать, к какому именно направлению относится та или иная величина апертуры. Несмотря на конечность апертуры, интегрирование в соотношении (2.2) удобнее выполнять всегда в одних и тех же бесконечных пределах, так как при этом легче переходить к преобразованию Фурье и другим известным интегральным преобразованиям, которые принято определять в бесконечных пределах. Сохранить бесконечные пределы в (2.2) удастся благодаря тому, что множителям  $A(x, y)$ , приходящимся на ту часть плоскости раскрыва, которая находится за пределами апертуры антенны, придаются значения, равные нулю.

2. Множители  $A(x, y)$  в плоскости раскрыва антенны задаются жестко привязанными к этой плоскости. Изменение положения этой плоскости в пространстве происходит без изменения множителей в плоскости раскрыва. Изменение положения плоскости раскрыва антенны в пространстве необходимо для формирования изображений антенной. Эта операция имеет специальное название — *сканирование антенны*. Сканирование может быть угловым (когда плоскость раскрыва поворачивается вокруг неподвижной точки на углы  $\alpha$  и  $\beta$ ), либо пространственным, когда антенна изменяет свое положение в пространстве так, что плоскость ее раскрыва остается параллельной самой себе. Как уже было сказано выше, при сканировании антенны множители  $A(x, y)$  не должны изменяться (в системе координат, связанной с плоскостью раскрыва антенны) или, говоря иначе, ее диаграмма направленности не должна изменяться при сканировании.

Плоскость раскрыва антенны является воображаемой плоскостью для реальных антенн, в ней может не быть никаких реальных устройств. Реальные устройства, физически формирующие множители  $A(x, y)$ , могут размещаться в другом месте и даже не на плоскости. Однако обработка сигнала в антенне, позволяющая сформировать изображение, дает возможность привести все операции к плоскости раскрыва и к простому алгоритму (2.2).

В том клубке проблем, о котором шла речь выше, есть один узелок, связанный с параметрами антенн, который лучше пока не распутывать. Под этим узелком я подразумеваю следующее.

В литературе, посвященной антеннам, заранее до всякого рассмотрения их работы и в процессе такого рассмотрения, имеется определенная и безапелляционная точка зрения на те величины параметров антенн, которые являются приемлемыми или хорошими, и те значения параметров, которые для антенны являются не очень хорошими и даже совсем плохими. Чтобы не распутывать этот узел, т. е. не выяснять сейчас те причины, по которым



принято именно так оценивать параметры антенн, приведем эти оценки пока без комментариев.

Итак, что принято считать хорошим и плохим, оценивая качество антенн по значениям их параметров?

1. Величину апертуры антенны принято сравнивать с длиной волны. Если апертура много больше длины волны, то это хорошо. Апертуру антенны, существенно превышающую длину волны, принято считать большой. Если апертура антенны порядка длины волны и только немного ее превышает, то это антенна с малой величиной апертуры и это считается плохим. Если же апертура антенны меньше длины волны, то это не антенна.

2. Диаграмма направленности антенны должна быть как можно более узкой – это считается хорошим значением параметра. Чем уже ширина диаграммы направленности антенны, тем антенна лучше.

На основании (2.1) можно получить следующую связь между шириной диаграммы направленности антенны и ее апертурой:

$$\Delta u \geq 2\pi/D, \quad (2.3)$$

где  $\Delta u$  означает эффективную ширину диаграммы направленности антенны, определенную по какому-либо интегральному критерию.

Из (2.3) следует, что при данной величине апертуры существует предельно достижимая острота диаграммы направленности антенны. Антенну с большой величиной апертуры осуществить трудно, поэтому, чтобы получить как можно более узкую диаграмму направленности при заданной величине апертуры, конструкторы антенн стараются связать ширину диаграммы направленности и величину апертуры именно посредством предельного случая равенства в неравенстве (2.3). Это настолько принято, что иногда равенством вместо неравенства (2.3) пользуются даже без оговорок. На отказ от равенства в (2.3) конструкторов антенн толкают только чрезвычайные обстоятельства. В чем они могут состоять? Некоторые из них заключаются в стремлении придать диаграмме направленности антенны еще одно очень важное качество.

3. Конечность величины апертуры (обращение множителей  $A$  в нуль за пределами апертуры) приводит к появлению в форме диаграммы направленности особенностей. Эти особенности называются боковыми лепестками диаграммы направленности антенны. Диаграмма направленности антенны примерно имеет вид, который изображен на рис. 13 – видны максимумы, постепенно убывающие с ростом  $u$ . Эти максимумы называются боковыми лепестками антенны. С боковыми лепестками антенны принято бороться всеми доступными методами. Чем быстрее спадают боковые лепестки



Рис. 13. Примерный вид диаграммы направленности антенны

антенны, тем считается лучше. Еще лучше, чтобы боковых лепестков не было вовсе, но это невозможно при конечной апертуре антенны. Выбор множителей  $A$  в пределах апертуры влияет на степень спада боковых лепестков, но эта процедура однозначно влияет и на ширину диаграммы направленности, расширяя ее по сравнению с минимальным значением, определяемым (2.3). Специальная процедура, имеющая целью снизить уровень боковых лепестков с минимальным расширением диаграммы направленности антенны, называется *аподизацией*. С помощью аподизации удастся существенно снизить уровень боковых лепестков диаграммы с незначительным расширением основного лепестка, называемого главным максимумом диаграммы направленности антенны.

Таким образом, антенна будет тем лучше, чем уже ее диаграмма направленности и чем меньше у нее при этом уровень боковых лепестков.

Таковы общие соображения о качестве параметров антенн.

Рассмотрим формирование изображений с помощью антенн, решая конкретные задачи.

Чтобы включить антенну в решение задачи видения, разделим эту задачу на две – прямую и обратную, как это было в разд. 1.3.

Начнем с решения прямой задачи, которая формулируется следующим образом. В плоскости  $z = 0$  имеется заданное распределение комплексной амплитуды, которое в малоугловом приближении просто и однозначно связано с формой того объекта, который освещается волновым полем (плоской волной). Требуется найти выход антенны, расположенной на расстоянии  $z$  и сканируемой как по углу, так и в пространстве. Решив эту задачу, приступим к решению обратной, для которой прямая задача является вспомогательной.

При решении как прямой, так и обратной задач большое значение имеет расстояние, на которое отстоит антенна от той плоскости, где задано поле. Имеется некоторый масштаб, с которым

сравнивается это расстояние. Расстояние считается большим и говорят, что антенна находится в дальней зоне, если выполняется неравенство

$$z \gg D^2/\lambda. \quad (2.4)$$

Расстояние считается малым (ближняя зона антенны), если выполнено обратное неравенство. Напишем его, чтобы было удобно сделать ссылки:

$$z \ll D^2/\lambda. \quad (2.5)$$

Одно и то же расстояние может быть и малым, и большим для разных антенн. Апертура антенны в неравенства (2.4) и (2.5) входит в квадрате, так как это апертура, измеренная в длинах волн  $\lambda$ , и еще раз сама по себе, как некоторый пространственный масштаб. Неравенства (2.4) и (2.5) играют очень большую роль при рассмотрении работы антенн.

Прямую задачу решаем следующим путем. С помощью выражения (1.5) найдем поле, которое получается в месте расположения плоскости раскрыва антенны. Подставив это поле в (2.2), получим искомое решение задачи. Если проделать все эти операции и учесть необходимость малоуглового приближения, получится следующее выражение [1]:

$$\begin{aligned} p(x, y, z, \alpha, \beta) &= e^{ikz}/(4\pi^2) \times \\ &\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u_1, u_2) \Gamma(u_1 - k\alpha; u_2 - k\beta) \exp\left[-i \frac{z}{2k} (u_1^2 + u_2^2)\right] \times \\ &\times \exp(iu_1x + iu_2y) du_1 du_2. \end{aligned} \quad (2.6)$$

Выражение (2.6) замечательно тем, что оно справедливо для всех значений  $z$  независимо от того, относится  $z$  к неравенству (2.4) или к (2.5). Это очень важно. Ниже этим обстоятельством мы воспользуемся.

На основании выражения (2.6) можно изобразить функциональную схему решения прямой задачи, которая представляет собою последовательное соединение двух фильтров или двух линейных систем (рис. 14).

С одним из них мы уже встречались – это свободное пространство, а другой фильтр получается за счет тех процедур, которые совершает с полем антенна. Частотная характеристика этого фильтра представляет собою диаграмму направленности антенны. Скани-

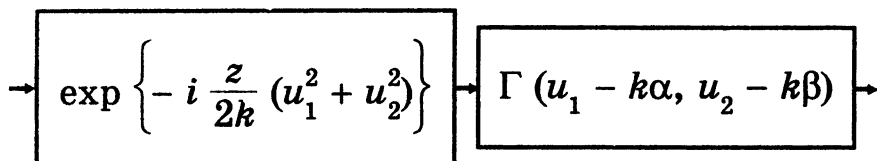


Рис. 14. Функциональная схема решения прямой задачи согласно (2.6)

рование антенны по углу приводит к перестройке этого фильтра на другие пространственные частоты.

## 2.2. Антенна с угловым сканированием

Рассмотрим сначала, как антенна формирует изображение, если она сканируется по углу. Такая система формирования изображения является в некоторых смыслах простейшей системой, и поэтому на примере этой системы легче всего показать, что же мы от нее вообще хотим получить. Это нам облегчит рассмотрение задачи более сложно устроенных или организуемых систем формирования изображений.

Решение прямой задачи в виде (2.6) является слишком сложным для углового сканирования, и оно никогда не применяется. При угловом сканировании всегда употребляется приближение, справедливое при условии (2.4), т. е. в дальней зоне. В этом приближении (2.6) можно представить в виде [1]:

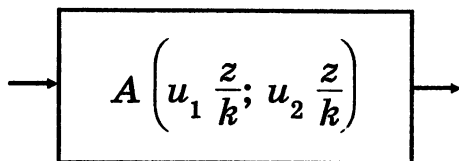
$$P(k\alpha; k\beta) = \frac{e^{ikz}}{i \lambda z} \iint_{-x}^x \tilde{p}(x', y') \Gamma\left(k\alpha - \frac{kx'}{z}; k\beta - \frac{ky'}{z}\right) dx' dy'. \quad (2.7)$$

Через  $\tilde{p}(x, y)$  в (2.7) обозначено распределение комплексных амплитуд на участке сферы радиуса  $z$ , центр которой совпадает с центром, вокруг которого вращается плоскость раскрыва антенны.

Выражение (2.7) имеет простую структуру. Его можно рассматривать как результат фильтрации входного сигнала  $\tilde{p}(x, y)$  фильтром, отклик которого описывается диаграммой направленности антенны при таких значениях аргументов  $u_1$  и  $u_2$ :

$$u_1 = (kx)/z, \quad u_2 = (ky)/z, \quad (2.8)$$

Рис. 15. Функциональная схема решения прямой задачи согласно (2.7)



функциональная схема, эквивалентная уравнению (2.7), изображена на рис. 15. В эту схему при условии того, что  $z$  достаточно велико или апертура антенны достаточно мала, переходит схема, изображенная на рис. 14.

Прежде чем перейти к решению обратной задачи, рассмотрим частный случай решения прямой задачи, который позволит существенно упростить ситуацию.

Пусть множители  $A(x, y)$ , заданные в плоскости разрыва антенны, в пределах апертуры антенны все одинаковы и равны единице. Эти множители согласно (2.7) и рис. 15 определяют частотную характеристику фильтра пространственных частот, через который входной сигнал идет на выход антенны. В том случае, который мы рассматриваем, форма частотной характеристики представляет собою прямоугольник. Все пространственные частоты, удовлетворяющие условию

$$u_1^2 + u_2^2 \leq \frac{k^2}{z^2} \frac{D^2}{4}, \quad (2.9)$$

при движении сигнала со входа на выход остаются без каких-либо изменений. Пространственные частоты, которые выше, чем граница, обозначенная в (2.9), на выход вообще не проходят. Кстати, если полоса пропускания антенны ограничивает полосу пространственных частот, то это не зависит от формы частотной характеристики, а определяется только тем, что апертура антенны конечна. Принятая нами форма частотной характеристики замечательна тем, что в условиях неизбежного ограничения полосы пропускания пространственных частот определяет наименьшие искажения, которым может подвергнуться сигнал, преобразованный антенной. Для того чтобы максимально упростить наши рассуждения, ближе подойти к самому для нас важному и сказать об этом с наибольшей ясностью, поступим следующим образом. Отбросим ту часть входного сигнала, спектр которой лежит выше пространственной частоты, определяемой неравенством (2.9). Все равно эта часть входного сигнала не пройдет через антенну. Под входным сигналом будем подразумевать только ту его часть, которая может пройти

через антенну; с этой оговоркой, весьма существенной, можно утверждать следующее. На выход антенны входной сигнал поступает без всяких искажений, т. е. входной сигнал и выходной сигнал полностью совпадают.

Решение обратной задачи, т. е. задачи нахождения входного сигнала по выходному сигналу, для этого случая не составляет никакого труда. Все, что мы наблюдаем на выходе, и есть то самое, что имеется на входе. Нет никаких искажений (кроме неизбежных, наличие которых заранее оговорено). В этом случае работу антенны можно уподобить работе осциллоскопа, служащего для того, чтобы наблюдать на экране форму сигнала. Такой осциллоскоп, так же как и антенна, обладает конечной полосой пропускания частот, и на его экране не может появиться сигнал, частота которого выше, чем полоса пропускания этого осциллоскопа. Идеальной частотной характеристикой для осциллоскопа является частотная характеристика, имеющая форму прямоугольника. В этом случае мы смотрим на экран этого осциллоскопа и решаем обратную задачу самым простым способом, путем отождествления входного и выходного сигналов. Естественно, таким же образом следует поступать и для антенны.

Столь подробный разбор приведен с той целью, чтобы показать, к чему можно и должно стремиться при решении задачи формирования изображения с помощью антенны. Надо стремиться к тому, чтобы по возможности отпала необходимость в решении обратной задачи как некоторой математической операции и вообще как некоторой операции. Выходной сигнал на выходе антенны должен сразу давать информацию о том, что делается на расстоянии  $z$  от места расположения антенны. Антенна должна дать возможность увидеть то поле, которое находится на расстоянии  $z$  от нее. В рассмотренном нами случае углового сканирования это расстояние должно обязательно удовлетворять условию (2.4).

А теперь очередной парадокс. Все, что было сказано выше об антенне с угловым сканированием, было сказано в педагогических целях и не имеет отношения к использованию антенн с целью видения. Педагогическая задача выполнена, если при этом удалось показать, как может быть построена работа антенны, и то, что она может и должна работать как осциллоскоп. Может, но практически не работает. Как говорится, теоретически лошадь, а практически не везет. Это выражение полностью соответствует той ситуации, которая складывается в задаче видения антенной с угловым сканированием с использованием высказанных выше соображений.

Сначала покажем, что это так, исходя из некоторых общих соображений, а потом уже рассмотрим конкретно, как на самом

деле ставится и решается задача видения с помощью подобной антенны. Посмотрим, насколько помогают характеристики антенны, отмеченные выше как хорошие, работе такой антенны-осциллоскопа, которая была только что рассмотрена.

Возьмем величину апертуры антенны. Чем больше апертура, тем лучше. А насколько помогает увеличение апертуры нашей антенне?

Прежде чем отвечать на такой вопрос, условимся качество работы антенны оценивать по той предельной полосе частот, которую она способна пропустить. Мы уже видели, что в пределах этой области частот никаких особых проблем не возникает, так как в пределах области пропускания антенны сигнал проходит без каких-либо искажений. Все искажения связаны только с ограничением полосы пропускания антенны. Поэтому качество изображения будем оценивать именно по этому параметру.

Но если оценивать работу антенны (а только так и можно) по изображению или по конечному продукту, то получается, что чем больше апертура антенны, тем изображение оказывается... хуже.

Действительно, антенна работает в условиях, когда должны выполняться соотношения (2.4) и (2.9). Решая эти уравнения совместно, получаем, что верхняя частотная граница изображения оказывается пропорциональной  $1/D$ . Отсюда и напрашивается вывод об ухудшении изображения с увеличением апертуры антенны.

Рассмотрим теперь влияние боковых лепестков антенны. Прямоугольная частотная характеристика, обеспечивающая неискаженное изображение в пределах тех частот, которые только и проходят, обладает очень большим уровнем боковых лепестков, но тем не менее это совсем не влияет на качество изображения, которое получается с помощью антенны, если ее рассматривать как осциллоскоп поля. Очень хорошая аподизированная антенна обладает гораздо худшей частотной характеристикой в качестве осциллоскопа поля.

Таким образом, мы пришли к вопиющему противоречию между требованиями к параметрам антенн, которые являются общепринятыми, и теми выводами относительно этих же параметров, которые следуют из решения конкретной задачи. Опять получился парадокс.

Имеется еще одно противоречие между тем, что было изложено выше, и тем, что сложилось в антенной теории. Сформулируем и этот парадокс, чтобы затем разрешить их все сразу.

Дело в том, что в теории антенн принято считать соотношение (2.7) не приближенным, а точным соотношением, но справедливым

только при выполнении условия (2.4) в силу того, что диаграмма направленности антенны, являясь сложной функцией расстояния от антенны до объекта, формируется при условии (2.4). Дальнейшее увеличение  $z$  сверх (2.4) уже не ведет к изменениям формы диаграммы направленности антенны. Соотношение (2.2) в теории антенн не фигурирует. Вся теория построена только на (2.7). При этом форма "несформировавшейся" диаграммы направленности антенны на расстояниях, меньших, чем (2.4), считается очень сложно зависящей от расстояния и никому не интересной. Это тоже нуждается в каком-то разумном объяснении.

Для того чтобы пояснить, в чем тут дело и как разрешить этот парадокс, посмотрим на рис. 16, *а*, где показана картина так называемого радиоизображения – это двумерное изображение карты некоторой местности, снятой сверху с самолета. В одном измерении изображение получалось с помощью антенны, работающей именно по алгоритму, который был нами только что рассмотрен. В другом направлении изображение получалось другим способом, обеспечивающим очень высокое пространственное разрешение. О качестве этого изображения можно судить, сравнив его с изображением, показанным на рис. 16, *б*, – это тот же участок местности, то же радиоизображение, полученное с помощью антенны, установленной на самолете, и длина волны та же. В одном из направлений, в котором было хорошее разрешение, ничего не изменялось – на рисунке *б* по сравнению с рисунком *а* изменен только способ обработки сигнала с антенны. Из сравнения рисунков *а* и *б* видно, что с помощью того алгоритма работы антенны, который мы рассмотрели, получаются изображения неудовлетворительного качества, на них не удастся ничего разобрать. Это объясняется тем, что в силу необходимости выполнения условия (2.4) применяемая антенна не может иметь достаточно большой апертуры, а соотношение (2.7), на котором основывается получение и формирование изображений, справедливо только при этом условии, с какой бы стороны мы к этому вопросу ни подходили – или от соотношения (2.2), или с учетом того, что диаграмма направленности антенны, входящая в (2.7), формируется только на расстояниях, удовлетворяющих условию (2.4).

В этом и состоит ответ на тот парадоксальный вопрос, который был поставлен вначале. Ответ будет следующий: по причине плохого качества изображений, в силу необходимости соблюдения условия (2.4) такой алгоритм работы антенны практически не применяется.

Однако, как и всегда в науке, получив ответ на один вопрос, мы получаем сразу даже не один, а несколько следующих вопросов.



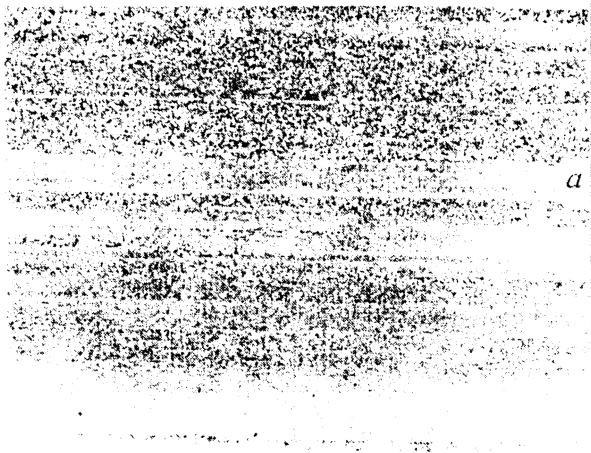


Рис. 16. Изображение карты местности, полученное антенной с угловым сканированием: *а* – в условиях выполнения (2.4) (по оси ординат – дальность – разрешение достигнуто с помощью импульсного режима работы антенны, по оси абсцисс – азимут – с помощью сканирования антенны по углу) и *б* – антенной с пространственным сканированием

Так и здесь, возникают следующие вопросы. Применяются ли вообще где-нибудь и когда-нибудь антенны с угловым сканированием для получения изображений? Или это только теоретическое рассмотрение с педагогическими целями? Как было получено изображение, приведенное на рис. 16, *б*?

Антенны с угловым сканированием для получения изображений в регулярных полях не применяются, однако они применяются для получения изображений в случайных полях. Представим себе, что наш объект наблюдения представляет собою ряд точечных источников, излучение которых некогерентно и не может интерферировать. Каждый из таких источников согласно (2.7) будет давать на выходе антенны отклик в виде ее диаграммы направленности. Так как разные источники не интерферируют между собой, то все такие отклики будут складываться по интенсивности. Получающаяся картина изображений будет представлять собой сумму откликов. На основе такого представления можно сделать следующие выводы. Отклик на данный источник будет тем *уже*, т. е. его координаты будут определены тем точнее, чем *уже* ширина главного лепестка диаграммы направленности антенны. Отсюда это требование к параметрам антенны становится понятным. Далее, следующие отклики будут наблюдаться на фоне боковых лепестков от этого отклика. Ясно, что чем быстрее будут спадать эти боковые лепестки, тем чище и лучше будет картина наблюдаемого изображения. Это требование к параметрам антенны тоже вполне естественно. Возникает вопрос: а почему такие рассуждения не применить бы и к построению изображений в регулярных полях? Это не даст результата, так как регулярные поля интерферируют, причем боковые лепестки могут взаимно не только усиливать, но и гасить друг друга, и подход к изображению как к сумме откликов становится неадекватным. Наш пример показывает большую разницу между подходами к построению изображений в регулярных и случайных полях. К этому вопросу мы вернемся во второй части книги; здесь его обойти не удалось, так как действительно в этом научном направлении все тесно переплетено.

Теперь рассмотрим, как же получить хорошее изображение в регулярном поле.

### 2.3. Фокусируемая антенна

Из предыдущего ясно, что для получения изображений хорошего качества следует отказаться от условия (2.4), ограничивающего размер апертуры антенны. Попробуем это сделать на основе решения прямой задачи более общего вида (2.2). Применив эту формулу, получим диаграмму направленности антенны, полностью сформированной на всех дальностях, и именно той, которая входит в выражение (2.7).

Поставим задачу: найти такой алгоритм решения прямой задачи, который бы позволил реализовать режим получения изображения как в осциллокопе, т. е. чтобы сигнал на выходе антенны был столь же близок к входному сигналу. Будем искать подходящую для этого форму диаграммы направленности антенны, которой можно распоряжаться. Взглянем еще раз на рис. 14, где изображена функциональная схема работы антенны в соответствии с соотношением (2.2). Здесь сигнал проходит последовательно два фильтра. Частотную характеристику одного из них, а именно того, у которого она представляет собой диаграмму направленности антенны, можно изменять как угодно, задавая разные формы диаграммы направленности антенны. Допустим, что можно распоряжаться формой диаграммы направленности антенны как только хочется. Цель состоит в том, чтобы сигнал прошел по возможности без искажений прямо со входа на выход антенны. Мы уже видели, насколько это хорошо для решения обратных задач. Существует такая форма диаграммы направленности антенны, которая полностью решает поставленную нами задачу. Для этого достаточно, чтобы форма диаграммы направленности антенны была такой, при которой получается фильтр, обратный по отношению к частотной характеристике свободного пространства. В малоугловом приближении частотная характеристика свободного пространства определяется соотношением (1.25). Для того чтобы "нейтрализовать" полностью влияние этого фильтра, форма диаграммы направленности должна иметь вид комплексно-сопряженной функции по отношению к (1.25):

$$\Gamma(u_1, u_2) = \Gamma(0, 0) \exp \left[ i \frac{k}{2z} (u_1^2 + u_2^2) \right]. \quad (2.10)$$

Такой вид диаграммы направленности антенны обеспечивает прохождение входного сигнала свободного пространства до выхода антенны без каких-либо изменений, т. е. поставленная нами задача решается наилучшим образом.

Заметим, что получающаяся идеальная диаграмма направленности не соответствует тем требованиям, которые были сформулированы вначале, потому что те требования исходили из решения задач, не связанных с регулярным полем. Мы же решаем задачу в регулярном поле, и полученные решения и алгоритмы могут совсем не подойти для нерегулярных полей. Диаграмма направленности вида (2.10) не только не острая, а никакой направленностью вообще не обладает.

А можно ли реализовать на практике диаграмму направленности вида (2.10)? Для ответа на этот вопрос найдем те множители А,

на которые надо умножить поле в плоскости раскрыва антенны. На основании (2.1) получим

$$A(x, y) = A(0, 0) \exp \left[ -i \frac{z}{2k} (x^2 + y^2) \right]. \quad (2.11)$$

Отсюда можно усмотреть только одну принципиальную трудность в создании такой антенны. Из (2.11) не следует, что ее апертура чем-то ограничена. Диаграмма направленности будет иметь вид (2.10) только в том случае, если поле в плоскости раскрыва антенны будет умножаться на множители (2.11) в пределах бесконечно большой апертуры. Это, разумеется, реализовать нельзя.

Что же можно реализовать? Можно произвести умножение на множители вида (2.11) не на бесконечной, а на достаточно большой апертуре. Предположим, что эти множители имеют следующий вид:

$$A(x, y) = q(x, y) \exp \left[ -i \frac{k}{2z} (x^2 + y^2) \right]. \quad (2.12)$$

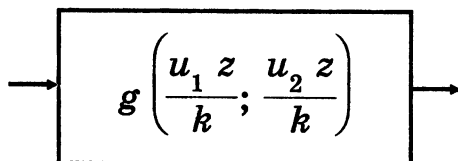
Конечная величина апертуры антенны в (2.12) обеспечивается тем, что функция  $q(x, y)$  обращается в нуль вне апертуры антенны. В пределах апертуры функция  $q(x, y)$  близка к единице. Функция  $q(x, y)$  позволяет учесть не только конечную величину апертуры антенны, но и те ошибки, которые возможны при практическом задании множителей вида (2.11) в пределах апертуры, так как вид этой функции никакими условиями не ограничивается.

Посмотрим, какое влияние окажет введение множителя  $q(x, y)$ , обеспечивающего конечность апертуры и реализуемость антенны, на форму ее диаграммы направленности. Еще раз воспользуемся соотношениями (2.1) и получим следующий результат:

$$\Gamma(u_1, u_2) = \Gamma(0, 0) q \left( \frac{u_1 z}{k}; \frac{u_2 z}{k} \right) \exp \left[ i \frac{z}{2k} (u_1^2 + u_2^2) \right]. \quad (2.13)$$

Выражение (2.13) в силу математического условия его получения [1] является приближенным и справедливым только в ближней зоне антенны, удовлетворяющей условию (2.5). Это смущать не должно, так как мы стремимся к тому, чтобы апертура антенны была как можно больше, а это способствует выполнению условия (2.5). Таким образом, выражение для диаграммы направленности (2.13) справедливо для конечной, но достаточно большой апертуры, удовлетворяющей условию (2.5).

Рис. 17. Функциональная схема фокусируемой антенны согласно (2.14)



Посмотрим теперь, что нам дает применение антенны, имеющей диаграмму направленности вида (2.13).

Подставив (2.13) в (2.6), получим

$$P(x, y, z) = e^{ikz}/(4\pi^2) \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u_1, u_2) g\left(\frac{u_1 z}{k}; \frac{u_2 z}{k}\right) \exp(iu_1 x + iu_2 y) du_1 du_2. \quad (2.14)$$

Выражение (2.14) допускает следующую интерпретацию. Это выход фильтра, на вход которого подается сигнал вида  $p(x, y, 0)$ , представляющий собой распределение комплексных амплитуд на плоскости  $z = 0$ . Частотная характеристика фильтра определяется функцией  $q(x, y)$ , которая должна быть, по нашему условию, равна единице в пределах апертуры антенны. Легко видеть, что реализуются все те же условия, которые мы приняли, рассматривая пример получения изображения антенной с угловым сканированием, основанный на соотношении (2.7). Функциональная схема, к которой сводится выражение (2.14), изображена на рис. 17.

Сравнивая эту схему со схемой, изображенной на рис. 15, можно убедиться в их идентичности. Вся разница заключается в том, что частотная характеристика фильтра в схеме рис. 15 определяется множителями  $A(x, y)$ , а на схеме рис. 17 частотная характеристика определяется множителями  $q(x, y)$ , которые связаны с множителями  $A(x, y)$  соотношением (2.12). Но эта разница относится только к тому, как осуществить антенну, т. е. как надо выбирать сами множители, а по отношению к конечному результату — к получившемуся изображению — никакой разницы между (2.7) и (2.14) нет.

Есть существенная разница в условиях применимости одной и той же формулы, связывающей искомый входной сигнал и выход антенны. Если формула (2.7) справедлива только при условии, что выполнено соотношение (2.4) и антенна находится в дальней зоне или апертура ее достаточно мала, то соотношение (2.14) справедливо при выполнении условия (2.5), противоположного (2.4) и требующего антенны с большой апертурой.

Таким образом, получен очень важный результат. Оказывается, возможно, сохранив очень хорошее свойство антенны воспроизводить на своем выходе искомый входной сигнал, отказаться при этом от условия (2.4), ограничивающего апертуру применяемой антенны, и заменить это условие прямо противоположным, которое требует как можно большей апертуры, удовлетворяющей условию (2.5).

Обратим теперь внимание не на принципиальную сторону дела, а на отличия при практическом применении такой антенны, которые могут возникать.

Прежде всего заметим, что такая антенна не имеет углового сканирования. Диаграмма направленности антенны вида (2.10) и (2.13) не зависит от угла, и изображение с помощью такой антенны можно получить только путем ее сканирования в пространстве. С помощью углового сканирования нельзя, в принципе, ничего хорошего получить путем увеличения апертуры антенны сверх того, что позволяет условие (2.4). Чтобы воспользоваться тем, что дает увеличение апертуры антенны, необходимо перейти от углового к пространственному сканированию.

Следующей особенностью такой антенны является то, что форма множителей  $A(x, y)$ , определяемая (2.12), зависит от расстояния, на которое антенна удалена от искомого распределения поля. Расстояние  $z$  в явном виде входит в формулу (2.12) и без его конкретизации такую антенну создать невозможно. Поэтому эти антенны получили название фокусируемых антенн, они должны быть заранее настроены (сфокусированы) на определенное расстояние до обозреваемого объекта.

Распределение множителей  $A(x, y)$ , которое требуется соотношением (2.12), может быть создано аналоговым способом, например путем применения сферического зеркала. Волны, отраженные от такого зеркала и наблюдаемые в некоторой точке, благодаря дополнительной разности фаз, которая вносится при отражении от зеркала, складываются с соблюдением соотношения (2.12). Параметр  $z$  определяется при этом как кривизной зеркала, так и выбором точки сложения. Однако такие антенны применяются довольно редко, так как их сложно сканировать. Сканирование надо осуществлять в двух плоскостях да еще с изменением в ряде случаев параметра  $z$ .

Попробуем представить себе, как выглядело бы получение радиоизображений с использованием фокусируемой антенны, осуществляемой аналоговым способом.

Требуется получить изображение местности в виде, представленном на рис. 16. Для этого надо изготовить достаточно большое сферическое зеркало, прицепить его к самолету или вертолету и сканировать его на постоянной высоте по местности. Всю местность следует разделить как бы на строки телевизионного кадра и обязательно пролететь над каждой такой строкой. Расстояние между строками должно составлять тот самый размер, который определяет разрешающую способность антенны. Чтобы в кадре иметь тысячу строк, надо вдоль каждой строки протащить антенну. Эта операция займет много времени, но есть и другая трудность. Антенна должна иметь очень большую апертуру. Диаметр зеркала должен удовлетворять условию (2.5). Все это выполнимо для низко летящего самолета или вертолета, но для спутника не применимо. И не только из-за технических трудностей это нереально и мало применяется — существует гораздо более эффективная и к тому же более простая процедура.

При этой процедуре все строки одного кадра получаются при одном и только при одном пролете вдоль строки. Такое упрощение процедуры совершается за счет использования принципа радиолокации. Полет с антенной производится не над тем участком местности, карту которого надо получить, а сбоку от него. Разделение изображений по строкам осуществляется по времени прихода отраженного сигнала. Кроме экономии времени на облет местности (при таком способе можно одновременно получить изображение в нескольких тысячах строк), достигается еще большая разрешающая способность. Изображения в разных строках разделяются в том случае, когда они отстоят на половину длины импульса радиолокатора в пространстве. Длина импульса может занимать во времени несколько колебаний, чему в пространстве соответствует несколько длин волн.

Такого разрешения (да еще не зависящего от дальности) вполне достаточно для получения хорошего изображения местности. Теперь антенну надо фокусировать только в одном направлении. Однако такую фокусировку с помощью простой аналоговой антенны осуществить нельзя. Пусть в нашем распоряжении имеется антенна с достаточно большой апертурой, которую можно фокусировать. Принципиальная трудность заключается в том, что каждая строка находится на известной, но различной дальности. Пока нет способа, позволяющего перестраивать антенну, изменяя ее фокусировку с такой же скоростью, с какой поступают отраженные импульсы с различных дальностей.

## 2.4. Синтез апертуры

На помощь приходит процедура, которая называется синтезом апертуры. Эта процедура осуществляется в два этапа.

На первом этапе производится запись сигналов комплексной амплитуды с хорошим разрешением между строками и с плохим разрешением вдоль каждой строки, так как на самолете устанавливается маленькая антенна, работающая в пределах выполнимости соотношения (2.4), т. е. в дальней зоне. Получающееся при этом изображение на промежуточном этапе приведено на рис. 16, а.

На втором этапе производится обработка полученного сигнала с целью увеличения размера апертуры по направлению вдоль строк и доведение ее до величины, соответствующей ближней зоне антенны, т. е. до условия (2.5). Эту операцию можно сделать с помощью вычислительного устройства или аналоговым способом. Значения комплексных амплитуд вдоль каждой строки умножаются на множители вида (2.12) с учетом того, что в каждой строке – свое значение  $z$ . После интегрирования полученных на всей выбранной апертуре значений произведений получается одна точка изображения сигнала в каждой строке. Для получения следующих точек необходимо сместить точки записи на величину разрешающей способности, после чего опять производится умножение и интегрирование в пределах выбранной апертуры.

Величина апертуры определяется при обработке сигнала и в принципе может быть очень большой. Отсюда и название этого способа обработки и получения радиоизображений: радиолокатор бокового обзора с синтезированной апертурой. Без синтеза апертуры радиоизображения высокое качество получить трудно.

С помощью синтеза апертуры получают радиоизображения со спутников Земли. Таким же способом было получено радиоизображение поверхности планеты Венера. Этот способ получения радиоизображений послужил мощным толчком для развития оптической голографии.

Вопросы обработки радиоизображений и связь этого метода с голографией рассмотрим в следующем разделе.

## 2.5. Радиоголография

Проблема получения радиоизображений была решена гораздо раньше, чем на сцену вышла оптическая голография. Изобретение Габора, предназначавшееся для совершенствования



изображений объектов при наблюдении их с помощью электронного микроскопа, не нашло распространения среди специалистов, не привело к какому-либо заметному интересу в этой области оптики, да к тому же и лазеров еще тогда не было и никто о них и не мечтал.

Как же была решена задача получения радиоизображений?

К тому времени была развита и добилась больших успехов радиолокация. В связи с задачей селекции движущихся целей в радиолокаторах появилась возможность измерения комплексной амплитуды принятого отраженного от объекта сигнала.

Одним из путей реализации такой возможности является применение так называемого когерентного гетеродина. Когерентный гетеродин представляет собой генератор, обладающий высокостабильной частотой (регулярное колебание в течение времени, за которое распространяется до объекта и назад излучаемый импульс), фаза колебаний которого каждый раз, т. е. при каждом новом излучении импульса, жестко привязывается к моменту излучения импульса. Пусть когерентный гетеродин генерирует сигнал вида

$$F_c(t) = \cos(\omega t), \quad (2.15)$$

который можно записать еще и в следующем виде:

$$F_s(t) = \sin(\omega t). \quad (2.16)$$

Импульс излучается каждый раз в момент начала отсчета времени  $t$  в (2.15), т. е. в момент  $t = 0$ .

Благодаря когерентному гетеродину появляется возможность измерять время распространения импульса туда и обратно в периодах и долях периода основной несущей частоты или измерять расстояние до объекта в длинах волн и долях длины волны. С этой целью принятый отраженный импульс складывается с колебанием когерентного гетеродина, сигнал которого имеет постоянную амплитуду вплоть до следующей посылки импульса. Принятый радиолокатором сигнал от дискретного одиночного отражателя имеет вид

$$S_c(t) = A_c \cos[\omega(t - \tau) - \varphi]. \quad (2.17)$$

В принятом сигнале тоже есть, как говорят, обе квадратуры, т. е. и *sinus*, и *cosinus*:

$$S_s(t) = A_c \sin[\omega(t - \tau) - \varphi]. \quad (2.18)$$

На основе (2.15)–(2.18) можно получить сигнал комплексной амплитуды следующим образом:

$$\text{Re}(P) = F_c S_c + F_s S_s, \quad (2.19)$$

$$\text{Im}(P) = F_s S_c - F_c S_s. \quad (2.20)$$

Формула (2.19) определяет действительную компоненту комплексной амплитуды  $P = A_r \exp(i\omega\tau - i\varphi)$ , а (2.20) – мнимую.

Тракт радиолокатора, позволяющий получать значение комплексной амплитуды, называется фазовым детектором в силу своей чувствительности к фазе отраженного сигнала. Основное назначение этого тракта – не синтез апертуры, а селекция движущихся целей. Если неподвижная цель или местный предмет будут находиться на одном и том же расстоянии от радиолокатора, что значит в приведенных выше формулах постоянство  $\tau$ , то выход фазового детектора в обоих квадратурных каналах будет одинаковым при разных посылках. Если же цель движется так, что ее расстояние до радиолокатора изменяется, то фаза все время будет изменяться от посылки к посылке и сигнал на выходе детектора квадратурных каналов будет переменным во времени. На выходе будет появляться сигнал с частотой Доплера. Это дает возможность различать подвижные и достаточно быстро движущиеся цели от медленно перемещающихся или неподвижных и убирать нежелательные объекты с экрана радиолокатора, оставляя только те, за которыми следует вести наблюдение.

С помощью фазового детектора можно получать значения комплексных амплитуд  $P$  при разных посылках. Следующей задачей является фиксация значений  $P$  для многих значений  $\tau$  (каналов дальности) и многих посылок (из которых в ходе обработки сигнала образуются каналы азимута).

В те времена, когда эта техника создавалась (это было в пятидесятые годы, т. е. около сорока лет тому назад), на пути осуществления задуманного было два серьезнейших препятствия. Первое из них заключалось в том, что трудно с точностью, обеспечивающей необходимую синхронность записи, фиксировать как действительную, так и мнимую части комплексной амплитуды.

Вторая трудность состояла в том, что в то время еще не существовала достаточно производительная вычислительная техника, которая бы могла обработать весь массив информации и непрерывно формировать радиоизображения. Насколько эта трудность была существенна в то время, говорит тот факт, что такой техники пока нет и сейчас, несмотря на то что она с тех пор прошла гигантский путь. Современная вычислительная техника позволяет получать отдельные кадры изображений, но обработать весь массив данных с нужным качеством она еще не в состоянии.

Как же были преодолены обе эти трудности?

Самое удивительное в этом преодолении заключается, на мой взгляд, в том, что вторая трудность была преодолена во многом благодаря тому, что первую преодолеть так и не удалось.

Итак, нет возможности синхронно фиксировать действительную и мнимую части комплексной амплитуды  $P$ . Ограничимся тем, что зафиксируем только одну компоненту  $P$ . Пусть это будет действительная компонента, определяемая (2.19). Оказалось, что эта уступка обстоятельствам открывает неожиданные и совершенно новые возможности для обработки радиосигналов, которых до этого специалисты по технике и теории антенн не знали. Оптики о таких возможностях обработки сигналов знали, но для оптического сигнала пока нет способа осуществить то, что делает фазовый детектор радиолокатора.

Рассмотрим эти возможности.

Отвлечемся от того, что сигнал на самолете принимается и излучается с помощью антенны, имеющей какую-то диаграмму направленности, и максимум этой диаграммы направленности может быть каким-то образом ориентирован в пространстве. Пусть сигналы излучаются и принимаются ненаправленным устройством. Такое устройство может иметь весьма малый размер, и можно считать, что сигнал комплексной амплитуды принимается в точке.

Принятый сигнал фиксируется с фазового детектора радиолокатора так, что мы располагаем только одной действительной составляющей полной комплексной амплитуды сигнала. Это можно представить в виде

$$\bar{P} = 2\text{Re } p(x, 0, z) = p(x, 0, z) + p^*(x, 0, z), \quad (2.21)$$

где действительная компонента комплексной амплитуды дана в виде суммы собственно комплексной амплитуды и ее комплексно-сопряженного члена.

Что получится, если теперь действительный сигнал  $\bar{P}$ , определяемый (2.21), обработать так, как следовало бы обработать запись полной комплексной амплитуды, включающей не только действительную, но и мнимую части. Вспомним, что мы с таким случаем уже встречались в разд. 1.5, а так как "повторение — мать учения", то целесообразно вернуться вновь к уже рассмотренной задаче.

В схеме голографии Габора (разд. 1.5) в уравнении (1.40) среди четырех членов была и сумма вида (2.21). Два члена этой суммы дают действительное и мнимое изображение. К выражению вида (2.14), т. е. к изображению исходного сигнала приводит обработка только одного слагаемого в (2.21). Какова же роль второго? У него может быть две роли. Если применяется алгоритм обработки сигнала с использованием множителей вида (2.12), то роль этого

слагаемого сводится к тому, что оно дает дополнительный шум, который накладывается на получаемое изображение. Здесь дело обстоит точно так же, как и в голографии, осуществляемой по схеме Габора.

Однако у этого слагаемого возможна и другая роль. Это слагаемое способно создавать так называемое действительное изображение. (В разд. 1.5 этот вопрос рассмотрен подробно). Если свет или другое волновое излучение промодулировать с помощью второго слагаемого (2.21), то это волновое излучение, пройдя путь  $z$  в свободном пространстве, сформирует изображение. В нашем случае изображением является выражение вида (2.14).

Подробности практического выполнения этой процедуры, связанные с переходом от радиодиапазона к световым волнам, здесь обсуждать не будем. Важно, что необходимая для синтеза апертуры математическая обработка сигнала может быть осуществлена аналоговым способом за счет распространения волны в свободном пространстве (применение оптических систем не вносит в эту процедуру чего-либо принципиально нового, а просто помогает технически осуществить эту процедуру).

Однако, если осуществить аналоговую обработку сигнала за счет второго члена (2.21), то там есть еще и первый член, который будет создавать помеху.

Таким образом, возникшая трудность в осуществлении синхронной записи действительной и мнимой частей комплексной амплитуды, при попытке обойти ее и записывать только одну действительную компоненту, привела к положительному результату – возможности аналоговой обработки сигнала, а также появлению новой трудности – помехи со стороны первого члена в выражении (2.21). С этой помехой нельзя не считаться, так как ее интенсивность равна интенсивности сигнала.

Оказывается, что можно преодолеть возникшую трудность и устранить влияние первого члена в (2.21) путем использования некоторых специальных приемов антенной техники, заплатив за это некоторым искажением изображения (устранением низких пространственных частот в изображении).

Как это делается?

Учтем теперь, что прием сигнала на борту самолета ведется с помощью антенны, имеющей хотя и малую, но вполне конечную апертуру, которая много больше длины волны. Для дальнейшего целесообразно будет уточнить нашу терминологию, относящуюся к изображению, так как изображение в разных координатах строится с использованием различных физических принципов.

Введем на изображении линии строк (как на телевизионном кадре). Эти строки, располагаясь рядом друг с другом, заполняют некоторый кадр, называемый изображением местности. Линии строк совпадают с направлением полета самолета. Информация с разных строк принимается и фиксируется отдельно благодаря импульсному режиму работы радиолокатора. Весь кадр объемом в несколько тысяч строк записывается одновременно за один пролет самолета вдоль строки. Плоскость раскрыва приемной антенны на самолете ориентирована вдоль линии строк, т. е. вдоль линии полета самолета. Так же ориентирована плоскость раскрыва и синтезируемой антенны, информация для которой накапливается в полете.

Напишем выражение для выхода малой антенны, имеющей апертуру  $D_m$ , с учетом того, что она во время полета сканируется вдоль координаты  $x$  (линии строк), а ее диаграмма направленности может составлять некоторый угол от нормали к плоскости раскрыва антенны. Поворот диаграммы направленности малой антенны полностью эквивалентен введению постоянного множителя вида  $\exp(ikx \sin \alpha)$  в плоскости раскрыва антенны, на который умножается падающее на антенну волновое поле. При малом  $\alpha$  этот множитель можно представить в виде  $\exp(ik\alpha x)$ . (Это малоугловое приближение, необходимость которого мы уже выяснили).

С учетом поворота диаграммы направленности малой антенны на угол распределение множителей  $A(x)$  в плоскости ее раскрыва запишем в виде

$$A(x) = \exp(ik\alpha x) \tilde{A}(x). \quad (2.22)$$

На основании (2.2) и (2.21) выход сканируемой вдоль оси  $x$  малой антенны будет

$$P_d(x, z) = \int_{-\infty}^{\infty} p(x', z) \tilde{A}(x - x') \exp[ik\alpha(x - x')] dx', \quad (2.23)$$

и его можно представить в виде

$$P_d(x, z) = \exp(ik\alpha x) \tilde{P}(x, z), \quad (2.24)$$

где  $\tilde{P}(x, z)$  — результат пространственной фильтрации входной комплексной амплитуды  $p(x, z)$  с помощью антенны:

$$\tilde{P}(x, z) = \int_{-\infty}^{\infty} p(x', z) \exp(-ik\alpha x') \tilde{A}(x - x') dx'. \quad (2.25)$$

Анализ выражения (2.25), описывающего пространственную фильтрацию входного поля антенной, повернутой на угол  $\alpha$ , временно отложим, займемся теми следствиями, которые вытекают из выражения (2.24).

На самолете во время полета фиксируется комплексная амплитуда, определяемая (2.23), а именно действительная часть ее, которая с учетом (2.23) и (2.20) выглядит следующим образом:

$$\frac{1}{2} \operatorname{Re} P_d(x, z) = \exp(ika\alpha) \tilde{p}(x, z) + \exp(-ika\alpha) \tilde{p}^*(x, z). \quad (2.26)$$

Сравним это выражение с (2.21) и с (1.40), по своему виду оно полностью совпадает с двумя последними членами выражения (1.40), определяющего оптическую голограмму. Два последних члена (1.40) определяют действительное и мнимое изображения. В разд. 1.5 уже было показано, что наличие экспоненциальных множителей вида  $\exp(ika\alpha)$ , имеющих разные знаки показателей, способствует разделению действительного и мнимого изображений. Из сравнения (2.26) и (1.40) следует, что запись действительной части выхода малой антенны, повернутой на угол  $\alpha$ , можно обработать точно так же, как и оптическую голограмму с возможностью (при некоторых условиях) разделения действительного и мнимого изображений.

Возможности разделения действительного и мнимого изображений определяются, как и в оптической голографии, двумя факторами – величиной смещения спектров сигналов  $\tilde{p}(x, z)$ , вызванного наличием экспоненциального множителя, и шириной этих спектров.

Для выяснения этих факторов вернемся к анализу выражения (2.25), который был нами временно отложен.

Соотношение (2.25) с учетом (1.10)–(1.12), (2.1) может быть записано в следующем виде:

$$\tilde{p}(x, z) = \int_{-\infty}^{\infty} g_z(u - ka) \tilde{\Gamma}(u) \exp(ixu) du. \quad (2.27)$$

Действительно, множитель  $\exp(-ika\alpha)$ , на который умножается входное поле, приводит к сдвигу спектра этого поля (гетероденированию), а все выражение (2.25) с учетом роли экспоненциального множителя представляет собой выход линейного фильтра с откликом  $A(x)$  и частотной характеристикой  $\tilde{\Gamma}(u)$ , представляющей собой диаграмму направленности антенны без учета ее поворота на угол  $\alpha$ .

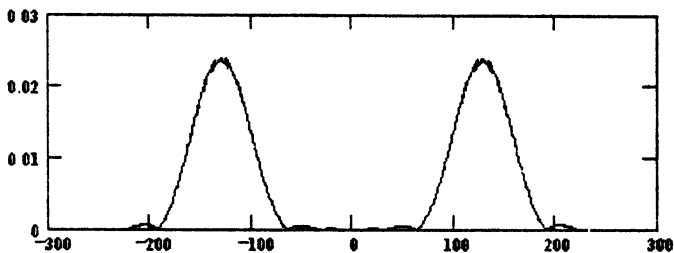


Рис. 18. Диаграммы направленности антенны в действительном и мнимом изображениях

Пространственный спектр сигнала на выходе фильтра не может быть шире полосы пропускания этого фильтра. В нашем случае полоса пропускания фильтра определяется шириной диаграммы направленности малой антенны. Принимая ширину диаграммы направленности за ширину полосы сигнала, получим следующее условие для разделения действительного и мнимого изображений: поворот антенны должен быть сделан на такой угол, который больше, чем ширина ее диаграммы направленности без учета поворота. При соблюдении этого условия взаимное проникновение действительного и мнимого изображений возможно только за счет пересечения сдвинутых диаграмм направленности, как показано на рис. 18. Это пересечение получается за счет боковых лепестков. Поэтому в данном случае борьба с боковыми лепестками является борьбой за получение изображения лучшего качества, свободного от влияния помехи за счет несфокусированного сигнала.

Каковы параметры получаемого изображения и чем они определяются?

При формировании изображения входной сигнал дважды подвергался пространственной фильтрации. Один раз при приеме на повернутую антенну согласно (2.27) и еще раз при восстановлении изображения согласно (2.19). Ширина фильтра пространственных частот в формуле (2.14) определяется величиной синтезируемой апертуры и составляет

$$\delta u = \frac{k}{z} D. \quad (2.28)$$

Ширина фильтра в (2.27) определяется шириной диаграммы направленности малой антенны и составляет  $\Delta u_d$ . Максимум диаграммы направленности повернут на угол  $\alpha$ , и фильтр пропускает только высокочастотную часть спектра входного сигнала, не про-

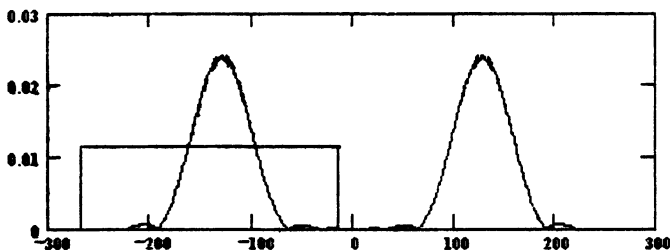


Рис. 19. Согласование областей пропускания фильтра (2.19) с пространственным спектром одного из изображений

пуская низких частот, по крайней мере до пространственной частоты  $\Delta u_d$ .

Смещая настройку фильтра, определяемого (2.19), путем смещения пределов апертуры, можно согласовать области пропускания обоих фильтров, как показано на рис. 19. Видно, как устраняется при согласовании полос пропускания мнимый несфокусированный сигнал (член, несущий информацию о мнимом изображении).

Чтобы искажения сигнала были меньше, желательно расширить полосу воспроизводимых пространственных частот как можно больше.

Увеличение величины синтезируемой апертуры будет приводить к увеличению полосы воспроизводимых пространственных частот, т. е. к росту разрешающей способности только до тех пор, пока не будет выполнено условие

$$\frac{kD}{z} \geq \Delta u_d. \quad (2.29)$$

При выполнении условия (2.29) дальнейший рост величины синтезируемой апертуры не будет приводить к увеличению разрешающей способности, так как она определится в этом случае шириной диаграммы направленности антенны, установленной на самолете, т. е. величиной  $\Delta u_d$ . Если на некоторой дальности выполнено условие (2.29), то на всех дальностях, меньших этой, разрешающая способность не будет зависеть от дальности и будет определяться только величиной  $\Delta u_d$ . Если при соблюдении условия (2.29) пересчитать угловую разрешающую способность на пространственное разрешение на местности, то разрешающая способность будет на местности определяться размером апертуры приемной антенны, установленной на самолете, в силу неравенства (2.3).



Вышеописанную систему получения радиоизображений вполне естественно назвать радиоголографической из-за ее сходства, если не сказать тождества, с голографией. Эта радиоголографическая система, как и всякая голографическая система, подвержена действию спекл-шума, с которым постоянно ведется борьба. В основе этой борьбы лежит избыточность получаемой информации об изображении с последующим исключением этой информации при одновременном усреднении спекла.

Сравним синтез апертуры с оптической голографией.

С момента фиксации значений  $P_d(x, z)$  вся обработка изображения ведется так же, как и в оптической голографии, причем даже в оптическом диапазоне с использованием лазеров.

В качестве принципиальных отличий можно указать следующие три.

1. Используется импульсный режим работы радиолокатора, позволяющий за один проход самолета тысячи голограмм для разных строк изображения, а разделение изображения на отдельные строки производится благодаря разделению принимаемых импульсов во времени.

2. Вместо опорного луча используется сигнал когерентного гетеродина, а вместо квадрата модуля комплексной амплитуды фиксируется ее действительная часть. В результате этого получают некоторые преимущества. В соотношении (2.26), аналогичном оптическому (1.42), остаются только два последних члена, представляющие информацию об изображении.

3. Условием разделения действительного и мнимого изображений является угол поворота приемной антенны на величину, превышающую ширину диаграммы направленности, а не соответствующий наклон опорного луча.

Так как вся эта разница не имеет принципиального характера (не оказывает влияния на качество изображения), то это направление антенной техники получило название "радиоголография".

Но так было не всегда. Создали радиоголографию в ее современном виде, включая и оптическую аналоговую обработку сигналов, те же Лейт и Упатниекс, которые дали путевку в жизнь оптической голографии, осуществив в ней эффективное разделение действительного и мнимого изображений. Однако оптическую голографию они создали в 1964 году, а радиоголографию в ее современном виде за 10 лет до этого. Они сами в то время не видели никакого сходства между получением радиоизображений с помощью радиолокаторов с синтезируемой апертурой и оптической голографией. В настоящее время тождество оптической и радиоголографии является общепризнанным, но тем не менее тот далекий

оптикам и близкий радиоинженерам язык, на котором велась разработка всей техники и идеологии построения антенн с синтезируемой апертурой, является чрезвычайно устойчивым. Мне не приходилось читать работ по апертурному синтезу, излагаемых так, как это было сделано выше с максимальным использованием языка и понятий оптической голографии.

Об устойчивости установившегося "радиоязыка" свидетельствует такой факт. На кафедре физики создавалась коллективная монография "Экспериментальная радиооптика". Будучи редактором этой монографии, я пытался добиться того, чтобы статьи, посвященные радиоголографии, были изложены не на традиционном "радиоязыке", а на языке оптической голографии. Так как мои устные и краткие письменные указания авторы игнорировали, то часто приходилось полностью изменять текст уже готовых статей, переписывая их заново и проводя свою оптическую линию примерно так, как это было изложено выше. Однако даже это несколько не помогло и соответствующие статьи в этой монографии остались написанными на традиционном радиоязыке. Что же это за язык? Почему он так устойчив?

Попробуем показать это на примере. Опишем на радиоязыке работу антенны с синтезируемой апертурой в дальней зоне, т. е. в условиях, когда справедливо неравенство (2.4). В этом приближении точечный источник, находящийся на расстоянии  $z$ , может считаться источником плоской волны. Угол ее падения на апертуру  $\alpha$  и будет направлением на этот источник. Пусть этот источник излучает монохроматическую волну частоты  $\omega$ . Двигаясь вдоль апертуры со скоростью  $v$ , примем излучение этого источника не на частоте  $\omega$ , а вследствие эффекта Доплера на частоте  $\omega + \Delta\omega$ , сдвинутой на  $\Delta\omega$  и определяемой соотношением

$$\Delta\omega = \omega \frac{v}{c} \sin \alpha \approx \omega \frac{v}{c} \alpha. \quad (2.30)$$

По величине смещения доплеровской частоты  $\Delta\omega$  можно определить угол  $\alpha$  или направление на источник, если таких источников несколько, то по частотному спектру можно на основании (2.30) определить угловой спектр расположения всех источников излучения.

Определим разрешающую способность такого метода. Точность нахождения частоты при спектральном анализе определяется временем наблюдения в соответствии с соотношением неопределенности:

$$\Delta(\Delta\omega) = 2\pi/T, \quad (2.31)$$

где  $T$  – время наблюдения.

С учетом (2.31) точность измерения угла  $\alpha$ , т. е. направления на источник, будет

$$\Delta\alpha = \frac{c}{\omega v} \Delta(\Delta\omega) = \frac{c}{\omega} \frac{2\pi}{vT} = \frac{\lambda}{D}, \quad (2.32)$$

где через  $D$  обозначен путь, который пройдет за время  $T$  приемник, двигаясь со скоростью  $v$ . Длина этого пути  $D$  эквивалентна некоторой апертуре  $D$ , а получившееся соотношение (2.32) является следствием более общего соотношения, аналогичного (2.7) для антенны с угловым сканированием.

Таким образом, на языке измерения доплеровских частот может быть описан синтез апертуры. Если апертура  $D$  не удовлетворяет условию (2.4), а существенно больше, то доплеровское смещение частоты не постоянно при движении вдоль всей апертуры и будет функцией не только угла, но и расстояния до источника. Такой сигнал может быть выделен соответствующей обработкой (включая, как будет показано ниже, и оптическую). Развивая эти соображения, можно получить соотношения (2.14) и (2.12). Так как когерентный радиолокатор был разработан именно для селекции движущихся целей на основе эффекта Доплера, то апертурный синтез, осуществляемый с помощью когерентного доплеровского радиолокатора, вполне естественно описывать именно на таком языке. Это объясняет его невероятную устойчивость и живучесть.

На этот радиоязык приходится переводить и всю идеологию последующей оптической обработки сигналов, что и было сделано Лейтом и Упатниексом. Вследствие такого описания какую-либо аналогию между апертурным синтезом и оптической голографией усмотреть было трудно и нет ничего удивительного в том, что Лейт и Упатниекс только через 10 лет работы пришли к идее оптической голографии, хотя она уже ничего нового в принципиальном отношении не содержала по сравнению с уже сделанной и давно внедренной радиоголографией.

Приведу еще один пример той разницы, которая имеется между оптическим и радиоязыком. Посмотрим, как выглядит на радиоязыке необходимость поворота приемной антенны на некоторый угол  $\alpha$  от оси полета. Если такого поворота не сделать и направить антенну (максимум ее диаграммы направленности) перпендикулярно к линии полета, то одинаковую доплеровскую частоту будут иметь сигналы, принимаемые под положительными и отрицательными углами в пределах диаграммы направленности антенны. По доплеровским частотам эти сигналы различить невозможно, они наложатся друг на друга и испортят изображение. Если же повер-

нуть антенну на некоторый угол, превышающий ее диаграмму направленности, то вследствие появления некоторого постоянного доплеровского смещения частоты, вызванного поворотом антенны, сигналы в пределах всей диаграммы направленности антенны будут иметь различные доплеровские частоты и наложения уже не будет. Так выглядит на радиоязыке проблема разделения действительного и мнимого изображений.

## 2.6. Проблемы теории и техники антенн

Выше мы познакомились с физическими основами формирования изображений с помощью антенн. Какие проблемы возникают в связи с этим в теории антенн и при их практическом создании?

Эти проблемы можно разделить на три группы.

1. Определение тех характеристик антенн, которые обеспечивают требуемые параметры изображений. Эта задача является по отношению к самой антенне внешней и ставится далеко не всегда. Часто требуемые характеристики антенны не задаются, а вытекают из тех возможностей, которыми располагает исследователь.

2. Выяснение возможных путей построения антенны с требуемыми характеристиками, вытекающими из решения задач первой группы. Такие задачи получили название синтеза антенн. Это крупное, развитое и продолжающее развиваться направление теории и техники антенн.

3. Вопросы, касающиеся измерения характеристик готовых антенн, определения соответствия того, что получилось практически в результате синтеза антенн, тому, что должно было получиться теоретически. Это тоже большое и все развивающееся направление.

Названные проблемы отнюдь не охватывают того, что подразумевается сейчас под антенной техникой и теорией антенн. Это только та часть антенной техники, которая связана с получением изображений и эквивалентна оптике, т. е. так называемая радиооптика. Область применения антенн много шире и выходит далеко за пределы темы этой книги.

Коснемся проблемы измерений в связи с тем, что она имеет непосредственное отношение к голографии.

Среди параметров, определяющих возможности антенны, есть такие, которые, несмотря на их важность, находятся только путем эксперимента. В качестве примера можно привести такой параметр, как уровень бокового поля антенны, под которым понимается

уровень боковых лепестков, существенно удаленных от основного. Этот параметр практически очень важен, так как даже очень малое боковое поле может являться существенной помехой при наличии сильного источника излучения, расположенного где-то сбоку вдаль от основного лепестка. Легко представить себе такую ситуацию, когда именно то ослабление, которое имеет антенна по боковому лепестку, определяет возможности антенны принимать слабое излучение.

Измерение уровня бокового поля и всей диаграммы направленности антенны осуществляется с помощью сильного точечного источника на основе соотношения (2.7), требующего для применения условий дальней зоны (2.4). Для больших антенн обязательное условие (2.4) приводит к неудобствам, связанным с необходимостью относить источник, по которому производится измерение, на весьма большое расстояние (даже поднимать вверх, в космос) и увеличивать в связи с этим мощность его излучения. Для того чтобы полностью исключить влияние нежелательных отражений, приходится принимать специальные меры, которые приводят к необходимости оборудования большого специального полигона для антенных измерений.

С появлением радиоголографии и оптической голографии был предложен принципиально новый способ для измерения диаграммы направленности антенн – в два этапа. На первом этапе предлагалось измерять не диаграмму направленности антенны, а распределение множителей  $A(x, y)$  на плоскости раскрыва антенны. На втором этапе определялась диаграмма направленности антенны путем вычислений по формуле (2.1).

На пути осуществления такого способа измерений встретились следующие трудности:

- положение плоскости раскрыва антенны никак не обозначено в пространстве, и ее следует еще определить путем каких-то измерений;

- когда положение плоскости раскрыва определено, то для осуществления измерений источник волн следует перемещать в пространстве по всей плоскости раскрыва антенны в пределах ее апертуры, в антеннах для этой цели не предусмотрено никаких приспособлений;

- при измерениях требуется фиксировать комплексный выход антенны, а не модуль, что тоже не предусмотрено при обычных режимах использования антенны.

Казалось бы, на этом перечень трудностей можно и закончить. Если их все удастся удовлетворительно преодолеть, то успех вполне обеспечен. Однако оказалось, что есть еще одна трудность, которую

при всей ее очевидности никто не предвидел, поэтому дружно взялись за преодоление вышеперечисленных трудностей, которых было достаточно для организации ударной работы. Путей преодоления было предложено и перепробовано много. Остановимся на одном, как представляется, достаточно эффективном пути.

Для исключения трудностей, связанных с отысканием положения плоскости раскрыва антенны и организацией сканирования по ней источником, был предложен следующий прием. Источник волн, по которому калибруется антенна, размещается на расстоянии от антенны, которое существенно превышает ее апертуру (в 5–10 раз), но не удовлетворяет условию (2.4), которое требует увеличения этого расстояния более, чем в  $D/\lambda$  раз по сравнению с апертурой антенны. Сканирование антенны в этом случае в процессе измерений осуществляется по углам  $\alpha$  и  $\beta$ , для чего у антенны имеются специальные устройства. Вычисление диаграммы направленности осуществляется на основе соотношения (2.6), которое справедливо при любом расстоянии от источника до антенны. Трудность, состоящая в требовании фиксировать значения комплексного выхода антенн, тоже оказалась легко преодолимой на основе того опыта, который был накоплен в голографии и радиоголографии. Был введен опорный сигнал, фаза которого изменялась вместе со сканированием антенны. Это позволило получить записи такие же, как и в радиоголографии.

Итак, все трудности оказались преодоленными. Что же получилось в результате?

На рис. 20 изображена диаграмма направленности антенны, которая определялась двумя способами. Один из них – это только что описанная радиоголографическая процедура с обработкой по соотношению (2.6). Второй – это классический способ с источником, расположенным в дальней зоне антенны, и угловым сканированием.

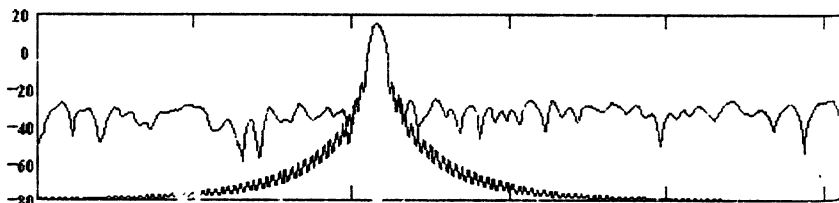


Рис. 20. Диаграмма направленности антенны, определенная в ближней зоне голографическим методом и в дальней зоне

Сравнивая эти две диаграммы направленности одной антенны, полученные двумя разными способами, можно видеть, что они хорошо совпадают только в области главного лепестка, а дальше резко расходятся. Там, где интенсивность боковых лепестков ниже, чем 20 дБ от основного лепестка, обе диаграммы резко различаются.

В чем дело и какая из диаграмм является истинной? Этот вопрос очень существен, так как измерение предназначено для определения уровня слабых боковых лепестков и должно давать правильный результат именно в этой области значений диаграммы направленности.

Здесь встретилась новая трудность, которая имеет принципиальное значение, на ней остановимся подробнее. Физическую природу этой трудности покажем на примере идеальной голографической схемы, в которой она проявляется в наибольшей степени.

Допустим, что мы измеряем распределение комплексных амплитуд и значения множителей  $A(x, y)$  в раскрытой антенне, положение которой точно известно. Пусть фиксируется и действительная, и мнимая части множителей  $A(x, y)$ . В этом случае диаграмма направленности  $\Gamma(u_1, u_2)$  определяется с помощью преобразований Фурье (2.1), которые являются однозначными. Выполняя их точно, без ошибок, можно, казалось бы, рассчитывать получить истинные значения  $\Gamma(u_1, u_2)$ , если действительно измерены истинные значения  $A(x, y)$ . Однако множители  $A(x, y)$  определяются нами не из теоретических соображений, а экспериментально — это означает, что вполне допустимы неточности и ошибки измерения. Они могут быть очень малы и играть лишь небольшую роль в значениях  $A(x, y)$ , но, тем не менее, нельзя определить экспериментально значения  $A(x, y)$  абсолютно точно. Это обстоятельство становится принципиальным. Допустим, что ошибки измерения можно представить в виде некоторого аддитивного шума, т. е. некоторой очень малой случайной добавки к значениям  $A(x, y)$ . Измеренное значение будет

$$\tilde{A}(x, y) = A(x, y) + a(x, y). \quad (2.33)$$

Сделаем два предположения относительно свойств ошибки или шума.

Эта ошибка в среднем одинакова для всех  $x$  и  $y$ , т. е. измеренное и истинное значения различаются примерно одинаково. Предположим также, что пространственный спектр шума  $a(x, y)$  тоже в среднем равномерен, т. е. в нем нет ярко выраженных частот. Это вполне естественные предположения, которые почти всегда выполняются.

На основании (2.1) и (2.33) для измеренной диаграммы направленности получим

$$\tilde{\Gamma}(u_1, u_2) = \Gamma(u_1, u_2) + \gamma(u_1, u_2). \quad (2.34)$$

Здесь  $\Gamma(u_1, u_2)$  — это истинная диаграмма направленности, а  $\gamma(u_1, u_2)$  — это пространственный спектр шума  $a(x, y)$ . Этот спектр более или менее равномерен. В нем равновероятны некоторые выбросы независимо от значений  $u_1$  и  $u_2$ . Истинная диаграмма обладает сильной неравномерностью. У нее есть большой главный максимум и боковые лепестки, которые намного меньше главного максимума, есть даже точки, где равны нулю.

Относительная точность измерения диаграммы направленности определяется как

$$\frac{|\Gamma(u_1, u_2) - \tilde{\Gamma}(u_1, u_2)|}{|\Gamma(u_1, u_2)|} = \frac{|\gamma(u_1, u_2)|}{|\Gamma(u_1, u_2)|}. \quad (2.35)$$

Из соотношения (2.35) видно, что искомая диаграмма направленности измеряется с высокой относительной точностью при больших значениях  $|\Gamma(u_1, u_2)|$ , т. е. в окрестности главного максимума. Для этой области значений  $u_1$  и  $u_2$  точность измерения диаграммы направленности с помощью описанного приема может существенно превосходить точность измерения исходных величин  $A(x, y)$ .

Однако в области значений  $u_1$  и  $u_2$ , где  $|\Gamma(u_1, u_2)|$  принимает весьма малые значения, ошибка измерений оказывается настолько большой, что измеренными с помощью описанного приема данными значениями  $|\Gamma(u_1, u_2)|$  вообще пользоваться нельзя, так как все получаемые значения целиком обусловлены только ошибками измерения.

Изложенные соображения объясняют полученные расхождения в измеренных двумя способами диаграммах направленности одной антенны, приведенные на рис. 20.

Подчеркнем еще раз, что особенность голографического процесса, связанная с необходимостью выполнения интегральных преобразований Фурье или Френеля, может приводить как к повышению точности измерений (диаграмма направленности в окрестности главного максимума), так и к катастрофическому ее падению (слабые боковые лепестки диаграммы направленности).

Все же эту проблему — измерение диаграммы направленности в ближнем поле антенны — удалось решить. Здесь не будем рассматривать этот вопрос во всех подробностях. Укажем только основные принципы, использованные при решении этой задачи.



Отметим, что необходимость выполнения (2.4) связана только с тем, что при этом условии точечный источник дает плоскую волну на всей апертуре антенны. Таким образом, условие (2.4) связано с тем, что мы располагаем только точечным источником, который нельзя располагать в ближней зоне антенны – в этих условиях он плоской волны не даст. Однако представим себе, что мы располагаем не только точечным источником, но и источником плоской волны, который дает плоскую волну на любом расстоянии от источника, включая и ближнее поле. Такой источник можно располагать и в ближней зоне антенны. Откликом антенны на подающую на ее вход (в плоскость раскрыва) плоскую волну будет значение диаграммы направленности при соответствующем угле. В этих условиях можно измерять в ближней зоне антенны и далекие малоинтенсивные лепестки диаграммы направленности.

Следовательно, задача заключается в том, чтобы создать источник плоской волны. Такая задача решается в оптике. С помощью объектива можно сферическую волну, идущую от точечного источника, преобразовать в плоскую. Такие оптические системы в оптике называются коллиматорами (делают лучи коллимированными, т. е. параллельными друг другу). В технике антенн для устройств, которые генерируют плоскую волну в своей ближней зоне, взяли оптическое название "коллиматор", а сам метод измерения диаграммы направленности антенны в ближней зоне с использованием плоской волны назвали коллиматорным методом.

Создать плоскую волну в ближней зоне можно различными способами, в том числе и тем, который используется в оптике. Однако в антенной технике возможны и свои, особые методы. Для возбуждения плоской волны в ближней зоне используются антенны с большой апертурой.

## 2.7. Математические программы по антенной тематике

### 2.7.1. Угловое сканирование антенны. Программа, моделирующая антенну с угловым сканированием, приведена на рис. 12П. Согласно (2.1) для углового сканирования антенны следует выполнить преобразование Фурье от распределения комплексных амплитуд в плоскости раскрыва антенны. При этом мы получим отклик как функцию пространственной частоты $u$ , которая зависит от угла в силу (1.9). Распределение в плоскости раскрыва антенны в данном примере задано функцией $A_k$ , представляющей собой одну

плоскую волну, падающую под некоторым углом, определяемым параметром  $f$ , и набор комплексных случайных чисел, которые добавлены к плоской волне. Плоскость раскрыва антенны дополнена нулями, причем интервал, занятый нулями, в 16 раз превышает интервал, занятый заданными значениями. Это имитирует реальную ситуацию, в которой предполагается, что плоскость раскрыва антенны бесконечна. Величина интервала, занятого нулями, определяет дискретизацию углового спектра. Чем больший интервал занят нулями, тем ближе могут быть расположены точки отсчета значений углового спектра. Интервал между ними равен  $2\pi/L$ , в то время как апертура антенны равна  $M = L/16$ . Особенность приведенной программы, на которую следует обратить внимание, — это специальная программа, позволяющая определить местоположение максимума результата углового сканирования антенны.

Такой программы в богатейшем арсенале "Mathcad" почему-то нет, и ее приходится составлять из "подручных средств". В качестве таких средств здесь использованы: программа, позволяющая определять величину максимума (или минимума) в числовой последовательности, и функция  $\Phi(k)$ , равная нулю для  $k < 0$  и единице для всех остальных значений  $k$ . В этой программе реализован следующий алгоритм. Разность значений последовательности и значения ее максимума равна нулю всего в одной точке: в той, где достигается максимальное значение. Функция  $\phi$  от этой разности равна единице всего в одной точке, а она совпадает с значением максимума. Если теперь образовать функцию, обозначенную на рис. 12П в виде  $ky_n$ , то значение ее максимума будет равно значению координаты максимума последовательности. В известном руководстве по программированию [13] содержится много различных программ поиска максимума функции, но такой программы, которая использована здесь, там нет, хотя она чрезвычайно проста и удобна. Даже при наличии интенсивного шума антенна в состоянии определить точно направление на источник излучения и может сделать это много точнее, чем ширина максимума. В программе, представленной на рис. 13П, показано, за счет чего и каким образом можно реализовать эти возможности. В предыдущем примере направление на источник излучения по координате максимума определялось с точностью до целого числа. Эта точность связана не с принципиальной возможностью определить направление максимума излучения, а с той дискретизацией, которой подвергался угловой спектр, определяемый до целого числа. Для увеличения точности определения направления в рамках той же программы необходимо увеличить число  $L$ . Такое увеличение можно сделать в два-три раза, но

не в 100 же раз. Здесь мы продемонстрируем иную возможность увеличения дискретности, которая допускает даже существенное уменьшение значения  $L$ . Именно эта программа приведена на рис. 13П. Обратите внимание на результат. Координата максимума определена с точностью до пяти знаков и все они верны. В данном случае точность лимитируется уже уровнем шумов в плоскости раскрыва антенны, а не дискретностью отсчетов углового спектра. Значение  $L$  при этом оставлено прежним.

Эта хитрость удалась благодаря тому, что отсчеты углового спектра в этой программе не распределены вдоль всех углов, а сконцентрированы только вблизи координаты максимума. Это сделано ценой отказа от быстрого преобразования Фурье и замены его обычным преобразованием, допускающим вычисление значений спектра не во всей спектральной области, а в любом месте и с любой частотой, как это нужно. Это преобразование осуществлено при вычислении функции  $G_{nm}$ . Благодаря достаточно частому сканированию в окрестности максимума его координата определяется с точностью, не зависящей от шага дискретизации по углу.

Координата максимума получающейся последовательности определяется по алгоритму, приведенному на рис. 12П и описанному выше.

**2.7.2. Выделение слабой волны на фоне сильной.** В программе, показанной на рис. 14П, поле в плоскости раскрыва антенны задано в виде суммы двух плоских волн  $A_k$ . Одна из этих волн в 100 раз (на 40 дБ) меньше другой. Особенностью программы является наличие множителя  $UP_n$ , на который умножается поле антенны на апертуре  $A_n$ . Этот множитель расширяет максимум отклика антенны, но и снижает уровень боковых лепестков. Благодаря существенному снижению уровня боковых лепестков слабый сигнал на отклике антенны отчетливо виден. Для сравнения приведен рис. 15П, на котором аподизирующий множитель равен единице. На рис. 16П с помощью аподизирующего множителя выделен шум, добавленный в плоскости раскрыва антенны. Эта программа позволяет определить, какой уровень шума (или ошибок при считывании поля в плоскости раскрыва антенны) препятствует измерению уровня боковых лепестков диаграммы направленности антенны. Иной, гораздо более мощный способ выделения слабого сигнала на фоне сильного приведен в программе рис. 17П. Здесь мощный сигнал, пространственная частота которого точно известна, полностью устраняется. Для того чтобы устранить сигнал, сначала определяется его комплексная амплитуда (комплексное число  $T$ ). В результате применения этой программы отчетливо виден сигнал, уровень которого меньше сильного сигнала на 60 дБ! При этом не

требуется никакой аподизации и угловое разрешение антенны полностью сохранено.

Следующая программа, изображенная на рис. 18П, показывает, как можно полностью устранить сильный сигнал, не зная его пространственной частоты. Она определяется методом, описанным при обсуждении программы рис. 12П. После этого этот сигнал устраняется точно так же, как и в программе рис. 17П. Для точного определения положения устраняемого сигнала применен прием, описанный в программе рис. 13П. В результате появляется возможность наблюдать сигнал, который слабее сильного на те же 60 дБ. Заметим, что классические методы адаптивных антенн [14] применимы только для случайных полей. В этой программе мы показываем возможность применения адаптивных (самонастраивающихся) алгоритмов для регулярных полей.

**2.7.3. Апертурный синтез в регулярном поле.** В программе, иллюстрирующей возможности апертурного синтеза, которая приведена на рис. 12П, есть все то, что содержится в предыдущих иллюстрациях. В ней последовательно повторяются все действия, описанные в разд. 2.4 "Радиоголография". Вначале приведен массив данных  $LM$  и апертура малой антенны  $L$ , установленной на самолете. Конкретные числовые данные можно изменять. Расстояния определены в длинах волн, а сама длина волны ( $p = 2$ ) составляет в данном примере минимально возможную величину, допускающую однозначность углового спектра. Вверху справа определена функция  $ZW$ , позволяющая выводить результаты в логарифмическом масштабе в децибелах.

Моделируется один источник, излучающий сферическую волну. Активная локация с освещением источника заменена пассивной, при которой источник является самосветящимся. В вопросах получения изображения нет разницы между пассивной и активной локацией, изображение строится в обоих случаях только при приеме сигнала. Поэтому для сокращения объема вычислений и экономной записи только самого существенного мы моделируем не активную локацию, а пассивный прием точечного излучателя.

Этап вычисления поля от излучателя приведен в программе, показанной на рис. 7П, а в этой программе сразу записано поле цилиндрической волны в виде  $A_k$ .

Здесь вновь пользуемся цилиндрическими координатами, позволяющими существенно сократить рассматриваемый объем данных и объем вычислений, показав все физические закономерности.

В правой части рисунка показано формирование аподизируемой антенны, установленной на самолете. Аподизирующий множи-

тель выбран в виде функции  $Y_m$ . Поле в раскрыве малой антенны представлено в виде плоской волны (сферичность ее фронта не сказывается в пределах апертуры антенны). Поворот антенны определяется параметром  $d$ . Его можно изменять при выполнении программы. Вид диаграммы малой антенны показан на рисунке в виде функции  $AC_k$ . Эта антенна осуществляет сканирование по пространству вдоль линии полета, что записано в виде функции  $S_k$ .

Далее функция  $S2_k$  изображает получаемую в полете радиоголограмму, представляющую собой выход одного канала фазового детектора. Вид этой голограммы от одного точечного источника показан на рисунке.

Дальнейшая программа повторяет то, что было в программах рис. 7П–9П – восстановление голограмм. Радиоголограмма восстанавливается точно так, как и оптические. Она освещается светом так, что форма комплексной амплитуды волны повторяет вид голограммы.

Спектр радиоголограммы показан на рисунке 19П в виде функции  $Q_k$  (сплошная линия). На том же графике показан фильтр  $e_k$ , образованный при синтезе апертуры. Результат восстановления радиоголограммы с фильтром показан на рисунке в виде функции  $C_k$ . Далее по описанному выше алгоритму определяется положение изображения, которое можно сравнить с тем, что было задано в виде функции  $x_k$ . Второе расплывчатое изображение существенно ослаблено фильтром.

Функция  $A1_k$  показывает вид радиоголограммы, записанной не малой, ориентированной особым образом антенной, а точечным ненаправленным приемником. Радиоголограмма  $S2_k$  представляет собой часть всей функции  $A1_k$ .

Так как эта программа включает в себя почти все предыдущее, то с ней интересно поиграть, изменяя значения параметров. Примерный вариант такой игры описан в виде задания.

*Задание:*

1. Попробуйте придавать различные наклоны антенне, изменяя параметр  $d$ .
2. Что будет, если убрать фильтр или изменить его положение?
3. Измените апертуру антенны  $L$ .
4. Измените координату точечного источника  $x$  и  $R$ .
5. Выясните роль аподизирующего множителя в антенне.
6. Во сколько раз произошло "сжатие" изображения в результате использования апертурного синтеза по сравнению с изображением, получаемым сканируемой антенной. От каких параметров задачи зависит величина этого сжатия?

### 3. ИЗОБРАЖЕНИЯ В ОПТИЧЕСКИХ СИСТЕМАХ

#### 3.1. Оптическая система

*Оптической системой* называются два участка свободного пространства, между которыми помещен оптический объектив. Под объективом условимся понимать плоскость, в которой комплексные амплитуды волнового поля умножаются на комплексную функцию вида  $L(x, y)$ . Здесь  $x$  и  $y$  – координаты точек плоскости, в которой происходит умножение. Вид функции  $L(x, y)$  определяет объектив.

Такое определение объектива сходно с определением антенны. Там тоже есть плоскость, в которой происходит умножение комплексных амплитуд на комплексные множители. Различие между объективом и антенной состоит в том, что в антенне производится не только умножение комплексных амплитуд поля на множители, а объектив ограничивается только этой операцией. Данное нами определение объектива (как и антенны) не совсем соответствует тому, что делает с полем реальный объектив. Определение касается только того, что реальный объектив должен делать, чтобы оптическая система могла формировать изображение. Отклонения в математических операциях, которые совершает реальный объектив от принятого нами определения, называются *абберациями*, их необходимо учитывать при описании работы реальных оптических систем.

Нашей целью является решение прямой и обратной задач для оптической системы. Под прямой задачей будем понимать преобразование заданного распределения комплексных амплитуд оптической системой, а обратная задача будет состоять в нахождении входного распределения комплексных амплитуд по тому полю, которое преобразовано оптической системой. Эта задача решается на основании связи между входными и выходными значениями распределений комплексных амплитуд, которые устанавливаются на основе решения прямой задачи.

Схема оптической системы приведена на рис. 21. Пользуясь для наглядности этой схемой, введем ряд определений. Линию, ориентированную вдоль оси  $z$  так, что на ней  $x = 0$  и  $y = 0$ , показанную на рис. 21, назовем оптической осью системы. Плоскость объектива перпендикулярна к оптической оси. Кроме плоскости объектива, обозначим еще две плоскости, также ориентированные нормально к оптической оси, которые назовем входной плоскостью

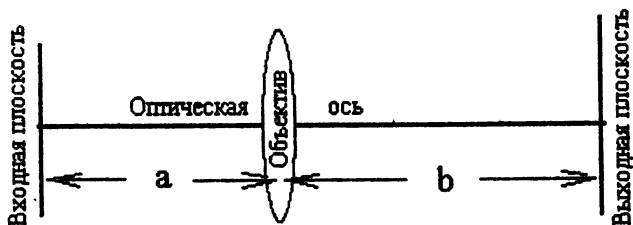


Рис. 21. Схема оптической системы

оптической системы и выходной плоскостью этой системы (см. рис. 21). Входная плоскость может быть проведена на любом расстоянии, но до объектива (при меньших значениях  $z$ ), а выходная плоскость проводится в любом месте после объектива на расстоянии  $b$  от него (при больших значениях  $z$ ). Распределение комплексных амплитуд, наблюдаемое в выходной плоскости при каком-то распределении комплексных амплитуд во входной плоскости, будем называть изображением входной плоскости в выходной плоскости. Найдем это изображение, опираясь на определение оптической системы.

### 3.2. Преобразование полей оптической системой

Обратимся к оптической системе, изображенной на рис. 21. Пусть на входной плоскости этой системы задано входное изображение в виде распределения комплексных амплитуд  $p(x, y)$ . На расстоянии равном  $a$  от входной плоскости располагается объектив оптической системы. Входное изображение, пройдя путь  $a$  до объектива, преобразуется свободным пространством согласно (1.5) или (1.18), которое запишем в виде

$$p_a(x, y) = \iint_{\infty} p(\zeta, \eta) H(x - \zeta, \eta - y) d\zeta d\eta, \quad (3.1)$$

где  $H(x, y)$  определяется (1.17) и имеет следующий вид:

$$H(x, y) = \frac{1}{i\lambda} \frac{\exp(ikR)}{R},$$

$$R = (x^2 + y^2 + a^2)^{1/2}. \quad (3.2)$$

Следующий этап преобразования поля – операция, выполняемая объективом. Условимся значения комплексных амплитуд поля, преобразованных объективом (прошедших через объектив), обозначать прописными буквами в виде  $P$ , а отсчет расстояний производить после этого от плоскости объектива. На выходе объектива распределение поля будет

$$P(x, y) = p_a(x, y) L(x, y). \quad (3.3)$$

После объектива есть еще один участок свободного пространства, имеющий длину  $b$ . Искомое распределение комплексных амплитуд на выходе оптической системы определится в виде следующей формулы:

$$p_b(x, y) = \iint_{\infty} p(\zeta, \eta) H_b(x - \zeta, y - \eta) d\zeta d\eta. \quad (3.4)$$

$H_b(x, y)$  определяется соотношением (3.2) с заменой расстояния  $a$  расстоянием  $b$ . С учетом (3.3) и (3.1) соотношение (3.4) можно записать в виде единой формулы, связывающей входное и выходное изображения:

$$P_b(x, y) = \iiint L(\zeta, \eta) p(v, \mu) H_a(\zeta - v, \eta - \mu) \times \\ \times H_b(x - \zeta, y - \eta) d\zeta d\eta dv d\mu. \quad (3.5)$$

В выражении (3.5) можно выделить операцию, не зависящую от входного сигнала, а определяемую целиком оптической системой. Она представляет собой интегрирование по переменным  $\zeta, \eta$ , от которых входное изображение не зависит.

Выделим эту операцию:

$$I(x, y, v, \mu) = \iint L(\zeta, \eta) H_a(\zeta - v, \eta - \mu) \times \\ \times H_b(x - \zeta, y - \eta) d\zeta d\eta. \quad (3.6)$$

Назовем (3.6) откликом оптической системы. С учетом (3.6) связь между входным и выходным изображениями можно записать в более компактном виде сравнительно с (3.5):

$$P_b(x, y) = \iint p(v, \mu) I(x, y, v, \mu) dv d\mu. \quad (3.7)$$

Выражение (3.7) решает задачу, сформулированную нами выше, связывая входное изображение и выходное. Это выражение решает полностью прямую задачу. Такое фундаментальное издание по оптике, как [15], не идет далее этого соотношения.



Однако нас это соотношение не устраивает. Из полученной формулы не видно, какие конкретные преобразования совершает с изображением оптическая система. Предлагается математическая задача – преобразовать выражение (3.7) так, чтобы выявить последовательность математических операций, которые совершает оптическая система, преобразуя входное изображение в выходное. Надо найти такое выражение, эквивалентное (3.7), из которого была бы ясна физическая картина преобразований, которым подвергается комплексная амплитуда поля в оптической системе. С этой целью необходимо ввести особое понятие, которое не вводится в литературе, посвященной описанию преобразований, совершаемых в оптических системах. Это *идеальная оптическая система*. Мы придаем этому понятию столь важное значение, что посвящаем ему самостоятельный раздел.

### 3.3. Идеальная оптическая система

Условимся идеальной считать оптическую систему, у которой объектив имеет функцию  $L(x, y)$  следующего специального вида:

$$L(x, y) = \exp \left[ -i \frac{k}{2F} (x^2 + y^2) \right] L(0, 0). \quad (3.8)$$

В дальнейшем будем опускать множитель  $L(0, 0)$ , полагая его равным единице.

Параметр  $F$ , входящий в (3.8), называется *фокусным расстоянием объектива*. Идеальная оптическая система является чисто воображаемым устройством, ее нельзя практически осуществить, потому что ее объектив имеет бесконечно протяженную апертуру.

В дальнейшем введем специальный множитель, превращающий идеальную оптическую систему в практически реализуемую наподобие того, как в разд. 2.3 был введен аналогичный множитель для фокусируемой антенны.

Для идеальной оптической системы выражение (3.6) примет вид

$$E(x, y, v, \mu) = \iint_{\infty} \exp \left[ -i \frac{k}{2F} (\zeta^2 + \eta^2) \right] \times \\ \times H_a(\zeta - v, \eta - \mu) H_b(x - \zeta, y - \eta) d\zeta d\eta. \quad (3.9)$$

Через  $E$  условимся обозначать отклик идеальной оптической системы.

На основе (3.9) мы получаем возможность ввести еще один необычайно важный параметр оптической системы, без которого поставленная нами выше задача не может быть решена. Для введения этого параметра еще раз перепишем (3.9), заменив входящие в него обе функции  $H$  их приближениями вида (1.28), в котором опустим постоянный фазовый множитель, не зависящий от координат:

$$E(x, y, v, \mu) = -\frac{1}{\lambda^2 ab} \exp\left[i \frac{k}{2} \varphi(x, y, v, \mu)\right] \times \\ \times \iint \exp\left[i \frac{k}{2} \Delta(\zeta^2 + \eta^2) - ik\psi(\zeta, \eta)\right] d\zeta d\eta, \quad (3.10)$$

где

$$\varphi(x, y, v, \mu) = \frac{x^2 + y^2}{b} + \frac{v^2 + \mu^2}{a}, \quad (3.11)$$

$$\psi(\zeta, \eta) = \frac{v\zeta + \eta\mu}{a} + \frac{\zeta x + \eta y}{b}. \quad (3.12)$$

Соотношение (3.10) в силу приближения (1.28) хорошо описывает случай больших  $a$  или  $b$  и не описывает случай, когда они обращаются в нуль. Для этого случая требуется особое рассмотрение. Оно проводится путем замены того  $H$ , которое не участвует в преобразовании поля из-за равенства нулю расстояния, тождественным преобразованием.

При  $a = 0$  получим

$$E_{a=0} = \exp\left[-i \frac{k}{2F} (v^2 + \mu^2)\right] H_b(x - v, y - \mu), \quad (3.13)$$

при  $b = 0$

$$E_{b=0} = \exp\left[-i \frac{k}{2F} (x^2 + y^2)\right] H_a(x - v, y - \mu). \quad (3.14)$$

В этих случаях функция  $E$  интегральных преобразований не содержит.

В том приближении, в котором мы рассматриваем задачу, переменные  $x$  и  $y$  полностью разделяются. Поэтому можно не выписывать формулы, содержащие обе переменные, а достаточно написать выражение, содержащее только  $x$ . Полное выражение всегда можно написать путем добавления сомножителя, в котором координата  $x$  заменена координатой  $y$ .

С учетом сделанного замечания выражению (3.10) придадим следующий компактный вид:

$$E(x, y, v, \mu) = E_x(x, v) E_y(y, \mu), \quad (3.15)$$

где

$$E_x(x, v) = \frac{i}{\lambda(ab)^{1/2}} \exp \left[ i \frac{k}{2} \left( \frac{v^2}{a} + \frac{x^2}{b} \right) \right] \times \\ \times \int \exp \left[ ik \frac{v^2}{2} \Delta - ik\zeta \left( \frac{v}{a} + \frac{x}{b} \right) \right] d\zeta. \quad (3.16)$$

Через  $\Delta$  в (3.16) и (3.10) обозначен очень важный параметр оптической системы, который назовем *настройкой оптической системы*. Этот параметр выражается через другие параметры системы следующим образом:

$$\Delta = \frac{1}{a} + \frac{1}{b} - \frac{1}{F}. \quad (3.17)$$

Величина этого параметра определяет класс преобразований, осуществляемых оптической системой. Неудобство формулы (3.7), побудившее нас двинуться дальше, состоит в том, что в ней параметр  $\Delta$  не выделен.

### 3.4. Изображения, формируемые идеальной оптической системой

Сначала обратим внимание на два особых частных случая, которые следует рассматривать отдельно, на их основе, опираясь на [1], придем к общему выводу.

Первый частный случай заключается в условии  $\Delta = 0$ , имеющем специальное название – оптическое сопряжение входного и выходного изображения. Подставив  $\Delta = 0$  в (3.10), получим [1]

$$PE_b(x) = i \sqrt{ab} \exp \left( ikx^2 \frac{a+b}{2b^2} \right) p \left( -\frac{a}{b} x \right). \quad (3.18)$$

Главная особенность полученного соотношения – однозначное соответствие между точками входного и выходного изображений. Сами изображения отличаются одно от другого, и мы остановимся на этих отличиях, но если входное изображение изменить в одной точке, то выходное изображение изменится тоже только в одной

точке. Точки входного и выходного изображений, связанные однозначно, называются *оптически сопряженными*. Множитель, изменяющий амплитуду выходного изображения, обеспечивает безусловное сохранение энергии при получаемом преобразовании. Этот множитель может быть получен из энергетических соображений. В дальнейшем мы не будем уделять внимание этому множителю, так как он не зависит от координат  $x$ ,  $y$  и в образовании изображения не участвует. Следующий множитель изменяет только фазу выходного изображения. Вид этого множителя точно такой, как и функция  $L(x, y)$  объектива идеальной оптической системы, определяемой (3.8) (для одной переменной). Этот фазовый множитель представляет собой объектив идеальной оптической системы с отрицательным значением фокусного расстояния  $F_v$ :

$$F_v = -\frac{b^2}{a+b}. \quad (3.19)$$

Этот результат давно известен в оптике. Изображение, полученное в сопряженной плоскости, ведет себя так, как будто бы оно проходит через дополнительный объектив оптической системы. Чтобы нейтрализовать действие этого объектива, называемого *виртуальным*, в тех случаях, когда оно нежелательно (виртуальный объектив сокращает область видимости изображения), используют окуляр (окуляром называют объектив, устанавливаемый в плоскости изображения или вблизи нее) Гюйгенса, имеющий фокусное расстояние, обратное (3.19).

Последний сомножитель (3.18) устанавливает однозначное соответствие между входным и выходным изображениями. Такая связь тоже давно известна в оптике при наличии оптического сопряжения. Выходное изображение оказывается по сравнению с входным перевернутым и с измененным масштабом в соответствии с отношением длин  $a/b$ .

Другой частный случай, требующий отдельного рассмотрения, получается при условии

$$b = F, \quad \Delta = \frac{1}{a}. \quad (3.20)$$

В этом случае (3.10) переходит в

$$E_F(x) = \left(\frac{k}{2\pi F}\right)^{1/2} \exp\left[ikx^2\left(\frac{F-a}{F}\right)\right] g\left(\frac{k}{F}x\right) \quad (3.21)$$

(см. [1]), и наблюдается изменение амплитуды и виртуальный объектив, обладающий фокусным расстоянием:

$$F_w = - \frac{F}{2(F-a)}. \quad (3.22)$$

В (3.21) тоже наблюдается однозначное соответствие двух функций, но теперь однозначное соответствие с измененным масштабом с выходным изображением имеет уже не само входное изображение, а его пространственный спектр. Теперь изменение в одной-единственной точке пространственного спектра входного изображения будет давать изменение в одной точке выходного изображения, сопряженной с этой точкой спектра.

Это тоже известнейший результат. В некоторых книгах (см., например, [16–17]) считается, что в плоскости, отстоящей от объектива на его фокусное расстояние (эта плоскость называется фокальной плоскостью объектива), входное распределение обращается в свой спектр, но не принято упоминать, что этот результат относится только к идеальной оптической системе, которую создать невозможно.

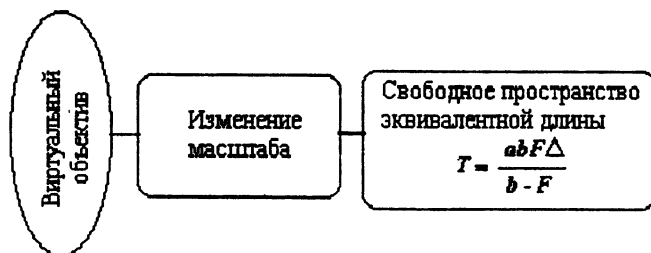
Структура (3.21) такова, что допускает подавление виртуального объектива без применения окуляра Гюйгенса. Чтобы не образовывался виртуальный объектив, достаточно входную плоскость разместить, как и выходную, на расстоянии  $F$  от объектива. Утверждение, что изображения по обе стороны объектива, удаленные на расстояние  $F$  от него, сопряжены по Фурье, весьма популярно в литературе [16–18]. Ниже рассмотрим, при каких условиях это действительно имеет место.

Теперь перейдем к общему случаю, когда нет оптического сопряжения и  $b \neq F$ .

В общем случае тоже есть все три компонента: амплитудный множитель, обеспечивающий сохранение энергии волны, виртуальный объектив и однозначное соответствие выходного изображения и входного с измененным масштабом, прошедшего некоторый эквивалентный путь в свободном пространстве  $T$ . Величина эквивалентного пути  $T$  определяется следующей формулой:

$$T = \frac{abF\Delta}{b-F}. \quad (3.23)$$

В итоге преобразование полей, осуществляемое идеальной оптической системой, можно представить в виде функциональной схемы (рис. 22), которая описывает и частные случаи, рассмотренные нами отдельно. Действительно, если  $\Delta = 0$ , то и  $T = 0$ , что соответствует оптическому сопряжению. При  $b = F$  эквивалентное расстояние  $T = \infty$ , при котором изображение согласно (1.35) переходит в свой спектр.



**Рис. 22.** Функциональная схема преобразования полей в идеальной оптической системе

Таким образом, идеальная оптическая система нами полностью описана, но она не реальна. Теперь поставим задачу описать аналогичным образом реальную оптическую систему. Для решения этой задачи воспользуемся идеальной оптической системой, как некоторым физическим понятием, с помощью которого решим более общую задачу.

### 3.5. Реальная оптическая система – это идеальная система + фильтр

Для того чтобы придать реальность идеальной оптической системе, ограничим ее апертуру. Введем некоторый комплексный множитель  $M(x, y)$ , который обращается в нуль вне апертуры. С помощью этого множителя формулу (3.8) запишем так, чтобы система с этим объективом была бы реальной:

$$L(x, y) = \exp \left[ -i \frac{k}{2F} (x^2 + y^2) \right] M(x, y). \quad (3.24)$$

Не будем пока накладывать на форму  $M(x, y)$  каких-либо ограничений за исключением условия обращения в нуль вне апертуры. Для упрощения структуры формул допустим, что в форме  $M(x, y)$  переменные разделяются, т. е. ее можно представить в виде произведения соответствующих функций только одного переменного. Тогда с учетом (3.24) выражение (3.10) можно записать для реальной оптической системы в виде [1]

$$P_b = i \sqrt{\frac{1}{\lambda ab \Delta}} \int_{\infty} p(v) M \left[ \frac{1}{\Delta} \left( \frac{v}{a} + \frac{x}{b} \right) \right] \exp [\psi(v, x)] dv,$$

$$\psi(v, x) = ik \frac{v^2}{2b} \left( 1 - \frac{1}{a\Delta} \right) + ik \frac{x^2}{2b} \left( 1 - \frac{1}{b\Delta} \right) - ik \frac{vx}{ab\Delta}. \quad (3.25)$$

Соотношение (3.25) интерпретировать трудно, так как оно содержит сразу две неопределенные функции – это входное изображение  $p(v)$  и  $M(v, x)$ .

Однако у нас есть возможность написать еще одно уравнение – отклик идеальной оптической системы. Это уравнение получается из (3.25) простой заменой  $M(v, x)$  на единицу. Исключив из этих двух уравнений  $p(v)$ , можно выразить изображение, полученное реальной оптической системой, через изображение, полученное идеальной оптической системой, и функцию  $M(v, x)$ .

Эти действия выполнены в [1], и искомая связь получилась в виде

$$P_b(x) = \frac{1}{\lambda b} \int E_b(\zeta) C_M \left[ \frac{k}{b} (x - \zeta) \right] \exp \left[ i \frac{k}{2b} (\zeta^2 - x^2) \right] d\zeta. \quad (3.26)$$

Здесь  $C_M(u)$  – пространственный фурье-спектр функции  $M(x)$ .

В таком виде это соотношение получено в [1], но там оно по-настоящему не интерпретировано. Перепишем (3.26) в виде

$$P_b(x) \exp \left( i \frac{k}{2b} x^2 \right) = \frac{1}{\lambda b} \int_{\infty} E_b(\zeta) \exp \left( i \frac{k}{2b} \zeta^2 \right) C_M \left[ \frac{k}{b} (x - \zeta) \right] d\zeta. \quad (3.27)$$

Полученное соотношение можно интерпретировать. Экспоненциальные множители имеют вид виртуальных объективов. Их можно убрать из реальной оптической системы с помощью окуляров Гюйгенса. Если это сделать, введя дополнительные объективы в оптическую систему (в [1] такая возможность показана применительно к оптическому сопряжению), то полученное соотношение легко может быть интерпретировано, так как оно становится полностью аналогичным отклику линейной системы, представленному в виде (1.11). Выход идеальной оптической системы играет роль входного сигнала, а функция  $C_M(u)$  – роль импульсной характеристики линейного фильтра. В результате получается, что функциональную схему реальной оптической системы можно изобразить в виде, показанном на рис. 23.

Таким образом, реальная оптическая система преобразует изображение не так, как это делает идеальная система. Преобразова-

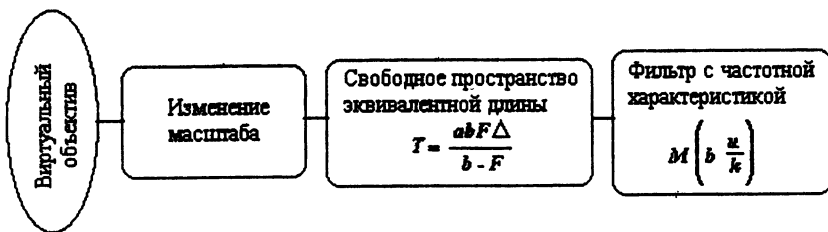


Рис. 23. Функциональная схема преобразования полей в реальной оптической системе

ние, совершаемое реальной системой, может практически не отличаться от того, что делает идеальная система, при двух условиях. Во-первых, в оптической схеме должен быть соответствующий окуляр Гюйгенса. Практические схемы оптических устройств такие элементы в своем составе имеют. Примером может служить фотоувеличитель, в котором так называемая конденсорная линза играет именно такую роль. Без этой линзы или при неправильной ее настройке поле изображения сужается. Допустим, что такой элемент в оптической системе имеется. Но есть еще и линейный фильтр. Его отклик представляет собой спектр функции  $M(x)$ , а частотная характеристика этого фильтра на основании (1.12) будет  $M\left(b \frac{u}{k}\right)$ . Если изображение, сформированное идеальной оптической системой, проходит через этот фильтр, не изменяясь или так изменяясь, что этими изменениями можно пренебречь, то тогда реальная оптическая система будет работать, как идеальная.

### 3.6. Как оптическая система преобразует изображение

На рис. 23 изображена функциональная схема оптической системы, к которой с помощью вышеприведенных рассуждений сведена формула (3.7). Эта схема показывает, какие преобразования с входным изображением позволяет совершать оптическая система. Но прежде чем мы приступим к обсуждению этих преобразований, подчеркнем еще раз, что эта схема описывает только регулярное поле, которое называется когерентным. Для обычного освещения, с которым мы имеем дело, эта схема не годится. Для получения регулярного поля в оптике необходимо использовать



лазер. При обычном освещении *та же самая оптическая система преобразует изображения иначе.*

Первые три звена оптической системы описывают преобразования, совершаемые введенной нами идеальной оптической системой. Эти звенья замечательны тем, что их при желании можно свести к тождественному преобразованию и тем самым полностью выключить, что всегда и делается с первым звеном – виртуальным объективом или виртуальными объективами, если их несколько. Оптическая система практически никогда не служит средством создания виртуального объектива, и действие таких объективов принято компенсировать. Поэтому не будем подробно обсуждать действие первого звена.

Следующее звено – изменение масштаба изображения – используется по своему назначению. Однако в схемах когерентного освещения эта операция редко бывает главной. Примером, когда используется именно это звено, являются фотоувеличитель или эквивалентная ему проекционная система с точечным источником света. Бинокль и подзорная труба не описываются схемой, изображенной на рис. 23, так как они работают с распределенными источниками света и не в регулярных полях. В увеличителе или проекционной системе не обязательно должен быть применен лазер, чтобы его работа описывалась схемой рис. 23. Такое применение когерентной схемы для фотоувеличителя позволяет без труда объяснить назначение всех деталей оптической схемы фотоувеличителя, особенности его работы и настройки, но это не является характерным примером использования когерентного света в оптических системах. Не будем более останавливаться на оптической схеме фотоувеличителя в конфигурации рис. 23, она подробно рассмотрена в [11].

Следующее звено характерно только для регулярного поля или лазерного освещения. Это звено является эквивалентом прохождению сигнала в свободном пространстве пути  $T$ , определяемого посредством (3.23). Замечательным свойством этого звена, благодаря которому оно и используется, является возможность как изменения знака свободного пространства, проходимого волновым полем, так и "сжатие" свободного пространства. Под последним мы понимаем существенное сокращение реальных размеров оптической системы  $(a + b)$  по сравнению с эквивалентной длиной  $T$ , которая может быть даже бесконечной. Это звено позволяет восстанавливать голограммы, записанные как в виде действительного, так и в виде мнимого изображений (это получается либо при  $F > b$ , либо при  $F < b$ ). При  $F = b$  мы имеем бесконечную эквивалентную длину свободного пространства, а преобразование изобра-

жений, которое совершается при этом оптической системой, является преобразованием Фурье.

Это свойство оптической системы, работающей в регулярном когерентном поле, предлагается использовать для оптического метода спектрального анализа сигналов, включая двумерные [19]. Исследователей привлекает скорость и точность осуществляемых преобразований. Основной трудностью на пути практического применения оптических методов спектрального анализа сигналов в когерентном свете является сложная система ввода и вывода информации. С этой целью разрабатываются специальные кристаллы, способные модулировать регулярное световое поле заданным электрическим сигналом [20]. Для вывода информации предлагаются телевизионные системы [21] или специальные полупроводниковые структуры [22]. Рассматриваемое звено выключается путем настройки оптической системы на оптическое сопряжение.

Последнее, четвертое звено выключить нельзя, так как ни одна оптическая система не может иметь бесконечную апертуру. Это звено представляет собой линейный фильтр. Форму частотной характеристики фильтра можно изменять, варьируя форму функции  $M(x, y)$ . Эта функция определяет величину и форму апертуры объектива оптической системы и отличия фазы и амплитуды функции  $L(x, y)$  реального объектива от идеального, описываемого формулой (3.8).

С подобным или в точности таким же фильтром мы уже встречались при рассмотрении антенн в разд. 2.6 и особенно фокусируемых антенн в разд. 2.7. Дополнительно к сказанному там о функциях  $M(x, y)$  (или  $q(x, y)$  в разд. 2.7) отметим возможность осуществления пространственной фильтрации с помощью этого фильтра — см. работы, посвященные оптическим методам обработки информации, опубликованные сразу же после появления лазеров [23].

В приложении (рис. 20П) приведена математическая программа, позволяющая моделировать преобразования, осуществляемые оптической системой. Это моделирование использует непосредственно соотношения (3.1)–(3.7) и определение объектива (3.8), а в основу положена схема, изображенная на рис. 21, с существенным дополнением — еще одним объективом, помещенным во входную плоскость системы. Этот объектив исключает соответствующий виртуальный объектив, позволяя описывать преобразования, осуществляемые оптической системой в полном соответствии с функциональной схемой рис. 23.

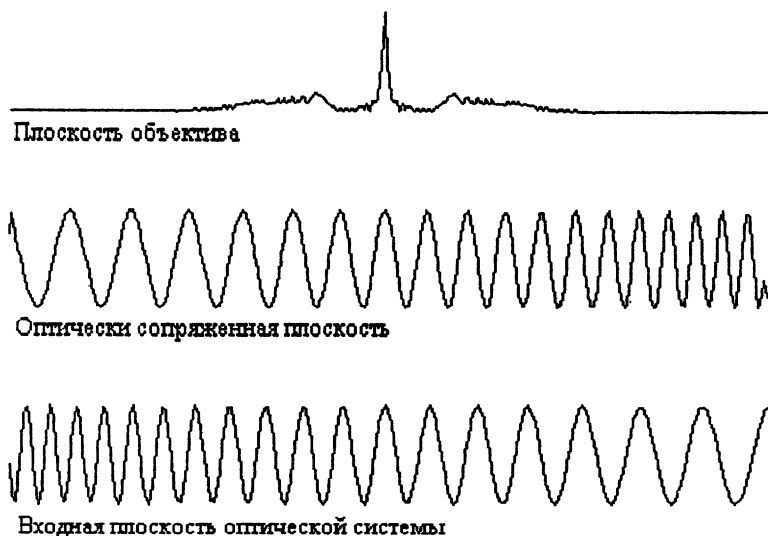


Рис. 24. Модули изображений в различных плоскостях оптической системы при  $a = b = 2F$

Проиллюстрируем это примерами. На рис. 24 приведены модули комплексных амплитуд, вычисленные в трех плоскостях: плоскости расположения основного объектива, плоскости, удаленной от него на удвоенное фокусное расстояние, и во входной плоскости, тоже удаленной от объектива (в другую сторону) на удвоенное фокусное расстояние. В этом случае входная и выходная плоскости получают оптически сопряженными. В плоскости объектива наблюдается модуль пространственного спектра входного сигнала. Входным сигналом служит действительное частотно-модулированное колебание, к которому добавлена постоянная составляющая. Получившиеся сигналы во входной и выходной плоскостях являются точно подобными, но перевернутыми в точности так, как полагается.

На рис. 25 показаны те же сигналы, в той же оптической системе, что и на предыдущем рисунке, но, в отличие от него, показаны не модули изображений, а их реальные части. При этом между входным и выходным сигналами никакого соответствия не наблюдается. Выходной сигнал несколько не похож на входной. Это объясняется влиянием виртуального объектива, формируемого в выходной плоскости системы. На виде модуля виртуальный

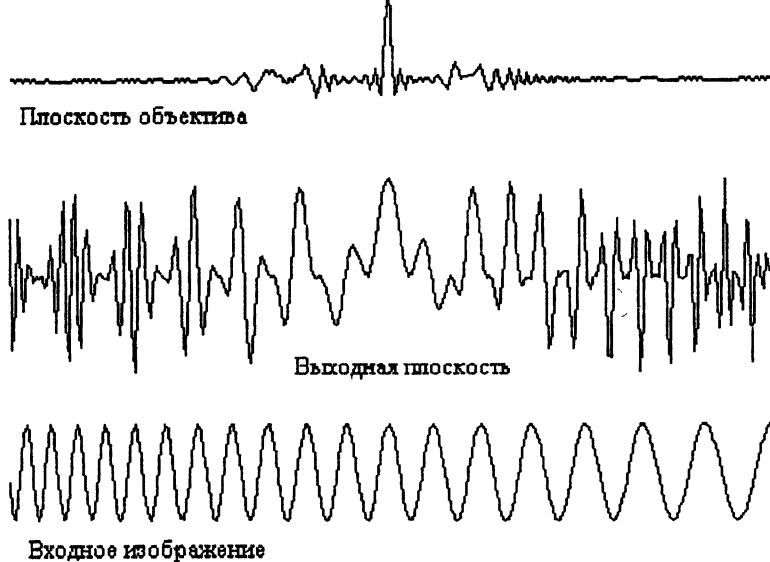


Рис. 25. Реальные части изображений, модули которых показаны на рис. 24

объектив не сказывается, а реальную часть он сильно искажает. Вид реальной части этого виртуального объектива показан на рис. 26, полученном в той же программе (см. рис. 20П), в которой в качестве входного сигнала взята одна постоянная составляющая.

Несколько слов скажем о математической программе, с помощью которой получены приведенные иллюстрации. Как уже отмечалось выше, эта программа построена целиком на формулах, входящих в определение оптической системы, причем все формулы имеют тот же вид, что и в тексте. По этим формулам компьютер работает существенно медленнее, чем по иному варианту тех же формул, приведенных на рис. 21П. В этой программе формулы свертки вида (3.1) записаны через трансформанты Фурье. Благодаря такой записи (математически совершенно эквивалентной), появляется возможность существенно ускорить вычисления, применяя быстрое преобразование Фурье. В примерах, приведенных в тексте программ, тоже показаны модули изображений в тех же трех плоскостях. В качестве входного сигнала взята симметричная функция косинус. Поэтому входное и выходное изображения при их наложении полностью совпали, а в плоскости объектива отчетливо виден классический модуль спектра косинуса.

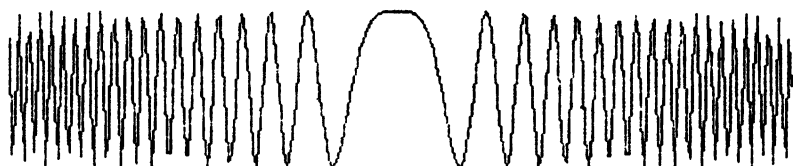


Рис. 26. Реальная часть изображения в выходной плоскости оптической системы, когда на ее входе постоянное поле

Свободное пространство длиной  $2F$  за объективом в этой оптической схеме формирует обратное преобразование Фурье от комплексного сигнала в плоскости основного объектива. Если вмешаться в этот сигнал, прикрыв его часть или изменив фазу на каком-то участке этой плоскости, то это скажется на изображении, образуемом в выходной плоскости. Это изменение изображения можно трактовать как процесс фильтрации — изменение спектра сигнала. С помощью оптической системы пространственный фильтр с заданной полосой пропускания можно получать с помощью ножиц, вырезая ими маски, закрывающие отдельные области частот. Эту возможность демонстрирует рис. 27, на котором показан результат вырезания постоянной составляющей из спектра сигнала вида  $1 + \cos(fx)$ . Изображение, сформированное оставшимися боковыми частотами, получается двойной частоты.

В теории линейной фильтрации рассматривается так называемый согласованный фильтр, обладающий целым рядом интересных и полезных свойств. Такой фильтр называется согласованным с заданным сигналом  $x(t)$  в результате того, что комплексная частотная характеристика такого фильтра равна комплексно-сопряженному спектру заданного сигнала.

Пусть спектр сигнала  $x(t)$  будет согласно (1.13)  $C_x(\omega)$ . Тогда частотная характеристика фильтра, согласованного с сигналом  $x(t)$ , будет

$$Z_x(\omega) = \overline{g_x(\omega)}, \quad (3.28)$$

где черта сверху означает комплексное сопряжение.

Интерес к согласованным фильтрам основывается на следующем свойстве отклика этого фильтра на сигнал, с которым он согласован.

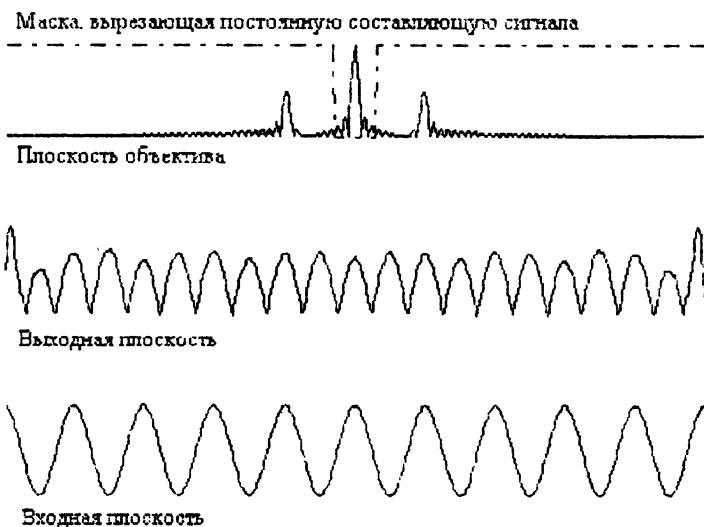


Рис. 27. Преобразование спектра сигналов в оптической системе

Этот отклик представляет собой так называемую простую функцию, у которой, в отличие от функции сложной, произведение величин эффективной длительности и эффективной ширины спектра минимально (порядка  $2\pi$ ). При этом эффективная ширина спектра на выходе согласованного фильтра такая же, как и на входе. Сохранение эффективной ширины спектра при согласованной фильтрации своего сигнала следует из тождественного равенства модулей спектра самой функции  $x(t)$  и модуля частотной характеристики фильтра. Такой фильтр не в состоянии существенно подавить ни одну спектральную составляющую сигнала, с которым он согласован. Поэтому, если произведение длительности исходного сигнала на длительность его спектра составляет

$$\Delta t \Delta \omega \approx 2\pi N, \quad (3.29)$$

где  $N$  – сложность сигнала, то отклик фильтра на сигнал, с которым он согласован, сожмется в своей эффективной длительности в  $N$  раз и составит

$$\tau = \frac{\Delta t}{N}. \quad (3.30)$$

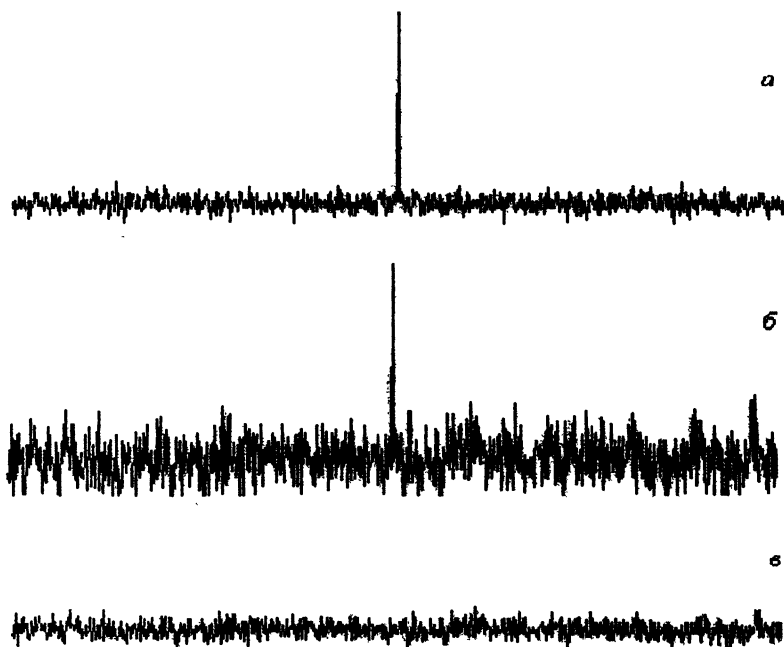


Рис. 28. Отклик согласованного фильтра на "свой" сигнал (а), на "чужой" (б) и на смесь сигналов (в)

Благодаря этому свойству отклик фильтра на сигнал, с которым он согласован, представляет собою такой короткий импульс, который может сформировать фильтр, обладающий шириной полосы  $\Delta\omega$ . Свойства и применение такого фильтра по выделению сложных сигналов определенной формы на фоне аддитивных шумов иллюстрируются на рис. 28 и 22П.

На рис. 28, а показан отклик фильтра на сигнал  $x(t)$ , с которым он согласован; на рис. 28, б — отклик того же фильтра на сигнал  $g(t)$ ; на рис. 28, в — отклик фильтра на сумму сигналов  $x(t) + n g(t)$ , где  $n = 10$ , а средние интенсивности сигналов  $x(t)$  и  $g(t)$  примерно равны. В этой смеси сигналов присутствие сигнала  $x(t)$  можно обнаружить на выходе фильтра, согласованного с этим сигналом. На рис. 22П приведена математическая программа, позволяющая выделять "свой" сигнал на фоне многих других.

Оптическая система благодаря своему четвертому звену (см. рис. 23) позволяет осуществлять двумерную согласованную фильтрацию изображений. Для этого надо в плоскость расположения объектива оптической системы поместить маску, прозрачность которой представляет собой двумерный комплексно-сопряженный спектр того сигнала, который собираемa фильтровать. Эта маска должна быть согласована как по своему масштабу, так и по положению и повороту с фильтруемым сигналом. Преимущество именно оптической системы согласованной фильтрации состоит в возможности перебора масштабов, положений и поворотов всего при одной-единственной маске и возможности осуществлять этот перебор сколь угодно быстро.

Необходимые технические детали осуществления оптической согласованной фильтрации и примеры приведены в статье [23]. Ограничимся указанием на эту статью и приведем математическую программу, моделирующую оптическую согласованную фильтрацию – рис. 23П. В качестве сигнала принята последовательность случайных чисел, обладающих нормальным распределением, нулевым средним и единичной дисперсией. Фильтр стоит в плоскости основного объектива. В той же плоскости помещен окуляр Гюйгенса, нейтрализующий образующийся виртуальный объектив. Согласованным фильтром является комплексно-сопряженный спектр сигнала, полученный в той же оптической системе. Это обеспечивает согласование масштабов.

В заключение отметим еще раз: все, что сказано выше об оптической системе и об антеннах, справедливо только для регулярных полей. При этом под изображением\*мы понимали невидимое глазом и неулавливаемое датчиками оптического изображения распределение комплексных амплитуд волнового поля.

Весь предыдущий материал является фундаментом для рассмотрения практических задач формирования изображений, понимаемых как распределение интенсивностей волновых полей и формируемых не только регулярными, но и случайными волновыми полями при разных статистических предположениях. Этим вопросам посвящена вторая часть книги.



## **ИЗОБРАЖЕНИЯ В СЛУЧАЙНЫХ ПОЛЯХ**

### **4. ОПТИЧЕСКАЯ СИСТЕМА В СЛУЧАЙНОМ ПОЛЕ**

#### **4.1. Какие поля считаются случайными**

В реальных условиях при естественном освещении волновые поля, создающие изображения, не являются регулярными. Определить условия, при которых получается регулярное поле, просто – не надо учитывать статистический характер полей, образующих изображение. Поэтому выделим условия регулярности поля и будем считать, что во всех остальных случаях учитывать статистику необходимо.

Чтобы поле было регулярным, достаточно выполнения одного условия: поле должно быть монохроматическим. Чисто монохроматический сигнал не может быть случайным, но это условие должно быть достаточно жестким. Если поле не является в нужной степени монохроматическим, то оно может быть и регулярным, если его источник имеет достаточно малые размеры, чтобы считаться точечным. Мы уже отмечали выше, что оптические проекционные системы с точечным источником белого света могут рассматриваться как системы, обладающие регулярным полем.

Эти два условия – точечность источника и точечность спектра – в какой-то степени нужны для получения регулярного поля. Оба условия можно, в принципе, выполнить путем ограничения размера источника или ограничения его спектра. Однако это неприемлемо по причине ограничения мощности источника. Габор говорил, что голографией, требующей регулярного поля, можно заниматься и без лазера, ограничив апертуру и спектр обычного источника света, но при этом экспозиция фотопластинки, фиксирующей голограмму, должна иметь длительность от Рождества Христова до

наших дней. Только лазер изменил ситуацию в оптике, а в радиодиапазоне видение осуществлялось с самого начала на основе мощного монохроматического излучения достаточно малых антенн. Поэтому до лазеров, а в безлазерной оптике и по сию пору, оптические изображения формируются существенно иначе, чем радиоизображения, так как в радио есть мощные регулярные поля, а нерегулярные весьма слабы.

В акустике ситуация более разнообразна. Там есть практически все. Есть мощные источники распределенного шума и малые по волновым размерам источники такого шума, можно создавать достаточно интенсивные монохроматические колебания и наблюдать мощный распределенный в пространстве шум.

Таким образом, рассмотрением только регулярных полей можно ограничиться для радиоизображений, лазерной оптики и ряда акустических задач. Это не мало, но совершенно не достаточно для многих случаев.

Посмотрим на сложившуюся ситуацию с нерегулярным полем с точки зрения теории регулярных полей. Необходимо выяснить, что нужно уточнить, чтобы рассмотреть формирование полей во всем объеме и во всех случаях. Заранее оговоримся, что не будем касаться влияния неоднородностей среды на формирование изображений, так как этот вопрос рассмотрен в литературе достаточно полно ([7, 28] и вторая часть монографии [28], написанная С. М. Рытовым, Ю. А. Кравцовым и В. И. Татарским).

#### **4.2. Какие параметры случайного поля влияют на формирование изображения**

Основой для нашего рассмотрения служит известный факт, что любые поля являются регулярными в течение очень короткого промежутка времени, когда их амплитуды и фазы не успевают измениться. На этом основывается условие регулярности поля, связанное с его монохроматичностью, которое уже упоминалось выше. Пусть промежуток времени настолько мал, что рассматриваемое нами поле регулярно. Любая система, формирующая изображение, содержит обязательный элемент в виде линейного фильтра, кроме идеальной системы с бесконечной апертурой, которую невозможно реализовать. Разница в формировании изображений регулярными и нерегулярными полями получается благодаря

наличию этого фильтра. Рассмотрим, как формируется выход линейного фильтра в условиях нерегулярных колебаний.

На основе (1.11) выход фильтра напишем в следующем виде:

$$y(x) = \int_{-\infty}^{\infty} s(\zeta) h(x - \zeta) d\zeta. \quad (4.1)$$

Пусть сигнал, наблюдаемый на выходе этого фильтра, флуктуирует вследствие нерегулярности входной функции  $s(x)$ , фаза и амплитуда которой изменяются во времени. Пусть эти изменения настолько быстрые и глубокие, что следить за ними не имеет смысла. Однако средняя интенсивность колебания, получающегося на выходе фильтра, достаточно стабильна за большое время и представляет интерес.

Такова общая философия перехода к нерегулярным полям от регулярных. Поскольку нет смысла следить за изменениями амплитуд и фаз сигналов в силу их быстрой и основательной смены во времени, нет необходимости и в определении изображения через комплексную амплитуду поля. Можно наблюдать только усредненные величины, которые не флуктуируют в силу усреднения по очень большому числу отдельных реализаций случайного процесса.

Итак, пусть реально наблюдаемой величиной на выходе фильтра является функция

$$U = \langle y(x) \overline{y(x)} \rangle, \quad (4.2)$$

в которой, как и всюду далее, угловые скобки означают усреднение по времени, а черта сверху — комплексное сопряжение. Мы наблюдаем среднюю интенсивность сигнала как функцию пространственной переменной  $x$ , а мгновенные значения комплексной функции  $y(x)$ , усредненные за такое же время, для нас интереса не представляют. Подставим (4.1) в (4.2):

$$U = \iint_{-\infty}^{\infty} \langle s(\zeta) \overline{s(\xi)} \rangle h(x - \zeta) \overline{h(x - \xi)} d\zeta d\xi. \quad (4.3)$$

Мы получили чрезвычайно интересное и важное соотношение, являющееся основой теории формирования изображений в нерегулярных полях. Особенностью полученного соотношения является то, что средняя интенсивность сигнала на выходе линейного фильтра определяется *корреляционной функцией входного сигнала, а не его интенсивностью.*

Почему это так важно?

Это важно потому, что определение корреляционной функции входного сигнала *требует дополнительных статистических гипотез, без которых выход фильтра не определен. В зависимости от тех или иных статистических предположений о виде функции корреляции входного сигнала выход фильтра (4.3) изменяется кардинальным образом.*

Поэтому изображение в случайном поле в одной системе, его формирующей, может получиться *существенно различным.*

На формирование изображений в нерегулярных полях оказывает влияние только функция корреляции полей, имеющая вид

$$\Gamma(x, y, \zeta, \eta) = \left\langle p(x, y) \overline{p(x - \zeta, y - \eta)} \right\rangle. \quad (4.4)$$

Эта функция в связи со своей исключительной ролью в формировании изображений имеет специальное название – *функция когерентности.* Свойства функции когерентности применительно к случайным процессам в виде функции времени подробно рассмотрены С. М. Рытовым в [7].

Случайное волновое поле характеризуется флуктуациями амплитуд и фаз сигналов в различных точках. Полное описание флуктуаций поля требует знания не только функции когерентности, но и функции корреляции полей, в которой нет комплексного сопряжения одной из функций [1, 24]. Формирование изображений определяется только функцией когерентности.

### 4.3. Виды функций когерентности

Наиболее употребительны следующие три статистические модели функций когерентности, которые поясним для одного измерения. Обобщение на случай двух измерений вполне очевидно, а формулы для одного измерения получаются компактнее.

**Первая модель, называемая однородным волновым полем, выглядит следующим образом:**

$$\Gamma(x, \zeta) = \Gamma(\zeta). \quad (4.5)$$

Согласно модели (4.5) функция когерентности зависит только от разностей координат точек и не зависит от того места, в котором определяется эта разность. Такая модель функции когерентности весьма распространена при описании случайных явлений в виде функций времени. К этой модели даже стараются (и не без успеха) свести более общие случаи [25]. При описании функций времени

эта модель называется моделью стационарного случайного процесса. Для описания функций времени этой модели с небольшими усовершенствованиями, сделанными на базе [25], оказывается вполне достаточно.

С моделью (4.5) нам придется иметь дело, но только не при статистическом описании входных сигналов. Для их описания модель статистически однородного процесса принципиально неприменима и никакие усовершенствования не могут сделать ее приемлемой. Причиной этого является понятие изображения, которое теперь необходимо ввести иначе. Понимать под изображением будем только усредненную функцию. Таковой является интенсивность сигнала, излучаемого каждой точкой поверхности, когда изображение формируется в случайном поле. В модели однородного поля интенсивность излучения во всех точках поверхности постоянна и поэтому в рамках этой модели никакого изображения сформировать нельзя. Для того чтобы получить возможность вводить изображение как распределение средней интенсивности поля, введем следующую статистическую модель функции когерентности [1].

**Вторая статистическая модель функции когерентности:**

$$F(x, \zeta) = [W(x)]^2 \delta(\zeta). \quad (4.6)$$

Вторым сомножителем в (4.6) является дельта-функция Дирака, равная нулю всюду, кроме одной точки при аргументе, равном нулю, а интеграл от этой функции равен единице. Физически модель (4.6) означает такое освещение, при котором любые две соседние точки обладают полностью некоррелированными полями. В этой модели изображение на входной плоскости описывается функцией  $W(x, y)$ . Что касается предположения о полной некоррелированности источников излучения в соседних точках, то оно справедливо для самосветящихся источников типа экрана электронно-лучевой трубки. Поверхности, освещенные обычным белым рассеянным светом, таким свойством не обладают. Конечный интервал корреляции вместо нулевого влияет на направленность излучения, и в пределах достаточно малых углов эта идеализация является приемлемой [1].

Кроме перечисленных двух статистических моделей есть еще и третья модель, относящаяся к регулярному полю. Дело в том, что регулярные поля можно тоже наблюдать с длительным усреднением по их интенсивности. Поэтому формально понятие изображения, введенное для случайных полей, может быть применено и к регулярному полю при соответствующей форме функции когерентности. Вид функции когерентности для регулярного поля очевиден.

Третья статистическая модель функции когерентности имеет вид

$$\Gamma(x, \zeta) = p(x) \overline{p(x - \zeta)}, \quad (4.7)$$

который для регулярных полей получается в силу предположения об отсутствии флуктуаций поля и сохранении всех значений поля во времени неизменными (в течение всего времени наблюдения). Таким образом, для регулярного поля имеет место следующее равенство:

$$p(x) \overline{p(x - \zeta)} = \left\langle p(x) \overline{p(x - \zeta)} \right\rangle, \quad (4.8)$$

откуда следует (4.7).

#### 4.4. Частотно-контрастная характеристика оптической системы

Найдем отклик реальной оптической системы в условиях нерегулярного поля, пользуясь общей формулой (3.27) и статистической моделью самосветящегося экрана (4.6). Средний квадрат (3.27) запишется в виде

$$\begin{aligned} |P_b|^2 = & \iint_{\infty} \left\langle E_b(\zeta) \overline{E_b(\xi)} \right\rangle \exp [(k/2b)(\zeta^2 - \xi^2)] \times \\ & \times C_m [(k/b)(x - \zeta)] \overline{C_m [(k/b)(x - \zeta)]} d\zeta d\xi. \end{aligned} \quad (4.9)$$

Для того чтобы интеграл (4.9) можно было вычислить, зная параметры оптической системы, необходимо определить функцию когерентности на выходе идеальной оптической системы. Поступим следующим образом. Воспользуемся свойством идеальной оптической системы, настроенной на оптическое сопряжение, заключающееся в однозначном соответствии входного и выходного изображений. На основе этого свойства, выраженного в формуле (3.18), можно найти функцию когерентности на выходе идеальной оптической системы. С учетом (3.18) и (4.6) эта функция будет [1]

$$\begin{aligned} \left\langle E_b(\zeta) \overline{E_b(\xi)} \right\rangle = \\ = \frac{a}{b} \exp \left[ i \frac{k}{2} \frac{a+b}{b^2} (\zeta^2 - \xi^2) \right] W^2 \left( -\frac{a}{b} \zeta \right) \delta \left[ \frac{a}{b} (\zeta - \xi) \right]. \end{aligned} \quad (4.10)$$

Подставляя (4.10) в (4.9) и интегрируя по  $\xi$ , пользуясь свойствами  $\delta$ -функции, получаем [1]

$$|P_b(x)|^2 = \frac{k^2}{4\pi^2 b^2} \int_{-\infty}^{\infty} W^2 \left(-\frac{a}{b}\zeta\right) \left| C_m \left[ \frac{a}{b}(x - \zeta) \right] \right|^2 d\zeta. \quad (4.11)$$

Это выражение для отклика реальной оптической системы на входное изображение, заданное в виде распределения средней интенсивности волны. Интересно отметить, что все виртуальные объективы, свойственные регулярному полю, благополучно исчезли. Оптическую систему, работающую в условиях обычного, не лазерного света и не точечного, а распределенного (лучше всего самосвещающегося поля), не надо снабжать никакими дополнительными объективами или иными устройствами, чтобы иметь возможность описывать ее работу посредством (4.11).

Соотношение (4.11) можно без каких-либо приближений трактовать как отклик фильтра, так как структура (4.11) соответствует выражению (4.1).

Поэтому функциональная схема оптической системы в условиях нерегулярного освещения выглядит так, как это показано на рис. 29. Входной сигнал преобразуется по своему масштабу и

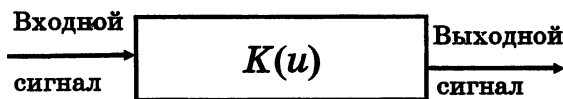


Рис. 29. Функциональная схема оптической системы в некогерентном свете

проходит через фильтр, частотная характеристика которого  $K(u)$  называется *частотно-контрастной характеристикой* (ЧКХ) оптической системы. На основании (4.11) отклик оптической системы (эта функция не имеет общепризнанного наименования) будет выглядеть следующим образом:

$$Q(\zeta) = \left| C_m \left[ \frac{k}{b}\zeta \right] \right|^2. \quad (4.12)$$

Частотно-контрастная характеристика оптической системы получается путем преобразования Фурье выражения (4.12). Учитывая, что  $C_m(u)$  является фурье-трансформантой функции  $M(x)$ , получаем

$$K(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} M \left( \frac{v}{b} \right) \overline{M \left[ \frac{v}{k} (v - u) \right]} dv. \quad (4.13)$$

Частотно-контрастная характеристика оптической системы является интегралом свертки и представляет собой так называемую положительно определенную функцию. Такие функции обладают абсолютным максимумом в нуле частот. Поэтому оптическая система всегда отдает предпочтение постоянной составляющей. Оптическая система, функционирующая в случайном поле, не может сделать такого преобразования сигналов, которое показано на рис. 27, когда постоянная составляющая сигнала полностью удалена. Такой операции (даже чуточку увеличить контраст какой-либо частоты) оптическая система в обычном свете сделать не в состоянии. Поэтому применение лазеров, сделавших возможными подобные операции в оптических системах, было воспринято как настоящая революция в оптике.

Частотно-контрастную характеристику оптической системы мы нашли только при условии оптического сопряжения. А как быть в общем случае? Ведь отыскать функцию когерентности при отсутствии оптического сопряжения не так просто.

Оказывается, для получения частотно-контрастной характеристики оптической системы в общем случае достаточно знать ее только в условиях оптического сопряжения [1]. Прием, позволяющий так сделать, основан на том, что функция  $M(x)$  произвольна. Поэтому ее можно ввести так, что оптическое сопряжение будет получаться при любой настройке  $\Delta$  оптической системы. Введем новую функцию  $J(x)$  следующим образом:

$$J(x) = M(x) \exp \left[ i \frac{k}{2} x^2 \Delta \right]. \quad (4.14)$$

Используя функцию (4.14) в определении объектива, всегда можно иметь оптическое сопряжение. Следовательно, эту функцию можно подставлять вместо  $M(x)$  в (4.13), получая выражение для частотно-контрастной характеристики в общем виде. Прделав эту операцию, получим [1]

$$K(u) = \frac{1}{2\pi} \exp \left( - \frac{b^2 u^2 \Delta}{2k} i \right) \int_{-\infty}^{\infty} M \left( \frac{v}{k} \right) \overline{M \left( \frac{v-u}{k} \right)} \times \\ \times \exp \left( - i \frac{b^2}{k} u v \Delta \right) dv. \quad (4.15)$$

Для функции  $M(x)$ , равной единице вне апертуры объектива, из (4.15) получаем [1]

$$K(u) = K_{\Delta=0}(u) \frac{\sin [ub\Delta (D - ub/k)]}{ub\Delta (D - ub/k)}. \quad (4.16)$$



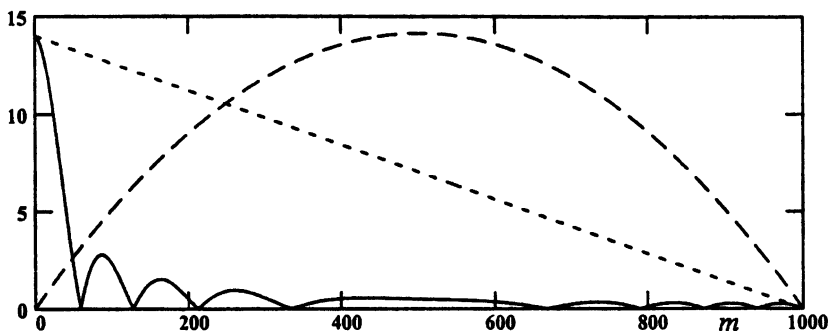


Рис. 30. Модуль частотно-контрастной характеристики неискажающего объектива при дефокусировке:  
 сплошная линия —  $\Delta = 9\lambda/D^2$ ,  
 мелкий пунктир — идеальная фокусировка,  
 крупный пунктир —  $\sin x/x$  в формуле (4.16)

График функции (4.16) приведен на рис. 30, из которого видны особенности частотно-контрастной характеристики оптической системы. Отклонение настройки от оптического сопряжения приводит к сужению характеристики, и благодаря этой особенности оптическую систему в условиях обычного освещения удастся настроить на оптическое сопряжение по "резкости" изображения, при котором ширина этого фильтра максимальна. В то же время обращает на себя внимание тот факт, что в случае предельно высоких частот, наблюдаемых в условиях оптического сопряжения с предельно малым контрастом, ширина фильтра слабо зависит от настройки оптической системы. По изображению, содержащему предельно высокие частоты, на "резкость" оптическую систему не наведешь.

Если сравнить преобразования изображений, получающиеся в одной и той же оптической системе в регулярном и случайном полях, то становится ясно, что общего мало. В самом деле, обе схемы содержат линейные фильтры, но в случае регулярного поля частотная характеристика фильтра никак не зависит от настройки оптической системы, а в случайном поле эта настройка является определяющим фактором, по которому обычно и настраивают "систему на фокус", т. е. добиваются оптического сопряжения. Функциональное звено в виде эквивалентного свободного пространства вообще отсутствует при нерегулярном поле, нет и виртуальных объективов, свойственных регулярному полю.

#### 4.5. Роль фокальной плоскости объектива в случайном поле

Остановимся еще на одной особенности функционирования оптической системы в случайном поле. Мы видели, что в регулярном поле выделяются две настройки оптической системы – это настройка на оптическое сопряжение и настройка, при которой  $b = F$  и которая называется настройкой на фокальную плоскость. В рассмотренной выше оптической системе настройка на фокальную плоскость не обладает какой-либо особенностью и ничем не выделена, что тоже является существенным отличием работы оптической системы в регулярном и нерегулярном полях.

Однако настройка на фокальную плоскость выделяется и в случайном поле при иной статистической модели функции когерентности [1]. Если для функции когерентности принять модель однородного поля, в которой справедливо условие (4.5) и корреляционная функция зависит только от разности координат, то настройка на фокальную плоскость становится аналогичной настройке на оптическое сопряжение, но только не для распределений интенсивности, а для угловых спектров полей.

При наблюдении в фокальной плоскости статистической модели для функции когерентности вида (4.5) получается следующее выражение для средней интенсивности поля на выходе оптической системы [1]:

$$\langle |P_F(x)|^2 \rangle = \frac{k^2}{16\pi^2 F^2} \int_{-\infty}^{\infty} G\left(\frac{k\zeta}{F}\right) \left| C_* \left[ \frac{k}{F}(x - \zeta) \right] \right|^2 d\zeta. \quad (4.17)$$

Здесь  $G(u)$  – спектр функции когерентности, заданной в виде (4.5), представляющей собой угловой спектр излучения входной плоскости. Этот спектр определяется посредством функциональной схемы, изображенной на рис. 29, с той же самой частотно-контрастной характеристикой. Для модели (4.5) имеет значение только настройка на фокальную плоскость, настройку на оптическое сопряжение эта модель игнорирует точно так же, как модель (4.6) игнорирует настройку на фокальную плоскость. Заметим, что в случае настройки на фокальную плоскость определяется не пространственный спектр какого-либо сигнала, а распределение излучения по углам. Пространственный спектр сигналов оптическая система в нерегулярном поле определять не может.

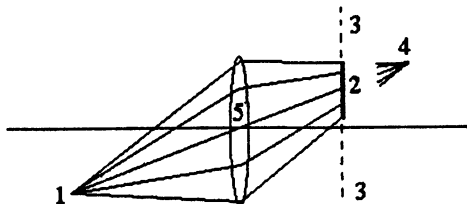
#### 4.6. Как вводится частотно-контрастная характеристика оптической системы в некогерентной оптике

До изобретения лазеров оптика была только некогерентной. Поэтому все понятия и соотношения, рассмотренные нами выше в разд. 3, оптиков не интересовали, и они их не вводили. Понятие частотно-контрастной характеристики существовало, но оно вводилось не так, как это сделано выше, и не могло быть введено таким образом, потому что преобразования регулярных полей оптиками не рассматривались.

Частотно-контрастная характеристика может быть введена следующим образом [26]. На основании статистической модели функции когерентности вида (4.5) можно считать, что сигналы, преобразованные оптической системой от каждой отдельной точки входной плоскости, суммируются по интенсивности, не интерферируя друг с другом. Таким образом, найдя отклик оптической системы на излучение отдельно взятой точки входной плоскости, можно найти отклик оптической системы от всей входной плоскости путем суммирования откликов от отдельных точек входной плоскости. Если при этом предположить, что форма отклика оптической системы не зависит от расположения точки, а только смещается, то отклик оптической системы на входной сигнал, заданный распределением интенсивности, выразится формулой вида (4.1). Роль импульсной характеристики фильтра будет играть отклик оптической системы на излучение отдельной точки входной плоскости.

В том случае, когда имеет место оптическое сопряжение, форма отклика определяется с помощью волновой оптики с учетом дифракционных явлений на апертуре объектива. Интенсивность этого отклика выражается функцией  $|C_M(u)|^2$ , а значит, для отклика оптической системы сразу же можно получить формулу (4.11), из которой следует формула (4.13). При отсутствии оптического сопряжения отклик на точечный источник может быть определен на основе геометрической оптики, как это показано на рис. 31. При расширении отклика частотно-контрастная характеристика сужается. Таким путем получают все характеристики оптических систем, работающих в условиях статистической модели некогерентного освещения (4.6).

Оптики не предполагали не только появления лазеров, но и того, что проекционные оптические системы необходимо описывать иначе. В книгах для фотолюбителей сообщалось, что фотоувеличитель с конденсором и точечным источником света дает более резкие



**Рис. 31.** Определение отклика на точечный источник при отсутствии оптического сопряжения:

1 – источник, 2 – отклик, 3 – плоскость изображения источника, 4 – сопряженная точка, 5 – объектив

отпечатки, чем увеличитель с тем же объективом, но с матовым стеклом, рассеивающим свет перед входным изображением. Однако никому в голову не приходило пытаться объяснить этот эффект изменением функциональной схемы всей оптической системы. В работе [11] подробно рассмотрена теория работы проекционных оптических систем, основанная на изложенных выше соображениях, содержащая анализ не только крайностей, когда свет либо когерентен, либо нет, но и промежуточной модели, которая легко получается путем плавного изменения размера источника света в проекционной системе. Изложенные соображения в [11] подтверждены экспериментом.

Оптики считают, что частотно-контрастная характеристика является исчерпывающей характеристикой объектива. На самом деле это не так. Исчерпывающей характеристикой объектива является его комплексная функция  $M(x)$ , или  $C_M(u)$ , но не  $K(u)$ , характеризующая объектив только для некогерентного света.

Входной сигнал на входе оптической системы, работающей в случайном поле, обязательно положителен, поэтому нельзя на вход системы подать чисто гармонический сигнал. Возможен гармонический сигнал вида синуса или косинуса в смеси с постоянной составляющей, причем отношение постоянной составляющей к переменной не может быть менее единицы. Это отношение называется контрастом. На выходе оптической системы контраст не возрастает, он может только убывать. Частотно-контрастная характеристика оптической системы показывает, как убывает контраст по мере увеличения частоты гармонической составляющей. Отсюда и происходит название этой характеристики – частотно-контрастная.

## 5. АПЕРТУРНЫЙ СИНТЕЗ В СЛУЧАЙНОМ ПОЛЕ

### 5.1. Функция когерентности в свободном пространстве

Рассмотрим, как изменяется вид функции когерентности при распространении волны в свободном пространстве. Расчет сделаем тем же путем, которым определялся выход оптической системы.

Примем, что в течение очень короткого времени поле можно считать регулярным, описываемым формулой (1.5) или аналогичной формулой для одномерного случая вида [1]

$$p(x, z) = \left(\frac{1}{i\lambda}\right)^{1/2} \int p(\zeta, 0) \exp\left[ikz + i\frac{k}{2z}(x - \zeta)^2\right] d\zeta \quad (5.1)$$

с учетом малоуглового приближения.

Будем искать так называемую поперечную корреляционную функцию, получающуюся при одинаковом  $z$  у каждого из сомножителей:

$$\Gamma(x, \zeta, z) = \langle p(x, z) \overline{p(x - \zeta, z)} \rangle. \quad (5.2)$$

Подставив (5.1) в (5.2), получим искомое соотношение:

$$\Gamma(x, \zeta, z) = \frac{1}{\lambda z} \iint \Gamma(\zeta', \zeta, 0) \times \\ \times \exp\left\{i\frac{k}{2z}\left[(x - \xi)^2 - (x - \zeta' - \xi - \zeta)^2\right]\right\} d\xi d\zeta. \quad (5.3)$$

В случае модели (4.6):

$$\Gamma(\xi, \zeta, 0) = (W(\xi)^2) \delta(\zeta). \quad (5.4)$$

Подставив (5.4) в (5.3) и проинтегрировав, получим

$$\Gamma\left(x, \frac{zu}{k}, z\right) = \frac{1}{\lambda z} \exp\left(iux - \frac{izu^2}{2k}\right) \int W^2(\zeta) \exp(iu\zeta) d\zeta. \quad (5.5)$$

Заметим, что здесь  $\zeta$  есть разность координат точек и отношение  $(zu)/k$ , являющееся новой переменной, тоже представляет собой разность координат точек, в которых определяется функция когерентности.

Введем следующую функцию, являющуюся пространственным спектром изображения:

$$G_w(u) = \frac{1}{\lambda z} \int W^2(\zeta) \exp(iu\zeta) d\zeta. \quad (5.6)$$

Подставив (5.6) в формулу (5.5), для точки  $x = 0$  получим

$$\Gamma\left(0, \frac{zu}{k}, z\right) = G_w(u) \exp\left(-i \frac{zu^2}{2k}\right). \quad (5.7)$$

Таким образом, мы получили искомую зависимость функции когерентности от расстояния  $z$ . Интересно отметить, что эта функция зависит от разности координат, т. е. представляет собой однородное поле. Значит, от модели однородного поля мы не ушли: она нас не устроила на входе, но все равно она "вылезла" на выходе.

## 5.2. Некогерентная голография

До сих пор мы считали, что голография получается только в когерентном свете, для нее требуется строго регулярное поле. Оказывается, что существует и некогерентная голография, основанная на свойствах функции когерентности случайного поля: удовлетворяющего условиям модели (4.6). Голография основана на свойствах комплексной амплитуды регулярного волнового поля. Покажем, что точно такие же свойства имеет функция когерентности. Запишем соотношение (5.1) для регулярного поля в виде

$$p(x, z) = \frac{e^{ikz}}{2\pi} \int g(u) \exp\left(-i \frac{z}{2k} u^2 + iux\right) du, \quad (5.8)$$

а затем перепишем в виде

$$p(x, z) = \frac{e^{ikz}}{2\pi} \int g(u, z) \exp(iux) du, \quad (5.9)$$

где  $g(u, z)$  представляет собою пространственный фурье-спектр комплексной амплитуды на расстоянии  $z$  от исходной плоскости. Эта функция согласно (5.8) и (5.9) имеет вид

$$g(u, z) = g(u, 0) \exp\left[-i \frac{z}{2k} u^2\right]. \quad (5.10)$$

Формула (5.10) справедлива для регулярного поля. Ее следует трактовать как преобразование спектра комплексной амплитуды свободным пространством при распространении регулярной волны.

А теперь сравним формулу (5.10) для регулярного поля с формулой (5.7) для нерегулярного, вполне случайного поля. Обе

формулы в точности совпадают. Это означает, что функция когерентности обладает свойствами регулярного поля. На основании (5.6) она связана с входным изображением, что дает возможность построить на функции когерентности голографию. Функция когерентности эквивалентна пространственному спектру комплексной амплитуды.

Таким образом, некогерентная голография, или голография в некогерентном свете, строится на тех же физических основах и на тех же формулах, что и голография в когерентном свете.

Представим, что нам удалось приготовить фотопластинку, прозрачность которой воспроизводит функцию когерентности. Эта пластинка будет отличаться от соответствующей записи значений комплексных амплитуд лишь тем, что является преобразованием Фурье. Это отличие несущественно в процессе восстановления изображения. Однако изображения, восстанавливаемые с некогерентных голограмм в когерентном свете, будут давать изображения интенсивностей источников, излучающих некогерентный свет, некоррелированный в соседних точках.

Кроме того, фотопластинку, имеющую такое же пропускание, как комплексная амплитуда, приготовить можно только в воображении, а в действительности надо прибегнуть к ухищрениям Габора и Лейта. Что касается функции когерентности, то ее можно иметь во всей ее комплексной красоте, не прибегая ни к методам Габора, ни к методам Лейта, хотя они тоже применимы в случае функции когерентности. Некогерентная голография осуществляется практически с помощью методов, называемых апертурным синтезом.

### 5.3. Некогерентный апертурный синтез

Немного истории. Развитие радиолокации в период второй мировой войны привело к зарождению и развитию радиоастрономии. В первые послевоенные годы американские инженеры разработали и создали крупнейший в мире радиотелескоп с полноповоротной антенной, зеркало которой было диаметром почти 300 метров. На зеркале такой площади можно проводить Олимпийские игры, как на настоящем стадионе. Было ясно, что такой огромный радиотелескоп не в состоянии даже повторить, не говоря о том, чтобы превзойти, ни одна страна, кроме Соединенных Штатов Америки.

Велико же было удивление ученых, когда всего через несколько лет в Англии Райлом [27] с помощью двух медленно раздвигаю-

щихся радиотелескопов с большими, но не выдающимися своими размерами антеннами был создан радиоастрономический инструмент, который по своей разрешающей способности был эквивалентен полноповоротному зеркалу с диаметром в целую милю. Это было сделано на основе использования некогерентного апертурного синтеза. Апертура инструмента синтезировалась по мере раздвижения антенн. Максимальная эквивалентная апертура определялась расстоянием, на которое удавалось раздвинуть две совместно работающие антенны.

Здесь следует оправдать уже проделанные нами ранее действия. Имеется в виду ограничение всего одним значением переменной  $x$  в формуле (5.5). Покажем, что это допустимо и ничего не изменится, если вместо точки  $x = 0$  взять любую иную. Чтобы убедиться в этом, достаточно (5.5) записать в виде

$$\Gamma\left(x, \frac{zu}{k}, z\right) \exp\left(-iux + \frac{izu^2}{2k}\right) = \frac{1}{i\lambda} \int W^2(\zeta) \exp(iu\zeta) d\zeta. \quad (5.11)$$

Обратим внимание, что левая часть (5.11) не зависит ни от  $x$ , ни от  $z$ . На этом основании можно написать

$$\Phi\left(\frac{zu}{k}\right) = \frac{1}{i\lambda} \int W^2(\zeta) \exp(iu\zeta) d\zeta. \quad (5.12)$$

Введенная нами функция  $\Phi(\zeta)$  однозначно определяется по функции когерентности  $\Gamma(x, \zeta, z)$  посредством равенства, следующего из сравнения (5.11) и (5.12):

$$\Phi\left(\frac{uz}{k}\right) = \Gamma\left(x, \frac{uz}{k}, z\right) \exp\left(\frac{izu^2}{2k} - iux\right). \quad (5.13)$$

Таким образом, для определения пространственного спектра изображения достаточно найти функцию когерентности только для разностей координат точек, приняв за точку отсчета любую из них.

Из соотношения (4.4), определяющего функцию когерентности, следует, что для нахождения значения  $\Gamma$  достаточно перемножить модули комплексных амплитуд сигналов в двух точках и определить разность фаз сигналов в этих точках. Усредняя эти значения в течение длительного времени, получим значение функции когерентности между этими точками.

Выражение (5.12) можно использовать для решения обратных задач, т. е. для нахождения изображений по значениям функции когерентности, измеряемым в точке приема случайных полей.

Если бы была возможность измерять функции когерентности  $\Gamma$  а также и функции  $\Phi$  для всех значений аргументов, то решение



обратных задач свелось бы просто к обращению интеграла Фурье (5.12). Однако на практике никогда нельзя измерить функцию  $\Gamma$ , а с ней и функцию  $\Phi$ , для всех значений разности координат. Все измерения можно провести только на ограниченном участке, линейные размеры которого называются апертурой. Эту апертуру можно расширять, добавляя измерения, но никогда нельзя сделать ее бесконечной, как того требует интеграл Фурье.

Посмотрим, каковы возможности решения обратных задач при ограниченной апертуре, т. е. при измерении функции  $\Phi$  в ограниченных точках. Как мы уже делали, конечность апертуры учтем введением специального множителя  $q(\zeta)$ . Введем этот множитель в правую часть уравнения (5.12), но, чтобы оно оставалось точным, надо изменить и левую часть этого уравнения. В нее подставим вместо истинного изображения  $W^2(\zeta)$  его образ  $W_R^2(\zeta)$ , который соответствует конечной апертуре — ее форма целиком определяется функцией  $q(\zeta)$ . Тогда получим

$$\Phi\left(\frac{zu}{k}\right) q\left(\frac{zu}{k}\right) = \int W_R^2(\zeta) \exp(iu\zeta) d\zeta. \quad (5.14)$$

Обращая этот интеграл Фурье, запишем следующий алгоритм получения образа входного изображения:

$$W_R^2(x) = \frac{1}{2\pi} \int \Phi\left(\frac{zu}{k}\right) q\left(\frac{zu}{k}\right) \exp(-iux) du. \quad (5.15)$$

Подставив в (5.15) значение функции когерентности из (5.13), найдем полный алгоритм, по которому можно получить образ входного изображения, располагая значениями функции когерентности на конечной апертуре. Такая процедура получения образа изображения называется *апертурный синтез*. Эта процедура не требует одномоментного измерения полей, для ее осуществления достаточно измерения функции когерентности в двух точках поля для двух значений разности координат. Чтобы иметь возможность производить такие измерения, достаточно располагать всего двумя точками приема излучения и перемещать эти точки приема в разные места, синтезируя тем самым апертуру. Такой способ синтеза апертуры по случайным источникам возник почти одновременно с синтезом апертуры в регулярных полях, где использовалось перемещение всего одного приемника поля.

Сравнивая выражения (5.15) с обращением интеграла Фурье (5.12), можно сделать вывод, что отличие образа изображения  $W_R^2(x)$  от самого изображения легко представить функциональной схемой, приведенной на рис. 32.

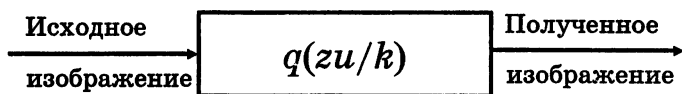


Рис. 32. Функциональная схема связи исходного и полученного изображений при апертурном синтезе

Действительно, функция  $q(\zeta)$  играет роль частотной характеристики и фильтра в выражении (5.15). Образ изображения отличается от самого изображения тем, что прошел через фильтр с частотной характеристикой, определяемой функцией  $q(\zeta)$ .

Как было показано выше, эта связь изображения и его образа, получаемого в результате решения обратной задачи, такая же, как и в регулярных полях. Поэтому апертурный синтез имеет еще одно название, очень для него подходящее – некогерентная голография. Математические соотношения, связывающие голограмму и изображение, которое можно с нее получить, как мы видели выше, в принципе совпадают с соотношениями, связывающими функцию когерентности и получаемые с нее изображения. Апертура  $q(\zeta)$  может быть любой формы, функция когерентности, в частности, может в одном измерении измеряться, а в другом нет.

Вращение Земли вокруг своей оси поспособствовало успеху апертурного синтеза Райла, но это возможно только при радиоастрономических наблюдениях. При каждом суточном повороте Земли антенны Райла, установленные неподвижно, совершали в системе координат, связанной с неподвижными звездами, один оборот относительно друг друга. При этом синтезировалась апертура в форме кольца, имевшего диаметр, равный расстоянию между антеннами в системе координат, связанной с неподвижными звездами. К следующим суткам расстояние между антеннами увеличивалось и синтезировалось кольцо иного радиуса. Так и была синтезирована апертура в виде заполненного круга, имеющего в диаметре целую милю.

Сканировать этим кругом не надо, так как стоит только заполнить всю апертуру, определив на ней значения функции когерентности, как появляется возможность определения углового спектра некогерентных между собой источников излучения.

Такой способ наблюдения с последовательным, точка за точкой, определением значений функции когерентности предполагает, что распределение источников, представляющих собой наблюдаемое изображение, со временем не изменяется. В радиоастрономии такое

предположение вполне оправданно и не влечет за собой существенных ограничений, хотя на самом деле они есть.

Дело в том, что, хотя колебания считаются случайными, они должны быть в определенной степени монохроматичными. Допустимая ширина полосы определяется величиной апертуры. Для радиоастрономии ширина полосы принимаемого сигнала является важнейшим параметром, так как она определяет помехоустойчивость приема [28]. Райл принимал меры к расширению полосы, однако в способе обработки сигналов, используемых при апертурном синтезе в радиоастрономии, есть принципиальное ограничение ширины полосы, которое преодолеть не удалось. В результате методом апертурного синтеза можно наблюдать только весьма интенсивные радиозвезды и в пределах узкого углового сектора. Сужение углового сектора обзора связано тоже с возможностью расширения полосы. Чем меньше сектор обзора, тем полоса может быть больше и тем больше могут быть размеры синтезируемой апертуры.

Пока область применения некогерентной голографии в виде апертурного синтеза ограничивается только радиоастрономией. В оптике нет устройств, обеспечивающих измерение значений функции когерентности. Это можно сделать только непрямыми путями, наблюдая интерференционную картину. Таким путем удается наблюдать пространственный спектр свечения электронно-лучевой трубки [29], что позволяет создать оптический двумерный спектроанализатор с удобным электронно-лучевым вводом сигнала в реальном времени [29].

Однако, как выяснилось сравнительно недавно, в живой природе апертурный синтез очень широко применяется. Этому вопросу мы посвящаем специальный раздел.

## **ОСНОВЫ СОВЕРШЕНСТВОВАНИЯ МЕТОДОВ ФОРМИРОВАНИЯ ИЗОБРАЖЕНИЙ**

### **6. АПЕРТУРНЫЙ СИНТЕЗ И БИНАУРАЛЬНЫЙ СЛУХ**

Этот раздел посвящен формированию изображений, создаваемых живыми организмами с помощью волновых полей. Речь будет идти только об акустике, так как глаз – это настоящая оптическая система, с которой все ясно, а с акустикой дело обстоит сложнее. Не будем касаться акустики летучих мышей, обладающих возможностью активной локации, способных строить изображение в регулярных полях и использовать апертурный синтез. Рассмотрим формирование акустических изображений у человека. Под таким формированием будем понимать определение направлений на каждый выделяемый человеком звуковой сигнал.

Устройство бинаурального слухового аппарата и его функционирование покажем путем выбора наиболее подходящего механизма формирования изображений из тех, которые известны науке. Для такого выбора надо подвести некоторые итоги сказанному выше, сформулировав характерные черты рассмотренных путей формирования изображений.

#### **6.1. Характерные особенности различных методов формирования изображений**

1. В настоящее время известны всего два алгоритма, позволяющие получать изображения, – это алгоритм, основанный на использовании направленности антенны с заполненной аперту-

рой, и второй алгоритм – апертурный синтез, основанный на измерении функции когерентности волнового поля.

2. Алгоритм получения изображений, основанный на использовании антенны с заполненной апертурой, не зависит от конкретной статистики излучаемых сигналов. Источники излучений, координаты которых определяются антенной, могут быть независимы полностью или частично коррелированы между собой. В любом из этих случаев их координаты будут определены с одинаковой точностью. Существенно то, что появление статистических связей или их изменение не приводят к появлению виртуальных источников излучения, т. е. таких изображений, которым не соответствуют реальные источники.

3. Алгоритм апертурного синтеза, основанный на свойствах функции когерентности, чувствителен к статистической модели источника излучения. Этот алгоритм дает правильные результаты только при одной статистической модели источника, которая описывается соотношением (5.6). Применение этого метода обусловлено полной некоррелированностью пеленгуемых источников. Если между источниками есть корреляционные связи, то алгоритм апертурного синтеза приводит к формированию в изображении дополнительных виртуальных источников, которым никакие реальные источники не соответствуют. Интенсивность и расположение виртуальных источников определяются корреляционными связями между колебаниями, излучаемыми реальными источниками.

4. Чтобы синтезировать апертуру, необходимо изменить пространственную частоту  $u$  в выражении (5.12). Для этого в классическом апертурном синтезе Райла меняют аргумент функции  $\Phi$  путем изменения разности координат точек  $\zeta$ . Очень существенно, что изменение расстояния между точками приема излучения не является единственным способом изменения пространственной частоты в (5.12). В самом деле, в (5.12) разность координат точек связана с пространственной частотой через аргумент функции  $\Phi$  следующим образом:

$$\zeta = \frac{uz}{k}. \quad (6.1)$$

Дифференцируя (6.1), получаем

$$\Delta u = \frac{k}{z} \Delta \zeta + \frac{\zeta}{z} \frac{\Delta \omega}{c}, \quad (6.2)$$

откуда следует, что изменения  $u$  можно добиться как изменением  $\zeta$  при постоянной  $\omega$ , так и изменением  $\omega$  при постоянной  $\zeta$ .

Изменения  $\omega$  и  $\zeta$ , приводящие к одинаковым изменениям  $u$ , связаны следующим соотношением:

$$\Delta\zeta = \frac{\Delta\omega}{\omega} \zeta. \quad (6.3)$$

Из (6.3) видно, что апертуру можно синтезировать путем изменения частоты (или работая на разных частотах одновременно).

5. На основании предыдущего замечания можно сформулировать еще одно, которое подчеркивает разницу между апертурным синтезом в регулярных полях и апертурным синтезом в случайных полях, или между когерентной и некогерентной голографией. Эта разница заключается в следующем. Апертурный синтез в регулярных полях, или голография, может быть осуществлен только в том случае, если излучается достаточно монохроматический сигнал. Нельзя придать сигналу нужной монохроматичности при его приеме. В когерентном апертурном синтезе принятый сигнал фильтровать нельзя. Такая фильтрация приводит к тому, что изображение пропадает.

В некогерентной голографии можно работать по сигналам, имеющим сплошной спектр, но нужной степени монохроматичности, кстати, существенно меньшей, чем в когерентной голографии, можно добиться при приеме сигнала путем его фильтрации. Поэтому можно принимать сигналы в широком спектре частот и фильтровать их так, чтобы обрабатывать на каждой частоте отдельно. Это дает возможность при работе по широкополосному излучению осуществлять апертурный синтез в частотной области одномоментно.

## 6.2. Физическая модель бинаурального слуха

Такая модель может быть основана на известных физических принципах, которые перечислены в разд. 6.1, или на каком-то еще не известном науке принципе. Последнюю возможность исключить нельзя, но принимать ее всерьез можно только в том случае, когда это подтверждает достоверный эксперимент. На сегодня, по-видимому, ситуация такова, что одна из возможностей, указанных в разд. 6.1, может быть положена в основу построения физической модели бинаурального слуха. Эта модель объясняет многие факты, но еще не все, так что и для дальнейшего совершенствования модели есть основания.

Физическую модель бинаурального слуха можно предложить на основе использования апертурного синтеза в случайных полях.

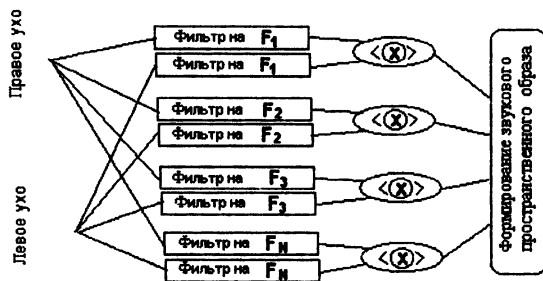


Рис. 33. Функциональная схема бинаурального слуха

Ее функциональная схема изображена на рис. 33. Она включает частотные фильтры, их наличие у человека согласуется с представлениями, которыми располагает физиологическая теория слуха. Такие фильтры получаются вследствие особого устройства внутреннего уха, когда разные частоты локализуются в различных местах и воспринимаются разными нервными волокнами. Такие фильтры работают независимо – как в правом ухе, так и в левом. После фильтрации сигналы одинаковых частот, принимаемых обоими ушами, перемножаются и усредняются, образуя функцию когерентности, зависящую от частоты.

Однако для получения слуховых изображений функции когерентности в виде функции частоты недостаточно. На каком же основании мы считаем эту схему полностью законченной функциональной схемой бинаурального слуха? Как и где полученные значения функции когерентности обрабатываются дальше? Ответы на эти вопросы, на наш взгляд, можно найти в особенностях поведения живых организмов, у которых индикаторами изображений являются непосредственно органы чувств, и в математических свойствах преобразования Фурье.

Располагая функцией когерентности, для получения слуховых образов следует согласно (6.3) и (5.12) выполнить следующее преобразование:

$$W^2(x) = \frac{\zeta}{cz2\pi} \int \Phi(\omega, \zeta) q(\omega) \exp\left(i\omega \frac{\zeta x}{cz}\right) d\omega, \quad (6.4)$$

где  $\zeta$  – расстояние между ушами слушателя, которое остается постоянным, так как синтез апертуры происходит благодаря изменению частоты приема сигнала;  $q(\omega)$  – пространственный фильтр изображения, эквивалентный соответствующему фильтру в антенне или оптической системе, образующийся благодаря конечности апертуры и фазовым и амплитудным искажениям.

Преобразование (6.4) представляет собой интеграл Фурье и является однозначным, что очень важно. Особенность биологической системы состоит в том, что индикатором вида итоговой функции изображения (6.4) служат ощущения расположения источника звука в определенном месте. Такое ощущение можно выработать столь же успешно не по самой функции  $W^2(x)$ , а по ее пространственному спектру, которым является функция когерентности  $\Phi(\omega, \zeta)$ . Короче говоря, в биологической системе может не производиться математическое преобразование (6.4) как некоторое математическое действие, эта процедура может быть осуществлена с помощью выработки соответствующих ощущений на основе функции когерентности. Если это принять, то предложенная модель бинаурального слуха является как полной, так и согласующейся с представлениями, выработанными физиологией слуха. Основными операциями, требуемыми нашей моделью, являются спектральное разложение входного сигнала, которое совершается улиткой внутреннего уха, перемножение сигналов одной частоты и их усреднение. На основе получившейся функции частоты органами чувств принимается решение о пространственном расположении слуховых образов.

Данная модель согласуется также с моделью слуха, предложенной ранее, в пятидесятых годах, Черри и Сайерсом [30]. Они предложили корреляционную теорию бинаурального слуха и свои рассуждения основывали на эффективности корреляционного анализа в радиолокации, считая, что природа вправе и должна была воспользоваться этими достижениями радиолокации. По теории Черри и Сайерса, бинауральный слух основан на измерении функции взаимной корреляции сигналов, принятых ушами.

Черри и Сайерс не ограничились высказыванием гипотезы о корреляционной природе слуха, а самым тщательным образом, весьма остроумно, проверили ее в опытах на людях. Несмотря на то что в то время не было абсолютно никаких физиологических оснований для этой гипотезы, Черри и Сайерс экспериментально подтвердили, что слушатель действительно "образует" функцию взаимной корреляции сигналов двух ушей.

Предлагаемую модель, изображенную на рис. 33, можно рассматривать как физиологическое основание справедливости модели Черри и Сайерса. Дело в том, что преобразование (6.4) математически эквивалентно функции взаимной корреляции сигналов двух ушей, если предположить, что с помощью набора фильтров (см. рис. 33) осуществляется преобразование Фурье. Таким образом, предлагаемая модель согласуется как с физиологией слуха, так и с



корреляционной теорией Черри и Сайерса. Однако фильтры внутреннего уха не обладают одинаковой шириной частотной характеристики, поэтому не могут образовывать настоящий фурье-спектр. В силу этого должны быть некоторые отклонения от корреляционной модели Черри и Сайерса, и такие отклонения на самом деле были обнаружены [31].

Таким образом, есть основания полагать, что бинауральный слух реализует алгоритм апертурного синтеза в частотной области, математически описывается формулой (6.4) и имеет функциональную схему рис. 33.

### 6.3. Естественная и художественная стереофония

В 30-х годах нашего века возникло радиовещание, которое дало возможность слушать речь и музыку, находясь далеко от того места, где располагался источник звуков. Бурный успех радиотехники вдохновил радиоинженеров на поиск таких методов для воспроизведения транслируемых звуков, которые бы не только сохраняли их качество, но и позволяли бы улучшить его. Руководящей идеей улучшения качества транслируемой программы было признанное музыкантами определяющее значение тембра звучания в красоте звука. Тембр представляет собой соотношение гармонических составляющих звука. Радиоинженеры сконструировали устройства, позволяющие в широких пределах изменять тембр транслируемых по радио программ (высококачественной звукозаписи тогда еще не было), считая это достаточным техническим средством существенного улучшения красоты звучания любого звука. Для дальнейших практических шагов в этом направлении радиоинженеры пригласили к себе известных музыкантов. Задачей этих музыкантов было убедиться в возможности выбора тембра, делающего транслируемый звук куда приятнее исходного. К удивлению всех участников этой работы, как инженеров, так и музыкантов, никакие изменения тембра звучания не позволяли улучшать качество звучания транслируемой программы по сравнению с исходной.

Известный дирижер Леопольд Стокковский высказал мысль, что чистоте опыта мешает одно обстоятельство: транслируемые звуки кажутся исходящими из одной точки, в то время как естественное звучание распределено в пространстве, поэтому даже при наличии занавеса очень легко отличить транслируемый звук

от того, который рождается прямо за занавесом. Это могло мешать правильному восприятию изменений тембра.

Высказанную гипотезу признали разумной и решили ее проверить экспериментально. Идея эксперимента заключалась в совершенствовании транслируемого сигнала настолько, чтобы сохранялась исходная информация о расположении источников звука. Это решено было сделать путем увеличения каналов трансляции, позволяющих передать сигнал ряда микрофонов, образующих целый "занавес" из микрофонов. Такой сигнал должен воспроизводиться также занавесом, состоящим из громкоговорителей, каждый из которых соединен трансляционной линией с соответствующей точкой занавеса, состоящего из микрофонов. Занавес из громкоговорителей устанавливался в зале для прослушивания транслируемых сигналов, а занавес из микрофонов устанавливался в том зале, откуда ведется трансляция. В зале для прослушивания располагался настоящий оркестр, который привлекался для сравнения естественного и транслируемого сигналов.

Такой опыт действительно был проведен в Америке в 1935 году. Опыт кончился невероятным успехом, просто триумфом. После этого он был повторен во многих столицах мира, в том числе и в Москве в том же 1935 году. Всюду эти опыты проходили с неизменным успехом. Слушатели, сидящие в зале, — известные музыканты — не могли отличить транслируемый сигнал от натурального звучания. Кто-то ходил по сцене и забивал гвозди, а слушатели не могли определить, когда эти звуки были натуральными, а когда транслируемыми.

Осталось только сказать, что по причине недостаточно развитой техники того времени занавесы состояли всего из трех громкоговорителей и трех микрофонов. Однако, так как это дало ожидаемый эффект, то никто не помышлял об увеличении количества каналов звуковоспроизведения. Наоборот, попробовали сократить количество каналов с трех до двух.

С этого времени возникла и стала развиваться *стереофония*, т. е. звуковоспроизведение, обладающее двумя каналами. Стерефонические запись и воспроизведение, при которых записывается (или транслируется) естественное, натуральное звучание с помощью двух пространственно разнесенных микрофонов, будем называть *естественной стереофонией*.

Современную стереофонию, которая основана на специальных эффектах, будем называть художественной стереофонией, так как основное их назначение — это добиться максимального художественного эффекта. Подбор эффектов и управление ими в процессе

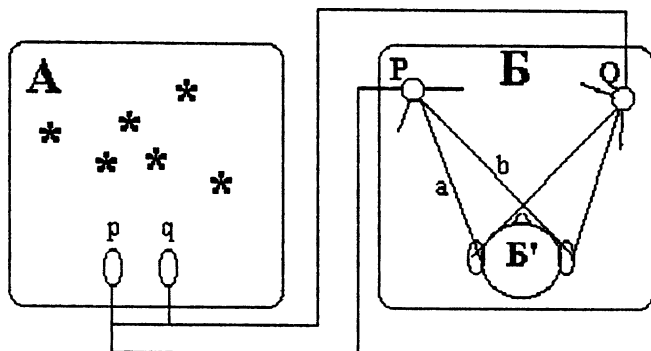


Рис. 34. Схема естественной стереофонии

записи производит человек особой квалификации – режиссер стереофонических программ.

Эта техника создавалась и развивалась исключительно на экспериментальной основе. Все, что было сделано, основано на методе проб и ошибок. Ниже попробуем подвести под эту работу некоторую теоретическую базу, основанную на изложенной выше модели бинаурального слуха.

#### 6.4. Физические основы естественной стереофонии

Рассмотрим схему естественной стереофонии в виде, изображенном на рис. 34, где показаны два помещения – А и Б. В помещении А находятся источники звука, помеченные звездочками, и установлены два микрофона, принимающие сигналы с комплексными амплитудами  $P$  и  $Q$ .

Примем, что сигналы, принятые в А микрофонами без каких-либо искажений и помех, достигают громкоговорителей и излучаются ими в Б. Громкоговорители излучают сигналы с комплексными амплитудами  $P$  и  $Q$ . В помещении Б находится слушатель в точке, равноудаленной от обоих громкоговорителей. Расстояния от громкоговорителей до ушей слушателя обозначены на рис. 22 и равны  $a$  и  $b$ .

Если пренебречь рассеянными и отраженными сигналами и считать, что слушатель воспринимает только прямые сигналы громкоговорителей, то комплексные амплитуды сигналов, поступа-

ющих на правое и левое ухо слушателя, на каждой частоте будут следующими:

$$PS_i = P \exp(ika) + \gamma Q \exp(ikb),$$

$$PS_p = \gamma P \exp(ikb) + Q \exp(ika). \quad (6.5)$$

Здесь учтено изменение фазы сигналов из-за расстояния, пройденного волной от громкоговорителя до ушей слушателя, и введен некоторый комплексный множитель  $\gamma$ , характеризующий дифракцию звука вокруг головы слушателя. Этот коэффициент  $\gamma$  индивидуален для каждого слушателя. Эксперимент убеждает, что голову нельзя аппроксимировать абсолютно жестким шаром, так как голова вносит более сильное ослабление в принимаемый сигнал, чем вносила бы дифракция на абсолютно жестком гладком шаре.

Согласно модели бинаурального слухового восприятия, принятой нами, слушатель из сигналов (6.5) "образует" корреляционную функцию, которая будет иметь вид

$$\Phi(\omega, \zeta) = \gamma \langle P\bar{P} \rangle \exp[ik(a-b)] + \gamma \langle Q\bar{Q} \rangle \exp[ik(b-a)] + \langle P\bar{Q} \rangle + \gamma \bar{\gamma} \langle \bar{P}Q \rangle. \quad (6.6)$$

Выражение (6.6) подвергается обработке по формуле (6.4), которая является линейной, и ее можно применить отдельно к каждому слагаемому (6.6). Воспользовавшись этим, выясним физический смысл каждого слагаемого (6.6) и результат его обработки по (6.4).

Для выяснения, чему соответствует первое слагаемое (6.6), выключим мысленно громкоговоритель, излучающий сигнал с комплексной амплитудой  $Q$ . Тогда в (6.6) останется единственный член, а именно первое слагаемое. Но в этом случае работает всего один громкоговоритель, излучающий сигнал  $P$ . У слушателя при этом должно возникнуть ощущение, что он слышит, как работает этот громкоговоритель. Следовательно, учитывая линейность обработки, можно утверждать, что первый член в (6.6) будет вызывать у слушателя слуховой образ громкоговорителя, излучающего сигнал  $P$ . Аналогичным образом можно показать, выключая мысленно сигнал  $P$ , что второй член (6.6) вызывает слуховой образ громкоговорителя, излучающего сигнал  $Q$ .

Рассмотрим теперь третий член в (6.6), который нам наиболее интересен, и вот почему. Если расположить слушателя в помещении А так, что его уши будут на местах установки микрофонов, то функция когерентности этого слушателя будет в точности совпадать

с третьим членом (6.6). Следовательно, этот член общего сигнала вызовет у слушателя слуховые образы тех звучащих источников, которые находятся в помещении А. Эти образы третий член один к одному "переносит" из помещения А в помещение Б. В помещении Б, где находится слушатель, источников звука нет, но этот третий член будет вызывать у слушателя слуховые образы, как если бы слушатель был в помещении А. Для помещения Б эти слуховые образы назовем виртуальными. Наконец, рассмотрим последний член в (6.6). Он также будет вызывать виртуальные образы, но они будут представлять собой зеркальное отражение истинных образов, которые имеются в помещении А. Однако эти зеркально отраженные виртуальные образы будут существенно ослаблены по интенсивности из-за наличия множителя  $|\gamma|^2$ . Согласно измерениям этот множитель может внести ослабление порядка 10 дБ. Поэтому в дальнейшем не будем принимать во внимание этот член. Вообще говоря, его наличие вносит несколько большее разнообразие в картину виртуальных источников, но на появление самих источников не влияет.

Подведем итог. Слушатель в помещении Б согласно нашей модели должен ощущать слуховые образы самих громкоговорителей и виртуальных источников, которые в точности соответствуют расположению источников в помещении А и ослабленным по интенсивности, зеркально расположенным источникам.

Однако все виртуальные источники по интенсивности одного порядка. Поэтому каждый отдельный виртуальный источник, если их много, будет звучать много слабее, чем слуховой образ громкоговорителя, создаваемый совокупным излучением всех виртуальных источников. Поэтому выделить в (6.6) виртуальные источники и даже обнаружить, что они имеются, сложно и практически доступно только тренированному слуху очень хорошего музыканта, а не обычного слушателя. Обычный слушатель, не будучи тренированным, не заметит виртуальных источников и будет ощущать только наличие двух громкоговорителей вместо одного при монофоническом звуковоспроизведении.

Это можно пояснить иначе. Два последних члена в (6.6), создающие виртуальные источники в помещении Б, представляют собой функцию взаимной корреляции сигналов, излучаемых двумя громкоговорителями. Если эта функция равна нулю, то эти члены в (6.6) пропадают, и слушатель слышит только реальные источники и ничего виртуального нет, т. е. слушатель в этом случае будет слышать только сами громкоговорители. При наличии корреляционных связей виртуальные источники появятся, но если корреляция мала, то эти источники будут очень слабыми. При большом

количестве источников корреляция будет весьма слабой, поэтому в условиях естественной стереофонии нетренированный массовый слушатель виртуальных источников не заметит и будет считать себя обманутым.

В условиях естественной стереофонии не создашь особо эффектных программ, рассчитанных на рекламу. Поэтому естественная стереофония не получила никакого распространения, несмотря на то что именно ей обязана своим рождением художественная стереофония. Успех первых опытов, в которых была использована естественная стереофония, объясняется тем, что первыми слушателями были музыканты, обладающие хорошо тренированным слухом, что позволило им выделять виртуальные источники звука и восхищаться ими. Самым эффектным моментом того эксперимента был звук забиваемых гвоздей. Этот виртуальный источник был в каждый момент один и поэтому уверенно прослушивался.

### 6.5. Художественная стереофония

Художественная стереофония основывается на результатах следующего опыта. Если взять два идентичных источника звука, которые излучают строго коррелированные сигналы (одни и те же сигналы), поместить их на место громкоговорителей в помещение Б (см. рис. 34), то у слушателя возникнет ощущение, что он слышит звук, который исходит из одной точки, расположенной на линии, соединяющей эти громкоговорители. Конкретное место расположения этого кажущегося источника звука определяется соотношением интенсивности сигналов, излучаемых источниками. При равных интенсивностях (когда слушатель воспринимает эти интенсивности как равные) источник звука находится прямо посередине линии, соединяющей громкоговорители, а если интенсивности не равны, то кажущийся источник смещается в сторону наиболее интенсивного громкоговорителя, сливаясь с ним при выключении. Этот опыт ставился многократно. Тщательно фиксировались соотношения интенсивностей источников, которые приводили к тому или иному смещению источника в сторону от центра. По этим данным составлены подробные таблицы.

Для целей художественной стереофонии результат этого опыта используется следующим образом.

В целях рекламы возможностей стереофонии записывается, к примеру, шум автомобиля всего одним микрофоном. Этот звук

переписывается на две дорожки магнитофона или пластинки так, чтобы интенсивность звука постепенно изменялась, чтобы звук в одном канале ослабевал, а в другом в это время возрастал. Получившаяся двухканальная запись объявляется стереофонической. Перед демонстрацией этой записи диктор объявляет, что микрофон установлен на улице и слушатель сейчас услышит, как автомобиль проедет мимо него по улице. Последующая демонстрация записи убеждает слушателя, что он слышит подлинное движение автомобиля, проезжающего мимо него по улице, так как звук, постепенно переходя от одного громкоговорителя к другому, вызывает ощущение перемещающегося источника. Этот эффект перемещающегося источника ощущает каждый слушатель, а не только тот, который специально для этого тренировался. Точно таким же образом представляется отрывок театральной постановки: разговор двух актеров записывается для каждого на свой отдельный канал, затем переписывается на два канала так, что голос каждого артиста попадает в оба канала в различных пропорциях. Это создает впечатление, что артисты, беседуя, перемещаются по сцене. Звукорежиссер, оперируя интенсивностями сигналов, заставляет артистов звучать из любого места между двумя громкоговорителями. Подобная реклама срабатывает безотказно, демонстрируя возможности стереофонического воспроизведения по сравнению с одноканальным монофоническим, которое такими возможностями не обладает.

Запись музыкальных программ производится по тому же принципу, но она значительно сложнее. Каждый инструмент или несколько инструментов оркестра записываются на свою дорожку в такой пропорции, чтобы каждый из них попал на свое, отведенное ему звукорежиссером место. При этом необходимо сохранить или получить то соотношение интенсивностей между различными инструментами, которое обеспечивает художественное воспроизведение, исходя из требований музыкальности. Эта работа требует высочайшей квалификации. Дирижер оркестра во время записи стереофонической программы не знает о том, как она прозвучит в интерпретации звукорежиссера.

Такая запись обеспечивает эффект локализации источников звука и позволяет в соответствии с первоначально высказанными гипотезами улучшить и звуковоспроизведение. Считается, что стереофоническая программа звучит лучше, чем монофоническая благодаря эффекту локализации источников звука. Ниже мы дадим иное объяснение улучшению качества звука.

## 6.6. Физические основы художественной стереофонии

Чтобы обосновать возможности, которые имеет и демонстрирует художественная стереофония, необходимо объяснить тот опыт, на котором она основана.

Для этого обратимся опять к соотношению (6.6) и к рис. 34. Начнем с рисунка. Чтобы сигнал  $P$  был идентичен сигналу  $Q$ , в помещении  $A$  должен находиться один источник звука и располагаться на одинаковом расстоянии от обоих микрофонов, на линии, которая пересекает соединяющую оба микрофона линию, в ее центре, направление – под прямым углом. Тогда из обоих громкоговорителей будет излучаться один и тот же сигнал. Этого можно добиться соединением обоих каналов друг с другом. Посмотрим, что будет представлять собою соотношение (6.6), если в него подставить одинаковые сигналы, т. е. если  $P = Q$ . Первые два члена уравнения, независимо ни от чего, должны давать образ громкоговорителей. Третий и четвертый члены должны давать образы виртуальных источников. Теперь виртуальный источник всего один, корреляция между  $P$  и  $Q$  полная, поэтому сигнал  $PQ$  по интенсивности будет превосходить сигналы, которые излучаются громкоговорителями. Таким образом, должно наблюдаться три источника звука: один, наиболее сильный, в центре и два по бокам. Боковые источники совпадают с самими громкоговорителями. Кажется бы, имеет место противоречие между тем, что наблюдается при опыте, на котором основывается вся стереофония. Это противоречие служило серьезным препятствием признания корреляционной теории Черри и Сайерса, так как из нее тоже получалось, что в этом опыте должно быть непременно три источника, а не один.

Это противоречие устранимо в рамках корреляционной теории, если учесть эффективную ширину получающихся звуковых образов и их интенсивности. Подробно этот вопрос рассмотрен в [31], где описаны результаты прямого эксперимента.

Опытным путем установлено, что эффекты, на которых основана художественная стереофония, наблюдаются в частотном диапазоне, ограниченном как сверху, так и снизу. Кроме этого, эффекты художественной стереофонии чувствительны к расположению громкоговорителей и слушателя. Есть оптимальное расположение громкоговорителя и слушателя, при котором обеспечивается максимальный эффект. По этому поводу имеются рекомендации, выработанные экспериментальным путем. Если их проанализировать, то получится следующая картина.



При оптимальной полосе частот и на оптимальном для слушателя месте все три источника звука, которые должен ощущать слушатель, согласно (6.6) сливаются в один из-за недостаточной разрешающей способности. Слившись в один, они образуют максимум в центре, где находится наиболее интенсивный звуковой образ. При изменении интенсивности сигналов происходит следующее. Пусть интенсивность одного из сигналов уменьшится в  $N$  раз. Тогда интенсивность крайнего изменится в  $N^2$  раз, а интенсивность среднего только в  $N$  раз, и максимум в результате этого переместится в сторону более интенсивно звучащего громкоговорителя. Количественные расчеты убеждают, что перемещение максимума происходит в полном соответствии с эмпирически установленными закономерностями [31]. Пользуясь вышеприведенными умозаключениями, можно обосновать эмпирические закономерности относительно лучшего места для слушателя и оптимального диапазона частот.

То, что существуют определенные точки или области, где только и можно наблюдать эффект перемещения источника, подтверждает следующее. В широкоформатных кинотеатрах используется эффект перемещения источника звука вдоль экрана. Для этого вроде бы было бы достаточно поставить всего два громкоговорителя по краям экрана и иметь всего два канала записи с изменяющейся интенсивностью. Однако, несмотря на дешевизну такого способа по сравнению с фактически применяемым, им не пользуются. Сзади экрана устанавливают целый ряд самостоятельных громкоговорителей, которые включаются в соответствии с программой. Это делается потому, что для всего зала широкоэкранный кинотеатр невозможно создать условия, при которых работает художественная стереофония.

Кроме того, был поставлен прямой опыт, в котором специально использовался широкополосный источник звука, позволяющий создать необходимую величину апертуры в частотной области. При этом в полном соответствии с ожиданиями наблюдаются все три источника, предсказываемые соотношением (6.6), которые не перемещаются при изменении интенсивности звука [31].

С помощью художественной стереофонии невозможно расставить по местам много источников звука. Они непременно сольются в единый звуковой образ. Возникает вопрос: если эффект локализации в художественной стереофонии ограничен, то чем объяснить тот эмоциональный и художественный эффект в смысле лучшего, более полного восприятия музыки? Ответ на этот вопрос будет дан ниже.

## 6.7. Бинауральная голография

Термином "бинауральная голография" обозначается такая запись, при которой слушатель оказывается в исходном звуковом поле, т. е. воспринимает предъявляемое ему волновое поле как исходное.

Этот эффект создается при естественной стереофонии, но в данном случае бинауральную голограмму надо рассматривать, а точнее, прослушивать на фоне сильной помехи, которую создают громкоговорители. Первую настоящую бинауральную голограмму предложил и осуществил Шредер [32]. Такую голограмму можно приготовить для одного слушателя, согласованную с его формой и материалом (акустическим) его головы. Эту звуковую голограмму надо прослушивать в заглушенной комнате и в одной точке, находясь при этом в одной позе.

Прежде чем рассказать, как такая бинауральная голограмма может быть получена и на чем основано ее получение и воспроизведение, поясним, для чего она может быть использована (и фактически была использована). Существует целая наука, которая называется архитектурной акустикой. У нее нет твердых физических основ, и бинауральная голография в том виде, в котором предложил и осуществил ее Шредер, предназначена для создания физических основ архитектурной акустики. Архитектурная акустика имеет дело с акустическими характеристиками помещений, такими как время реверберации звука на разных частотах, корреляционные характеристики звукового поля в различных точках помещения, характеристики ослабления звука между различными точками и др. Эти характеристики можно измерять в готовых помещениях; можно выработать определенные рекомендации, как их достичь при конструировании зданий, как их регулировать в процессе строительства и по его окончании. Неясно только одно, какими должны быть эти характеристики, чтобы люди, придя в зал и прослушав в нем музыку или речь, сказали по окончании: "В этом зале великолепная акустика". Шредер высказал идею о том, что решить эту задачу можно, для этой цели и понадобилась его бинауральная голография.

В акустике для решения многих задач принято проводить экспертные оценки. Таким путем развивалась и художественная стереофония. Все важные вопросы, такие как требуемая полоса частот, качество громкоговорителей, размещение слушателей и другие, решались не с помощью теоретических моделей, а с помощью опытов, в которых участвовали эксперты.

В связи с этим у Шредера возникла идея сравнить с помощью экспертов лучшие концертные залы Европы, выбрать среди них лучший и те, что похуже, а потом, измерив объективные акустические характеристики каждого зала, сопоставить их с его качеством. Такой анализ, по мнению Шредера (и трудно с ним не согласиться), должен был помочь выбору тех значений параметров, к которым следует стремиться, конструируя или ремонтируя тот или иной зал с целью улучшения в нем качества звучания музыкальных программ.

На пути реализации этой идеи стояла только одна явная трудность: как организовать работу экспертов так, чтобы они имели возможность объективно сравнивать различные залы между собой. До вынесения своего суждения они не должны были знать, какой зал они оценивают. Самое простое решение этой задачи – зашить экспертов в мешки и возить их в таком виде из одного города в другой, из зала в зал, чтобы они не имели представления о том, где они находятся, – явно не подошло. В неудобном положении и угнетенном состоянии эксперт уже ничего не сможет оценить. Кроме того, на переезд из города в город требуется время, которое тоже не способствует объективной оценке, так как свежее впечатление трудно сравнить со вчерашним или позавчерашним.

Бинауральная голография и явилась разумным решением этой задачи. Если записать естественное звучание в зале именно в том виде, в котором оно там происходит в действительности, то такие записи можно предъявлять экспертам для сравнения, чтобы были выполнены все требования объективности. В этом случае годится и такая система, которая обеспечивает нужное звучание всего для одного определенного человека и в одной точке, и в заглушенной камере. Такую возможность, в принципе, дает естественная стереофония. Третье слагаемое в (4.4) представляет как раз то звучание, которое требуется Шредеру. Но Шредер предложил другой путь – во-первых, он о существовании уравнения (6.6) не подозревал, а во-вторых, если бы он и знал о такой возможности, то вряд ли принял бы ее, так как очень велико влияние мешающих слагаемых. Шредер предложил и осуществил систему, в которой мешающие слагаемые могут быть подавлены.

Основную идею Шредера проще всего пояснить с помощью соотношений (6.5).

Рассмотрим схему естественной стереофонии, изображенную на рис. 34. Поместим между микрофонами и громкоговорителями специальный фильтр, который так изменяет излучаемые громкоговорителями сигналы, что к ушам слушателя подводятся не сигналы  $PS_1$  и  $PS_p$ , а сигналы  $P$  и  $Q$ , которые принимаются

микрофонами помещения А. При этом излучаться должны уже не сигналы  $P$  и  $Q$ , а другие, преобразованные фильтром сигналы, которые обозначим  $M$  и  $N$ .

На основании (6.5) получим следующие два уравнения для определения сигналов  $M$  и  $N$ :

$$\begin{aligned} P &= M \exp(ika) + \gamma N \exp(ikb), \\ Q &= \gamma M \exp(ikb) + N \exp(ika). \end{aligned} \quad (6.7)$$

Решая систему уравнений (6.7), находим для  $M$  и  $N$  следующие значения:

$$\begin{aligned} M \exp(ikb) &= \frac{\gamma P \exp[ik(b-a)] - Q}{\gamma^2 \exp[ik(b-a)] - \exp[ik(a-b)]}, \\ N \exp(ikb) &= \frac{P - \gamma Q \exp[ik(b-a)]}{\exp[ik(a-b)] - \gamma^2 \exp[ik(b-a)]}. \end{aligned} \quad (6.8)$$

В этих решениях учтено, что соответствующий фильтр можно реально сделать только в том случае, если учтено акустическое запаздывание сигналов, подводимых к ушам слушателя. Поэтому найдены значения, умноженные на экспоненциальные множители. Только такие значения являются физически реализуемыми.

Шредер разработал схему такого фильтра, который осуществляет преобразования сигналов, указанные в (6.8). Большую роль в этих преобразованиях играет коэффициент  $\gamma$ , являющийся индивидуальной характеристикой слушателя. Нужный эффект при прослушивании может быть достигнут только в полностью заглушенном помещении.

Шредер осуществил всю намеченную программу, сравнил между собой лучшие концертные залы Европы и сопоставил их акустические характеристики [32].

Работа по объективной оценке залов была организована следующим образом. Оркестр в заглушенной камере сыграл несколько классических произведений, исполнение было записано на два микрофона. Эти записи с помощью двух громкоговорителей (одних и тех же) воспроизводились в различных концертных залах и там опять записывались с помощью установленных в зале двух микрофонов (громкоговорители устанавливались на сцене). Полученные таким образом записи с помощью специальных фильтров [33], осуществляющих преобразования (6.8), воспроизводились в заглушенной комнате каждому эксперту.

После обработки мнений экспертов по статистическим критериям принималось решение об акустическом качестве зала.

Столь сложной процедуры, по-видимому, можно избежать, если основывать бинауральную голографию не на соотношениях (6.5), а на соотношении (6.6), которое написано с учетом определенной модели бинаурального слухового восприятия. Преимущество метода Шредера состоит в том, что он не зависит от модели слухового восприятия и будет работать в любом случае. Существенно более простая система, основанная на соотношении (6.6), даст нужный эффект только в том случае, если верна модель рис. 33.

### 6.8. Объемное и стереообъемное звучание

Из выражения (6.6) вытекает ряд следствий, которые можно проверить экспериментально и тем самым подтвердить или опровергнуть модель бинаурального слуха, изображенную на рис. 33.

Одним из таких следствий является возможность осуществления бинауральной голографии существенно более простым путем, чем это сделал Шредер. Однако такая возможность еще никем не реализована и поэтому здесь не место для ее изложения.

Другая возможность состоит в реализации объемного звучания на практике. Эффекта объемного звучания можно добиться, располагая всего одним монофоническим сигналом. Этот эффект следует из соотношения (6.6), если при идентичности спектров сигналов, излучаемых громкоговорителями, т. е. при равенстве  $\langle PP \rangle$  и  $\langle QQ \rangle$ , добиться того, чтобы эти сигналы были практически некоррелированы, т. е. было бы  $\langle PQ \rangle = 0$ . Этого можно добиться, применяя известные в радиотехнике схемы, некоторые из них очень просты. При отсутствии корреляции сигналов  $P$  и  $Q$ , в (6.6) остаются только два первых члена. Если при этом выполнены условия для оптимального прослушивания записей, полученных методом художественной стереофонии, то образы этих громкоговорителей будут размыты и сольются для слушателя в один объемный звуковой образ, который заполняет собой все пространство между двумя громкоговорителями.

Такой эффект действительно наблюдается. При опыте вначале включается обычное монофоническое звучание на два громкоговорителя, т. е. звучание кажется исходящим из центра линии, соединяющей громкоговорители. Когда это наблюдается, то реализуются условия прослушивания художественной стереофонии, которая основана на этом эффекте.

После этого включается устройство, которое, не изменяя ни громкости, ни спектров сигналов, делает их взаимно некоррелированными. Слушатель при этом ощущает размытие звукового образа, который до этого воспринимался в точке, находящейся между громкоговорителями в центре линии, их соединяющей.

Демонстрируя этот эффект разным слушателям, музыкантам и другим людям, мы заботились о том, ощущают они или нет размытие центрального звукового образа и появление эффекта объемного звучания. Оказалось, что наряду с этим всегда присутствует и другой эффект, которого мы никак не ожидали. Музыкальные программы в объемном звучании кажутся более естественными и приятными, нежели в монофоническом звучании. Особенно сильно это ощущали музыканты, которые предпочитали объемное звучание стереофоническому. Наиболее удачные стереофонические оркестровые записи по своему качеству звучания близки к объемному, а такие звуки, как звуки рояля или звучание голоса, в объемном звучании оказывались, по мнению музыкантов, значительно предпочтительнее. Этот эффект улучшения восприятия звуков был для нас неожиданным, потому что при объемном звучании эффекты локализации звуков отсутствовали, так как исходная запись являлась презренной монофонической и никакого звукорежиссера не было. Объемное звучание из любого монофонического можно сделать нажатием кнопки.

В настоящее время этот эффект используется как при записи стереофонических программ (в доме звукозаписи имеется соответствующая аппаратура, которую может включать звукорежиссер), так и при воспроизведении.

Наибольшего эффекта удалось добиться путем соединения стереофонического воспроизведения с объемным звучанием. Для этого в каждый канал стереозвучания включается объемное звучание. Громкоговорителей становится не два, а четыре, и качество воспроизведения звуков становится заметно лучше по сравнению со стереозвучанием, а с монофоническим тем более.

О полученном таким путем качестве звучания можно судить на основании такого факта. Эта аппаратура демонстрировалась на международной выставке, проходившей в Москве ("Связь-75"). По качеству звучания она превзошла все специально сделанные профессионалами стереофонические системы настолько, что заместитель министра прямо на выставке распорядился о немедленном промышленном внедрении этой аппаратуры. Он подписал приказ о строительстве в г. Ростове-на-Дону специального цеха по выпуску соответствующей аппаратуры. Это было сделано, и аппаратура "Ростов-Дон стерео" была выпущена.

## 6.9. Чем отличается естественное звучание

До сих пор принято считать, что естественное звучание отличается от монофонического тем, что имеет распределение звуковых образов. Стерефоническое звучание предпочтительнее, так как оно способствует, даже в виде художественной стереофонии, появлению эффектов локализации образов. Это все верно, но только отчасти. Отличие, по-видимому, состоит в статистической модели источника звука. Естественные звуки описываются статистической моделью (5.6), в то время как монофонический звук представляет собою регулярный сигнал, описываемый моделью (5.7).

Особенно это видно на примере объемного и стереообъемного звучания. При объемном звучании появляются два независимых источника звука, в случае стереообъемного звучания их уже четыре. Художественная стереофония в силу своих особенностей тоже не может дать большого числа независимо звучащих источников и, как и объемное звучание, ограничивается только двумя. Существенная разница в звучании различных объемных и стереообъемных программ объясняется, очевидно, разным числом независимо звучащих источников. В естественном звучании таких независимых источников может быть много. Их предельное число ограничивается не тем, сколько их расположено фактически, а возможностями их пространственного разрешения – бинауральным слухом.

Элементарные расчеты показывают, что таких независимо звучащих источников может быть порядка десяти или двадцати. Такое звучание должно быть существенно предпочтительнее, чем стереофоническое или стереообъемное.

С этой точки зрения бинауральная голография должна давать очень высокое качество звучания, что и было отмечено особо Шредером, который слушал такое звучание и описал свои ощущения как не сравнимые со стереофоническим воспроизведением. Можно сделать попытку дать физиологическое объяснение тому факту, что человек предпочитает слушать распределенный в пространстве звук, а не звук, исходящий из определенной точки пространства. Распределенный в пространстве звук дает узкую корреляционную функцию, поэтому на схеме рис. 33 много ячеек остается свободными. Наличие свободных ячеек, очевидно, является какой-то предпосылкой лучшего эмоционального восприятия звука. Эта предпосылка особенно используется музыкантами, так как разница в восприятии стереофонического или объемного звучания среди музыкантов и не музыкантов очень велика. Эти люди попросту не могут понять друг друга.

## 6.10. Проблемы и сомнения

В заключение мне хочется подчеркнуть, что проблема слухового восприятия еще далеко не решена до конца. Приведу примеры задач, ставящих исследователей в тупик.

Можно было бы сравнительно просто решить проблему бинауральной голографии, если попросту подвести сигналы из помещения А (см. рис. 33) к ушам слушателя посредством специальных стереофонических наушников. Такие наушники воспроизводят звук гораздо лучше громкоговорителей, так как они не обладают большой мощностью, что снимает ряд проблем. Однако такой способ оказывается неприемлемым. Неприемлемость такого заманчивого во многих отношениях способа воспроизведения стереофонической программы заключается в необъясненном пока эффекте, получившем специальное название *эффекта латеризации*. Этот эффект состоит в том, что все звуковые образы представляются слушателю локализованными внутри его собственной головы. Это является существенной помехой восприятию.

Гипотезы, относящиеся к причинам возникновения и физической сущности эффекта латеризации, есть, а теории, объясняющей этот эффект и проверенной экспериментально, до сих пор нет. Это явная и серьезная загадка слухового восприятия человека.

Другой загадкой является особенность локализации человеком звуковых образов или источников звука. Человек различает направления прихода звука, отличающиеся на 180 градусов. Это загадка, так как с помощью всего двух приемников (ушей) и даже любого числа приемников, расположенных в линию, можно различать направления только относительно этой линии. Угол поворота источника вокруг линии, соединяющей приемники (уши), не позволяет определить никакая система, которую может осуществить и объяснить современная наука. Для примера представим себе, что источник расположен прямо перед человеком. Угол с осью составляет 90 градусов. Это все, что можно определить с помощью двух ушей. Источник, находящийся в противоположном направлении, т. е. точно сзади, составляет все тот же угол с осью, т. е. 90 градусов. Отличить эти направления с помощью ушей теоретически нельзя. Однако человек практически умеет четко различать эти направления прихода звука. Каким образом это совершается, точно неизвестно. По этому поводу есть только различные гипотезы.

Такого рода гипотез можно выдвинуть множество. Например, возможно, что обе загадки имеют одну общую для них причину. То, что позволяет ликвидировать в восприятии человека латериза-



цию источников, возможно помогает уточнять местоположение источников излучения (находятся ли они спереди или позади слушателя). Конкретный механизм, позволяющий это сделать, тоже может быть предложен в качестве гипотезы.

Разбираться в существующих и вновь формируемых гипотезах относительно вышеприведенных особенностей бинаурального слухового восприятия чрезвычайно трудно. Главная трудность заключается в том, что необходим эксперимент, поставленный на человеке. Постановка такого эксперимента осложнена тем, что появляется новая весьма существенная проблема, связанная с особенностями субъективного восприятия. Человек способен вообразить и явно услышать то, чего он и не должен слышать, и объективно не слышит. Физическому эксперименту это не свойственно, а при эксперименте на человеке приходится считаться с этой особенностью, и потому часто физическому эксперименту придаются еще и черты следственного.

## **7. ФОРМИРОВАНИЕ ИЗОБРАЖЕНИЯ НА ОСНОВЕ ЕГО ЧИСЛЕННОЙ МОДЕЛИ**

### **7.1. Общее описание метода**

Этот раздел может рассматриваться как общее заключение автора, вводящее читателя в современные перспективные методы получения изображений в волновых полях.

На рис. 35 схематично представлены методы формирования изображений, рассмотренные выше, в первых шести разделах. Изображение формируется путем соответствующей фильтрации сигнала, принятого на некотором расстоянии от плоскости, в которой задан входной сигнал. Такое преобразование может быть осуществлено как численно, так и с помощью самого волнового поля. Входной сигнал задается либо в виде распределения комплексных амплитуд (см. разд. 1–3), либо в виде распределения интенсивности случайного некогерентного поля (см. разд. 4–6).

Иное дело — схема получения изображений, описываемая в настоящем разделе и приведенная на рис. 36. Она построена в два ряда, что является принципиальным. Верхний ряд почти повторяет схему, показанную на рис. 35, но не включает финальной операции формирования изображения. Этот ряд заканчивается получением

ряда параметров, которые еще не могут считаться изображением входного распределения. На основе этих параметров нельзя сформировать изображение, пользуясь традиционными методами. Соответствующая обратная задача решения не имеет, так как поля могут быть необратимо усреднены. Назовем этот ряд натурным рядом (НР). Формированию изображений служит второй ряд, который назовем компьютерным рядом (КР). КР начинается с того, что компьютер формирует начальное изображение, которое включает ряд параметров или функций, изменяющих его вид. Далее это изображение преобразуется так же, как и входное изображение верхнего ряда. После этого совершается обработка сигнала, повторяющая обработку, сделанную в верхнем ряду и приводящую к получению тех же параметров, которые могут иметь в КР иные значения и иной вид (если это функции), нежели в НР.

Изображение формируется в КР путем такого изменения начального входного компьютерного изображения, которое приводит к получению в конце КР параметров или функций, близких к тем, которые содержатся в НР. Эти процедуры подбора включают изменение параметров входного изображения, формируемого компьютером, с последующим сравнением результатов обработки сигналов в КР с теми же параметрами, полученными в НР.

Компьютеру отводится не только роль вычислителя, у него есть и его собственная роль, являющаяся главной в этой схеме. Окончательно изображение не формируется на основе входных данных, оно получается путем подбора подходящего компьютерного изображения. В этой схеме используется преобразование изображений, осуществляемое с помощью методов, формул и программ, содержащихся в вышеизложенных разделах книги.

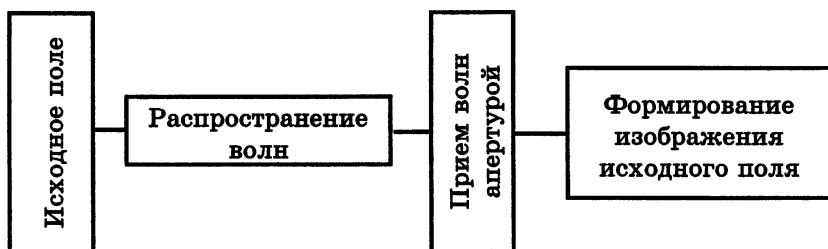


Рис. 35. Общая схема получения изображений в когерентном и в некогерентном поле. Изображение получается как решение обратной волновой задачи на основе принятого поля или его статистических характеристик

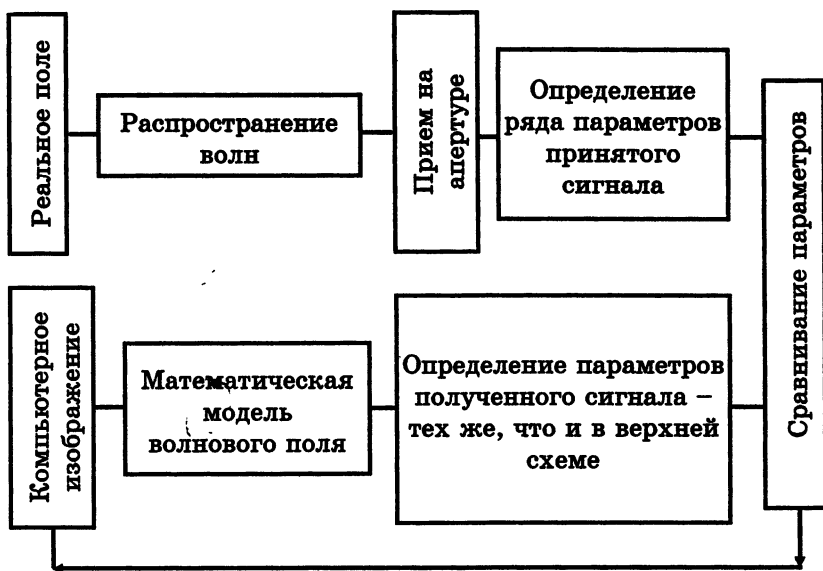


Рис. 36. Схема формирования изображения на основе численной модели

Приведенную общую схему мы наполним содержанием и более подробно поясним на примере конкретных задач.

## 7.2. Формирование изображения на входе антенны

Натурный эксперимент, использованный нами в качестве примера, опубликован в ведущем мировом журнале по обработке сигналов [34].

Постановку опыта опишем, в основном следуя тексту этой работы. Ее задачей являлось определение того, что могут дать современные методы обработки сигналов антенных решеток, подавляя сильный сигнал с целью приема слабого, в условиях проведенного опыта. Речь шла о распространении акустических сигналов в мелком море, глубина которого порядка 80 м с колебаниями порядка 20 м, на расстояние порядка 10 км. Сигнал мощного излучателя подавался на фиксированной частоте 237,5 Гц и принимался линейной антенной, имеющей 64 приемных элемента, расположенные через интервалы в половину длины волны. Сигнал был достаточно мощным, чтобы можно было пренебречь влиянием аддитивных

шумов. Чтобы определить, какое излучение можно принять на фоне сигнала мощного излучателя, в опыте использовался второй излучатель, который буксировался специальным судном и был движущимся. Сигнал этого излучателя имел частоту 235 Гц. Разность излучаемых частот позволяла при обработке отфильтровывать сигнал второго дополнительного излучателя и подмешивать его к сигналу мощного излучателя в любой пропорции. Сигналы, принятые гидрофонами антенны, перед их обработкой фильтровались в полосе 0,2 Гц около частоты излучателя.

Чтобы объективно характеризовать условия, в которых производился опыт, в [35] приведены результаты предварительной обработки сигнала мощного излучателя, принятого антенной. Эти результаты для нас будут служить выходным сигналом НР. Основной корреляционной зависимостью является:

$$A(j) = 1 - |k(j, 31)|, \quad (7.1)$$

где  $k(j, 31)$  – комплексный коэффициент корреляции между сигналами, принимаемыми приемными гидрофонами с номерами  $j$ , и сигналом, принимаемым гидрофоном с номером 31 (все гидрофоны антенны нумеруются по порядку их расположения от кончика носа до кончика хвоста). Эта картинка усреднена по 240 временным реализациям. В [35] приведена также зависимость:

$$B(j) = 1 - |k(j, j+1)| \quad (7.2)$$

и следующее отношение:

$$C(j) = \frac{1 - |k(j, j+2)|}{1 - |k(j, j+1)|}. \quad (7.3)$$

Эти зависимости объективно характеризуют условия опыта. Функции (7.2) и (7.3) построены так, что уточняют поведение (7.1) в ближайшей окрестности минимума. В отношении (7.2) это вполне понятно, а отношение (7.3) построено так, что должно быть в среднем равно 0,25 в том случае, если в ближайшей окрестности минимума функция (7.1) имеет квадратичный участок. Такой участок должен существовать в том случае, когда интервал корреляции волнового поля больше размера антенны независимо от конкретного вида корреляционной функции. Если же этот интервал меньше расстояния между отдельными гидрофонами антенны, то это отношение должно возрасти до единицы. Таким образом, это отношение характеризует величину интервала когерентности поля вдоль антенны.

Опыт состоял в определении отношения сигналов двух излучателей, при котором сигнал более слабого удаётся наблюдать на фоне

сигнала более сильного источника. Это отношение определялось для ряда самых современных методов выделения сигналов с помощью антенных решеток. В результате этого опыта было установлено, что такое отношение не превышает 30 дБ. Приему более слабого сигнала препятствует шум, вызываемый присутствием сильного сигнала. Этот шум является следствием флуктуаций сильного сигнала. Физический механизм возникновения флуктуационного шума в [35] не упоминается. Его физическая природа точно такая же, как и спекл-шума, описанного в разд. 1.6.

Покажем, как с помощью рассматриваемого здесь метода формирования изображений можно увидеть, откуда идет этот шум. Опубликованные в [35] корреляционные зависимости (7.1)–(7.3) будут для нас служить конечным этапом операции НР (см. рис. 36).

Далее следует согласно нашей схеме сформировать компьютерное изображение. Прежде всего определим, что же мы хотим увидеть с помощью компьютерного метода. Мы хотим увидеть, с каких направлений и какой интенсивности идет шум, который маскирует слабый сигнал. Для формирования соответствующего компьютерного изображения воспользуемся физической стороной явления.

Она заключается в следующем. Физической причиной флуктуационных явлений в данном случае является рассеяние акустических волн излучателя взволнованной поверхностью моря. Мелкое море представляет собою акустический волновод с неровной границей, рассеивающей падающие на нее волны. Это явление, применительно к данному опыту, подробно рассмотрено в [36], где содержится методика расчета этого явления, позволяющая получать численные оценки. С помощью приведенного в [36] элементарного расчета можно, пренебрегая интерференцией рассеянных волн, получить гладкую кривую, показывающую интенсивности рассеянных сигналов, приходящих с разных направлений. Типичное угловое распределение интенсивности волн, рассеянных поверхностью, показано на рис. 37. Это распределение мы примем за исходное компьютерное изображение.

Особенностью реального волнового поля является его монохроматичность, что приводит к интерференции рассеянных волн. Компьютерное изображение получено в предположении отсутствия интерференции. Поэтому сначала надо сформировать изображение так, чтобы интерференция в нем присутствовала. Такое поле можно, в принципе, сформировать численно, пользуясь формулами разд. 1 (например, (1.5)). Пример такого численного расчета приведен в приложении (см. рис. 4П). Однако для этого следует точки на поверхности задавать с интервалами не более, чем в  $\lambda/2$ . Пло-

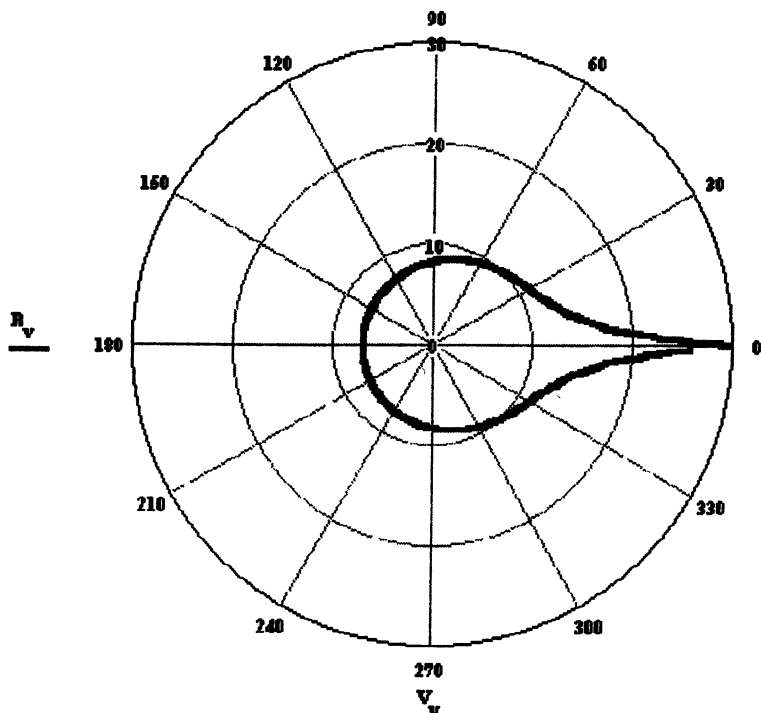


Рис. 37. Угловое распределение сигнала мощного излучателя, рассеянного морской поверхностью. Углы отсчитываются от направления от приемной антенны на излучатель. Вдоль радиуса отложен уровень рассеянного сигнала в децибелах

щадь рассеяния весьма велика, а значит, такой расчет потребует гигантского объема вычислений, который не влезет ни в один компьютер. Поэтому поступим так, как это сделано в работе [35], и учтем результат интерференции на конечном этапе расчета, выполненного без учета интерференции (что достигается при существенно меньшем числе точек). Такая методика обоснована нами в [36]. Здесь мы следуем описанной и обоснованной там методике. Сформируем комплексные случайные числа:

$$Q_{v,z} = A_{v+Lz} + iB_{v+Lz}, \quad (7.4)$$

где  $A$  и  $B$  – последовательности случайных чисел, генерируемых компьютером. Последовательности взяты достаточно протяженными, чтобы из них, как показано в формуле (7.4), можно было сформировать последовательности комплексных случайных чисел для целого ряда реализаций, каждая из которых обладает размером

360 значений, что можно сопоставить с градусами. Компьютерное усредненное изображение (см. рис. 37) представлено значениями, взятыми через градус.

Образуем случайные спектральные составляющие

$$q(v, z) = \sqrt{S(v)} Q(v, z) + ZO \delta(v). \quad (7.5)$$

Каждое значение  $v$ , выраженное в градусах, сопоставим с плоской волной, пользуясь формулой (1.2). Получим

$$W(x, v) = \exp[iu(v)x], \quad (7.6)$$

где волновой вектор  $u(v)$  имеет вид

$$u(v) = \frac{2\pi}{\lambda} \sin\left(\frac{2\pi}{360} v\right), \quad (7.7)$$

$x$  — координата точки пространства в плоскости  $z = 0$  в направлении оси  $x$ .

Для образования волнового поля следует сложить плоские волны (7.6), придав им случайные амплитуды (7.5). В результате получим волновое поле, в котором интерферируют плоские волны:

$$P(x, z) = \sum_v W(x, v) q(v, z). \quad (7.8)$$

В это поле можно помещать антенну, апертура которой линейна и ориентирована вдоль оси  $x$ , а также можно получать статистические характеристики этого поля, в частности комплексные коэффициенты корреляции.

Статистические характеристики волнового поля, полученные в эксперименте, определяются следующими соотношениями:

$$R(x) = 1 - \left| C(x) / \sqrt{CE(x) CE(31)} \right|, \quad (7.9)$$

где

$$C(x) = \frac{1}{N} \sum_z P(x, z) \overline{P(31, z)}, \quad CE(x) = \frac{1}{N} \sum_z \left[ |P(x, z)|^2 \right];$$

$$RJ(x) = 1 - \left| \frac{C_1(x)}{\sqrt{CE(x) CE(x+1)}} \right|, \quad (7.10)$$

где

$$C_1(x) = \frac{1}{N} \sum_z P(x, z) \overline{P(x+1, z)};$$

$$K(x) = \frac{R_1(x)}{R_2(x)}, \quad (7.11)$$

где

$$R_2(x) = 1 - \left| \frac{C_2(x)}{\sqrt{CE(x) CE(x+2)}} \right|,$$

$$C_2(x) = \frac{1}{N} \sum_z P(x, z) \overline{P(x+2, z)}.$$

Статистические характеристики поля (7.9), (7.10) и (7.11) определялись в эксперименте.

Таким образом, мы имеем возможность сравнивать эти характеристики с полученными экспериментально, и по их сходству или различию судить о сходствах или различиях изображения, имевшего место в опыте, и компьютерного.

В качестве первого шага формирования изображения рассеивающей области было взято изображение, приведенное в [36] и показанное на рис. 37. С помощью формул (7.5)–(7.8) были получены статистические характеристики (7.9), (7.11) и мгновенный отклик антенны и сопоставлены с полученными экспериментально. Результаты существенно разошлись, что говорило о несоответствии компьютерного изображения с искомым. Корреляция между значениями функции  $R(x)$  (7.9) была явно меньше корреляции между значениями функции  $A(j)$  НР. Это связано с тем, что в компьютерном изображении угловой спектр сигналов был очень широким. Изменение параметров среды, принятых в расчетах компьютерного изображения (поглощения, глубины волновода, расстояния между излучателем и приемником, характеристик степени убывания рассеянных волн), не приводило к изменению функции  $R(x)$  в сторону ее большего соответствия с  $A(j)$ . Поэтому деформацию компьютерного изображения мы осуществляли с помощью внешнего параметра, не рассматривавшегося в работе [36].

Как уже отмечалось выше, сигнал НР перед обработкой подвергался узкополосной фильтрации. Это могло привести к исключению или существенному ослаблению рассеяния из тех областей, в которых частоты рассеянных сигналов имели доплеровское смещение частоты, превышающее полосу пропускания фильтра.

Если взглянуть сверху на рассеивающую поверхность, увидим относительное распределение доплеровских смещений частот таким, каким оно показано на рис. 38. На рис. 38, а показаны изолинии доплеровского смещения частоты (в единицах максимального смещения) для случая, когда направление движения рассеивающей поверхности происходит перпендикулярно воображаемой линии, соединяющей приемник и излучатель. На рис. 38, б



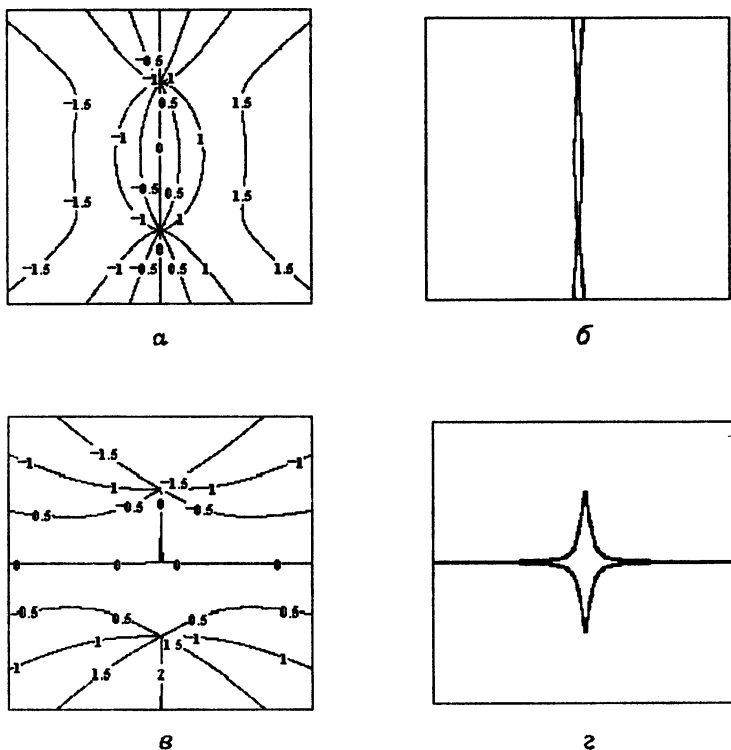


Рис. 38. Изолинии доплеровского смещения частоты  
рассеянного сигнала на поверхности

показана область, ограниченная некоторой полосой доплеровских частот. Применение частотной фильтрации приводит к подобному ограничению размера и изменению формы области, откуда могут приходиться рассеянные сигналы. На рисунке *б* – вид изолиний доплеровских частот при движении рассеивающей поверхности вдоль линии излучатель – приемник. На рисунке *г* показана область, откуда могут приходиться рассеянные сигналы, – в этом случае при ограничении полосы принимаемых частот.

Соответствующие изменения в связи с фильтрацией доплеровских частот вносились в исходное компьютерное изображение. В результате форма области, из которой могут приходиться рассеянные сигналы, деформировалась так, как показано на рис. 38. При этом было учтено, что никакой фильтр не в состоянии полностью убрать сигналы, находящиеся вне его полосы пропускания. Это было

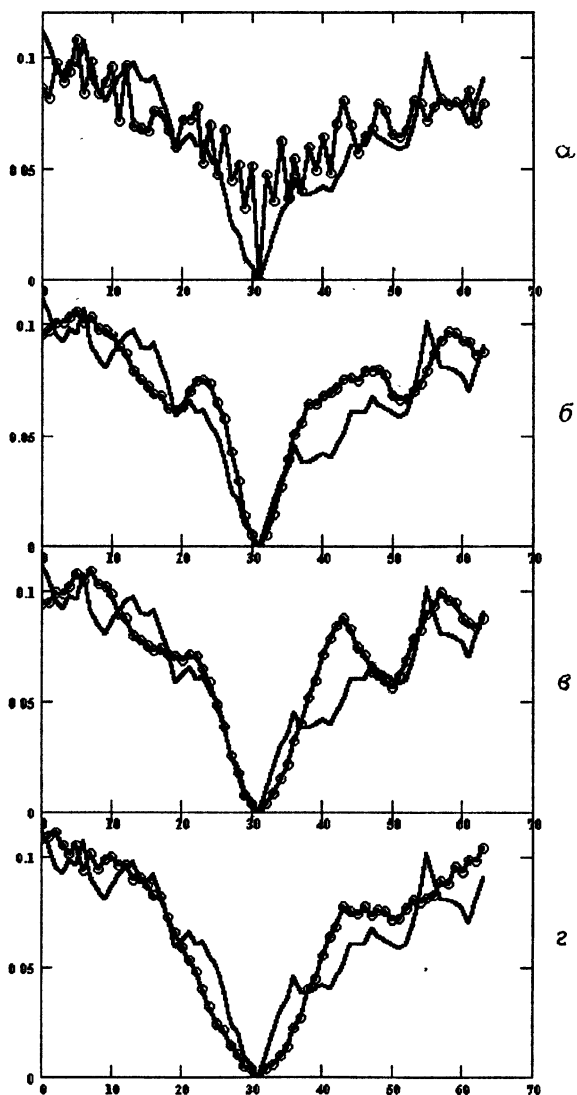


Рис. 39. Результат сравнения функций  $A(j)$  (сплошная линия) и  $R(j)$  (линия с точками) при различных компьютерных изображениях без фильтрации доплеровских частот (а) и при частотах среза фильтра доплеровских частот 0,1 (б); 0,075 (в); 0,05 (г) от максимального доплеровского смещения частоты в предположении, что рассеиватели двигаются поперек трассы

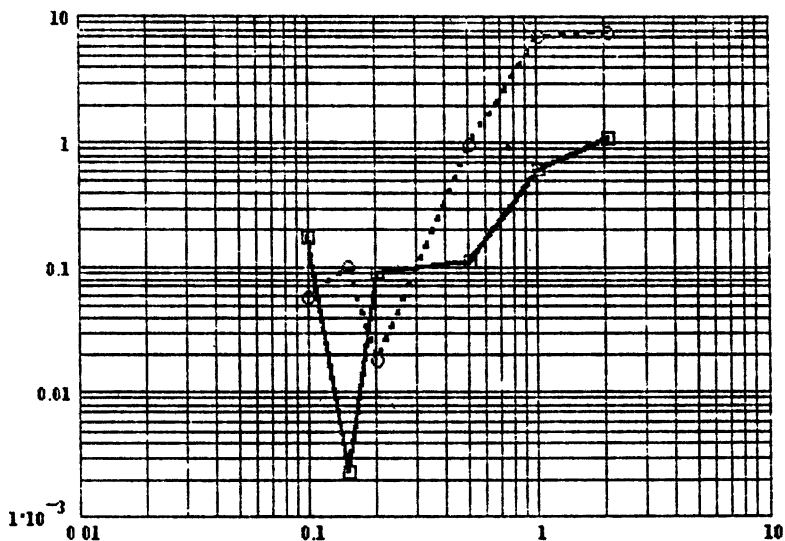


Рис. 40. Результат сравнения значений

$$\frac{|\langle C(f) \rangle - \langle R_1(f) \rangle|^2}{\langle C(f) \rangle \langle R_1(f) \rangle}$$

НР и КР, отмеченных кружочками, соединенными пунктиром, и аналогично представленных средних значений  $B(f)$  и  $R_2(f)$  тех же рядов в зависимости от частоты среза фильтра доплеровского смещения частоты.

Предполагается, что скорость движения рассеивателей направлена поперек трассы

учтено тем, что небольшая часть всего полного сигнала, не подверженного фильтрации (ослабленная на 25 дБ), прибавлялась к отфильтрованному сигналу.

Результаты представлены отдельно для направления скорости смещения поверхности поперек трассы или вдоль нее.

На рис. 39 представлены корреляционные функции для фильтра, срезающего доплеровские частоты на уровне 1; 0,1; 0,075; 0,05 от максимального доплеровского смещения. Это означает, что первая картинка представлена без всякого среза. Другие статистические параметры представлены на рис. 40. На них по оси абсцисс отложен срез фильтра в единицах удвоенного максимального смещения частоты (в логарифмическом масштабе). По оси ординат отложены разности значений параметров, усредненных вдоль антенны и нормированных на их произведение. На рис. 41 показаны компьютерные изображения, соответствующие выбранным границам среза фильтра.

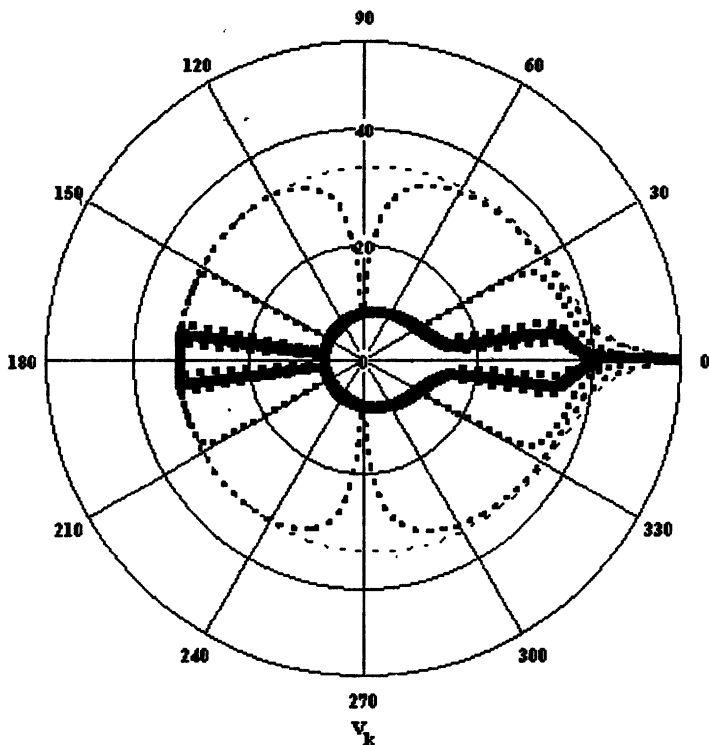
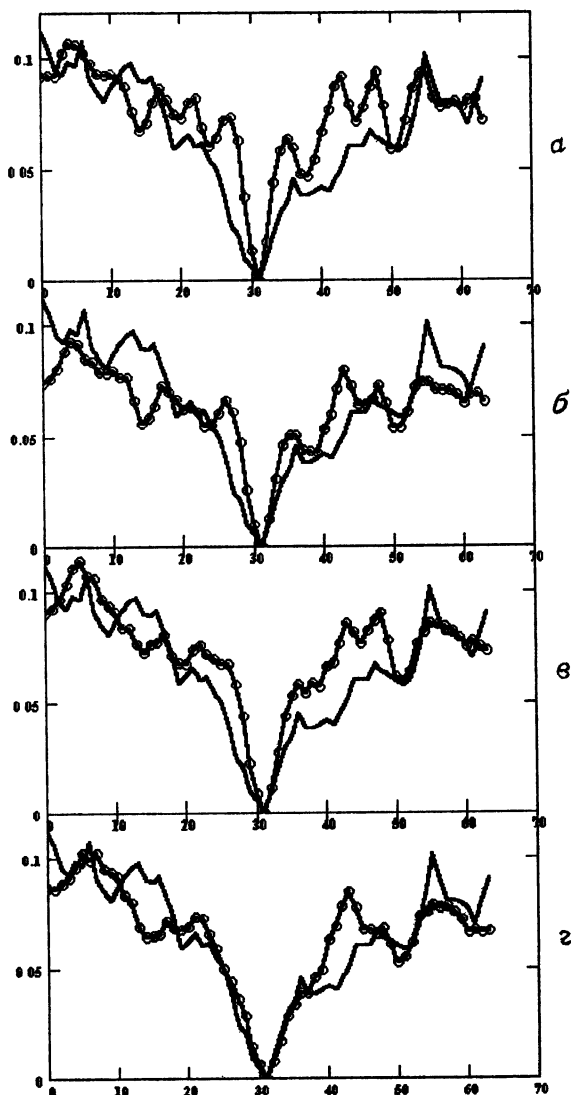


Рис. 41. Компьютерное изображение при различных частотах пропускания фильтра доплеровских частот

Как видно из рис. 39–40, вполне удовлетворительное согласие параметров НР и КР получается при срезе фильтра 0,1 от максимального доплеровского смещения частоты. Соответствующее этому значению среза фильтра компьютерное изображение на рис. 41 показано жирной сплошной линией. Изображения, соответствующие иным значениям среза фильтра, даны пунктиром, линия которого тем тоньше, чем шире соответствующий доплеровский фильтр. Эти изображения показывают разрешающую способность метода при определении той области поверхности, от которой исходит существенное рассеяние. Чтобы частота среза доплеровского фильтра 0,2 от максимального доплеровского смещения совпала с фактической частотой среза фильтра, указанной в [35] (0,2 Гц), надо, чтобы скорость волн на поверхности была порядка 5 м/с, что является реальным.



**Рис. 42.** Результат сравнения функций  $A(j)$  (сплошная линия) и  $R(j)$  (линия с точками) при различных компьютерных изображениях без фильтрации доплеровских частот (а) и при частотах среза фильтра доплеровских частот 0,1 (б); 0,01 (в); 0,005 (г) от максимального доплеровского смещения частоты в предположении, что рассеиватели движутся вдоль трассы

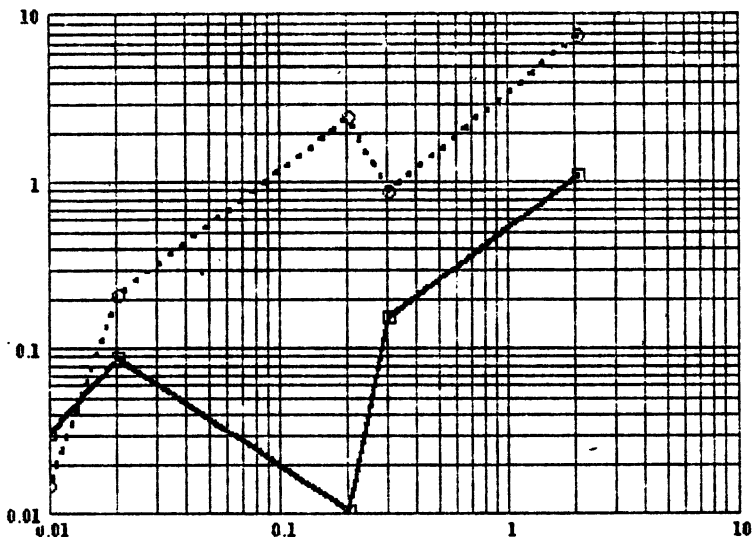


Рис. 43. Результат сравнения значений

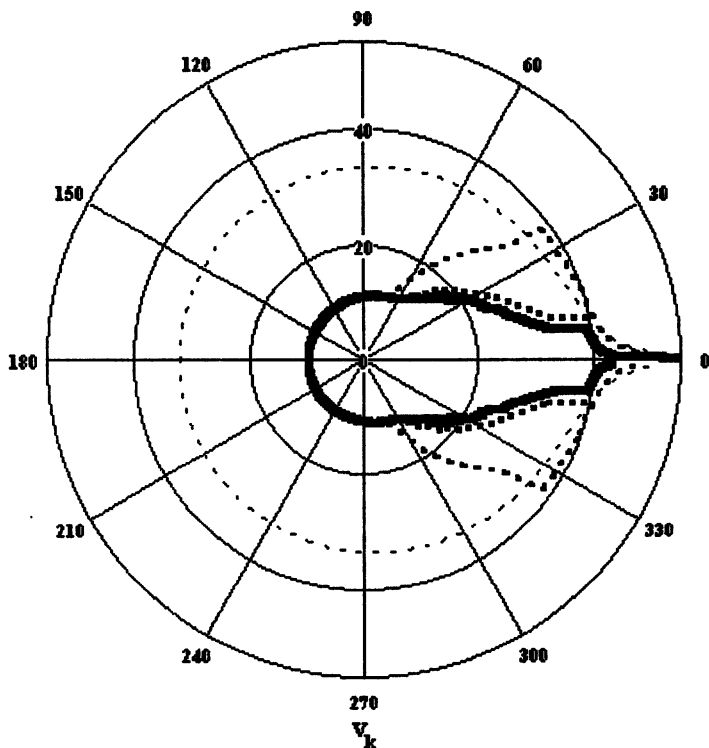
$$\frac{\langle C(f) \rangle - \langle R_1(f) \rangle^2}{\langle C(f) \rangle \langle R_1(f) \rangle}$$

НР и КР, отмеченных кружочками, соединенными пунктиром, и аналогично представленных средних значений  $B_1(j)$  и  $R_2(j)$  тех же рядов в зависимости от частоты среза фильтра доплеровского смещения частоты.

Предполагается, что скорость движения рассеивателей направлена вдоль трассы

То обстоятельство, что фильтрация в эксперименте осуществляется не до конца и добавление полного сигнала в какой-то пропорции необходимо, показывает мгновенный отклик антенны, на котором видны флуктуационные изменения (спекл-шум) при всех углах (см. далее рис. 47).

На рис. 42–44 показано то же самое сопоставление результатов НР и КР для направления ветра вдоль трассы. Здесь также содержится набор выходных параметров КР, которые отлично совпадают с параметрами НР. Однако частоты среза фильтра доплеровских частот (0,005 от максимума) в этом случае явно нереальны, так как требуют на порядок большей скорости волн. Однако при этом форма изображения поверхности хорошо совпадает с той, которая получается при ином направлении скорости движения рассеивателей. Это говорит о том, что определяются не физические параметры эксперимента, а только изображение области, с которой поступают рассеиваемые сигналы. Это изображение по условию формирования параметров НР и КР не должно отличать направления, различия



**Рис. 44.** Вид компьютерного изображения при различных частотах пропускания фильтра доплеровских частот в предположении, что рассеиватели движутся вдоль трассы. Жирной сплошной линией отмечено изображение при частоте среза фильтра 0,01 от максимума. Толстым пунктиром показаны изображения при частотах среза 0,2 и 0,005 от максимума, которым соответствуют небольшие, но вполне заметные изменения параметров КР. Тонкая пунктирная линия – изображение без фильтрации

на 180 градусов. Если это учесть, то изображение на рис. 44 можно считать похожим на изображение рис. 41 при ином направлении движения рассеивателей.

### 7.3. Формирование изображения на выходе антенны

Оптическая система формирует изображения в различных плоскостях. Рассматриваемый нами компьютерный метод также допускает такую возможность. В предыдущем разделе мы

сформировали изображение в плоскости входных сигналов до антенны, там, где происходит рассеяние. Теперь сформируем изображение, получающееся на выходе антенны. Его координатами будут пространственные частоты.

В этот раз компьютерное изображение представим в виде следующей математической формулы:

$$F(u) = \exp\left(-\frac{|u - 32|}{D}\right) + PD + ZO \delta(u - 32). \quad (7.12)$$

В этой формуле 3 параметра, значения которых надо подобрать так, чтобы максимально сблизить статистические характеристики НР и КР.

Прежде всего образуем необходимые нам статистические характеристики КР. Для этого запишем последовательность случайных чисел (7.4), взяв для переменной  $v$  число значений, равное числу элементов антенны (64). После этого спектр по переменной  $v$  этой последовательности умножим на компьютерное изображение (7.12) (за исключением последнего слагаемого, которое не является случайным).

Полученный ряд значений следует подставить в выражения (7.9)–(7.11), сформировав статистические характеристики КР.

Далее образуем следующие числовые характеристики, позволяющие сравнивать выходы рядов НР и КР:

$$W = (A/C)^{1/2}, \quad (7.13)$$

где

$$A = \sum_j [A(j) - C(j)]^2, \quad C = \sum_j [A(j) + C(j)]^2,$$

$$W' = \frac{\langle B(j) \rangle - \langle RJ(j) \rangle}{\langle B(j) \rangle + \langle RJ(j) \rangle}, \quad (7.14)$$

$$W'' = \frac{\langle C(j) \rangle - \langle K(j) \rangle}{\langle C(j) \rangle + \langle K(j) \rangle}. \quad (7.15)$$

Числа  $W$ ,  $W'$ ,  $W''$  рассматриваются как невязки между выходами КР и НР. Параметры компьютерного изображения подбираются так, чтобы минимизировать сумму этих невязок по абсолютной величине.

Для этой цели может быть применена соответствующая математическая технология. Например, мы поступили следующим об-

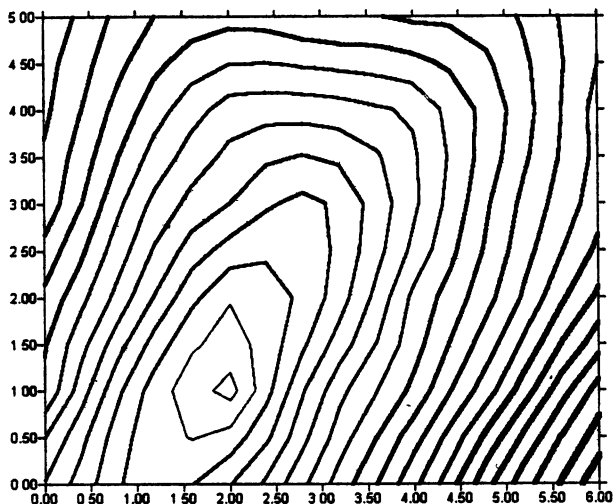


разом. Были получены значения всех трех чисел  $W, W', W''$  для некоторого диапазона изменения значений трех параметров  $D, PD, ZO$ . Затем для каждой пары значений  $PD, ZO$  был определен минимум суммарной невязки в зависимости от значения параметра  $D$ . Значение этого минимума было представлено в виде функции двух переменных  $PD$  и  $ZO$ . Полученная поверхность в виде изолиний показана на рис. 45. Искомыми значениями параметров являются значения, в которых поверхность имеет абсолютный минимум, точка абсолютного минимума четко выявляется. Это достигнуто благодаря тому, что значениям параметров  $W, W'$  и  $W''$  при суммировании невязок приданы различные веса (1; 0,1 и 0,1 соответственно). При меньших значениях весов для последних двух параметров точка абсолютного минимума не выделяется, а при больших значениях образуются локальные минимумы, не изменяющие координаты абсолютного минимума. Локальные минимумы лишают картину наглядности, если на ней не указывать значения проведенных линий уровня.

Приведем сопоставление выходов НР и КР для параметров изображения на выходе антенны (7.12) при значениях всех 3 параметров, соответствующих точке минимума поверхности невязок, показанной на рис. 45.

На рис. 46, а показан график функции  $B(j)$ , определяемой (7.2), как функции номера гидрофона  $j$ . Для сравнения с результатом КР (см. рис. 36) используется только среднее значение этой функции, которое обозначено на рисунке линией. На этом же рисунке показано (тоже в виде линии) среднее значение функции  $RJ(x)$ , определенной посредством (7.10), относящейся к КР. Разница между средними значениями НР и КР на рисунке заштрихована вертикальными линиями. На рис. 46, б аналогичная процедура показана применительно к функциям  $C(j)$  (определяемой (7.3)) натурального ряда и  $K(x)$  компьютерного ряда (определяемой (7.11)). На рисунке в – график функции  $A(j)$  НР (линия без точек), определяемый (7.1) и наложенный на него график функции  $R(x)$ , определяемый (7.9), на который нанесены круглые точки.

Как видно из рис. 46, результаты аналогичной обработки сигналов в НР и КР хорошо совпадают, что говорит о соответствии компьютерного изображения (7.12) с соответствующими параметрами. На рис. 47 показан мгновенный отклик антенны, полученный в тех же координатах КР, что и отклик НР. Само компьютерное изображение, соответствующее угловому отклику антенны, показано на рис. 48.



**Рис. 45.** Изолинии поверхности минимумов суммарной невязки параметров НР и КР.

Показанные минимумы получились в результате вариации параметра  $D$  – ширины спектра изображения. Поверхность построена в виде функции параметров  $PD$  (вдоль горизонтали) и  $ZO$  (вдоль вертикали).

Значения параметров  $x$  по оси абсцисс и  $y$  по оси ординат для получения их значений в изображении следует подставлять в формулы:

$$PD = 0,04 + 0,005x \text{ и } ZO = 7 + 0,5y$$

Таким образом, мы получили изображение углового отклика антенны, который имел место в эксперименте. Полученный результат интересен и важен тем, что позволяет получить все выводы статьи [35], используя для этого сигнал одного излучателя, без использования второго и всей примененной в статье хитроумной техники выделения слабого сигнала.

Компьютерное изображение, соответствующее угловому отклику антенны, приведено на рис. 48. На нем же показано изображение, в основе которого лежит иная формула (вместо линейной экспоненты использована квадратичная). В этом случае минимальная невязка получается существенно больше.

Действительно, спектральный уровень изображения закрывает весь акустический горизонт, показывая уровень шума, который несет с собой мощный сигнал. Уровень минимального сигнала, который может быть выделен на фоне сильного сигнала, определяется по компьютерному изображению, полученному на антенне, с помощью обычных статистических оценок, в основу которых может быть положен факт, что никакие хитроумные приемы не помогают

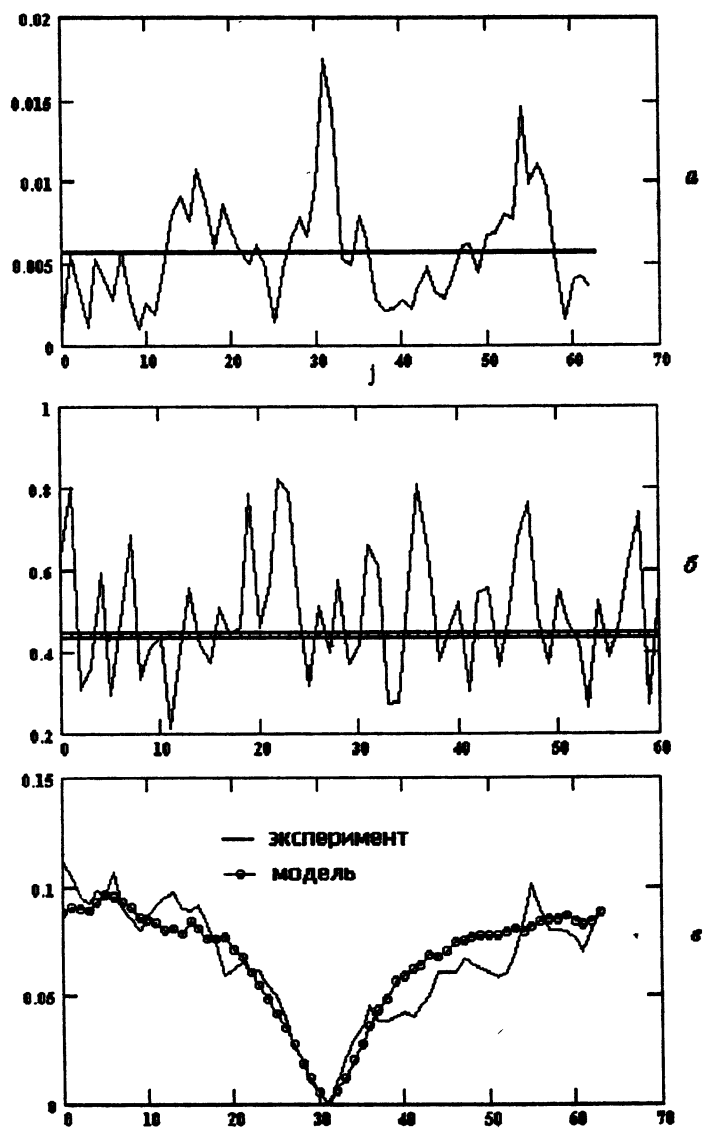


Рис. 46. Сравнение выходов НР и КР для параметров компьютерного изображения, соответствующих точке минимума поверхности, показанной на рис. 45

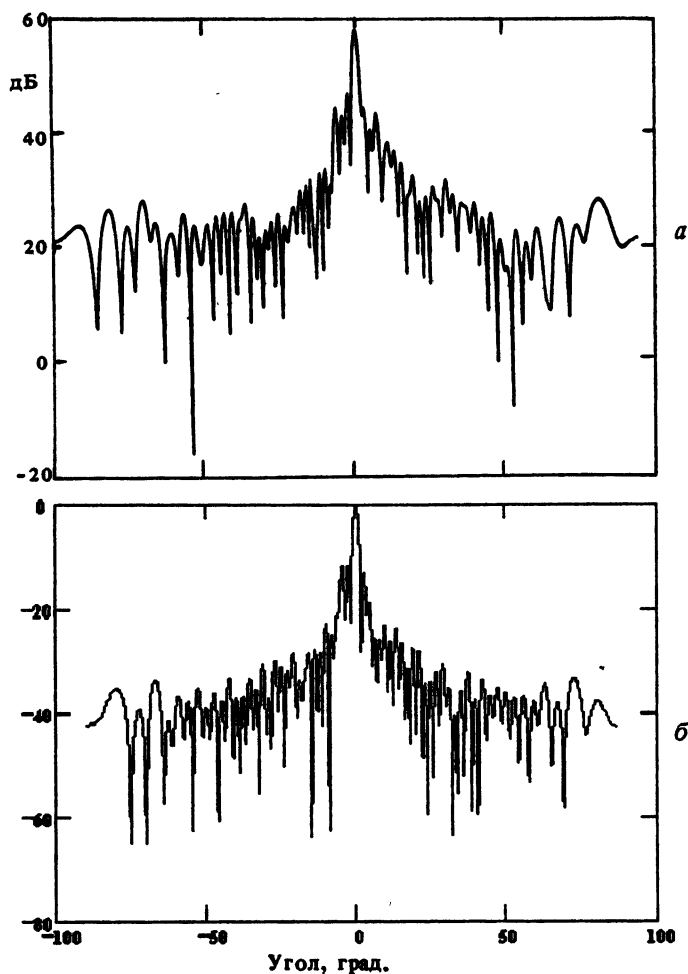
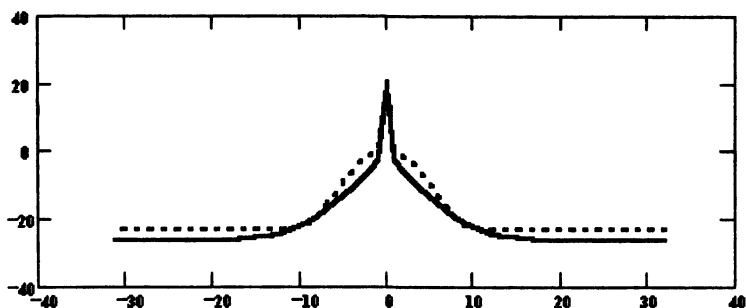


Рис. 47. Мгновенный отклик антенны в НР (а) и КР (б)

выделить сигнал из спекл-шума. Такой расчет может быть сделан для всех значений углов, а не только для тех, на которых в данный момент находится слабый излучатель.

Таким образом, построение компьютерных изображений дает возможность чуточку глубже проникнуть в физику явлений, это



**Рис. 48.** Форма отклика антенны на сигнал излучателя, прошедший через исследуемую среду. Вдоль оси абсцисс отложены пространственные частоты в единицах углового разрешения антенны. Вдоль оси ординат — уровень отклика антенны в децибелах. Сплошной линией показаны значения, полученные по формуле (7.12) при следующих значениях параметров:  $D = 2,9$ ;  $PD = 0,05$ ;  $ZO = 8$ . Минимальная сумма невязок составляет 0,1. Пунктиром показаны значения, полученные при замене в (7.12) линейной экспоненты на квадратичную при следующих значениях параметров:  $D = 20$ ;  $PD = 0,072$ ;  $ZO = 10$ . Минимальная сумма невязок составляет 0,3

происходит благодаря тому, что видна форма области, откуда приходят рассеянные сигналы при их фильтрации.

Практическое значение метода заключается в возможности наглядно увидеть получающуюся помехо-сигнальную ситуацию, создаваемую сильным сигналом в волноводе с неровными и движущимися границами.

Автор весьма благодарен рецензентам книги Н. С. Степанову и Л. С. Долину, взявшим на себя нелегкий труд проверки результатов автора и согласившимся разделить с автором ответственность за его возможные упущения и ошибки. Особую признательность выражаю редактору Наталье Николаевне Крапиной, которая своими многочисленными поправками внесла ясность и удобочитаемость в представленную ей рукопись. Работа по написанию книги и ее изданию поддержана Российским фондом фундаментальных исследований (гранты № 96-15-96603, № 96-02-18621 и № 97-02-17555).

# **ПРИЛОЖЕНИЕ**

## Знаменитое светлое пятно в центре тени

Диапазонные переменные

$$L := 256 \quad k := 0..L - 1 \quad n := 0..L - 1 \quad v := 0..0.5 \cdot L - 1 \quad u := 0..0.5 \cdot L - 1$$

Пространственные переменные

Длина волны

$$x_k := k - 0.5 \cdot L \quad y_n := n - 0.5 \cdot L \quad R := 16$$

$$p := 4$$

Формирование исходного поля

$$F(k, n) := \begin{cases} 0 & \text{if } R^2 - (x_k)^2 - (y_n)^2 > 0 \\ 1 & \text{otherwise} \end{cases}$$

$$z := \frac{R^2 \cdot 2}{p^2}$$

Дистанция в длинах волн  
до первой зоны Френеля

$$e := 0..0.25 \cdot L - 1$$

$$z = 32$$

$P_{k,n} := F(k, n)$  Исходное поле

$$w := 0..0.25 \cdot L - 1$$

Пространственные частоты

$$U1_k := \frac{k - 0.5 \cdot L}{L} \cdot 2 \cdot \pi \cdot p \quad U2_n := \frac{n - 0.5 \cdot L}{L} \cdot 2 \cdot \pi \cdot p$$

ЧХ свободного пространства

$$T_{k,n} := \exp \left[ i \cdot z \cdot \sqrt{(2 \cdot \pi)^2 - (U1_k)^2 - (U2_n)^2} \right]$$

Преобразование ЧХ к виду пространственного спектра в Mathcad

$$E_{v,u+0.5L} := T_{v+0.5L,u} \quad E_{v+0.5L,u} := T_{v,u+0.5L}$$

$$E_{v,u} := T_{v+0.5L,u+0.5L} \quad E_{v+0.5L,u+0.5L} := T_{v,u}$$

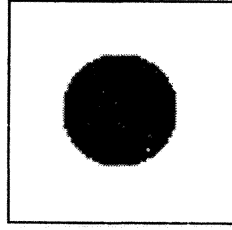
Вычисление поля на расстоянии z

$$g := \text{cfft}(P) \quad G_{k,n} := g_{k,n} \cdot E_{k,n} \quad PD := \text{icfft}(G)$$

Представление результатов

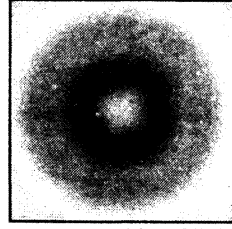
$$PR_{e,w} := PD_{0.5L+e-\frac{L}{8}, 0.5L+w-\frac{L}{8}}$$

Модуль исходного поля



$$PO_{e,w} := P_{0.5L+e-\frac{L}{8}, 0.5L+w-\frac{L}{8}}$$

Модуль поля на расстоянии одной зоны Френеля



PO

PR

Рис. 1П



## Дифракция волны на щели

Переменные различного диапазона изменений

L := 256    k := 0..L-1    q := 0..31    n := 0..0.5·L-1

Параметры задачи

R := 0    M := 16    p := 4    z<sub>q</sub> := R + 0.5·q

Пространственная частота

u<sub>k</sub> := 2· $\frac{\pi}{L}$ ·(k - 0.5·L)·p

Щель ширины M

ЧХ свободного пространства

$$O1(x) := \begin{cases} 1 & \text{if } 0.5 \cdot L - 0.5 \cdot M < x < 0.5 \cdot L + 0.5 \cdot M \\ 0 & \text{otherwise} \end{cases}$$

$$f_{k,q} := \exp[i \cdot z_q \cdot \sqrt{(2 \cdot \pi)^2 - (u_k)^2}]$$

Поле за щелью

Приведение ЧХ к виду пространственного спектра

PT<sub>k</sub> := O1(k)    ff<sub>(n,q)</sub> := f<sub>(0.5·L+n,q)</sub>    ff<sub>(0.5·L+n,q)</sub> := f<sub>n,q</sub>

Вычисление полей в зависимости от расстояния z

Спектры

Волновое поле

gt := cfft(PT)    GT<sub>k,q</sub> := gt<sub>k</sub>·ff<sub>k,q</sub>    PFT<sup><q></sup> := icfft(GT<sup><q></sup>)

Данные для построения графиков

Спектр поля на щели

Реальная часть ЧХ

Поле за щелью

$$g^1_k := |gt_k|$$

$$Z1_k := \text{Re}(ff_{k,16})$$

$$JPT_{(n,q)} := |PFT_{(0.5L-64+n,q)}|$$

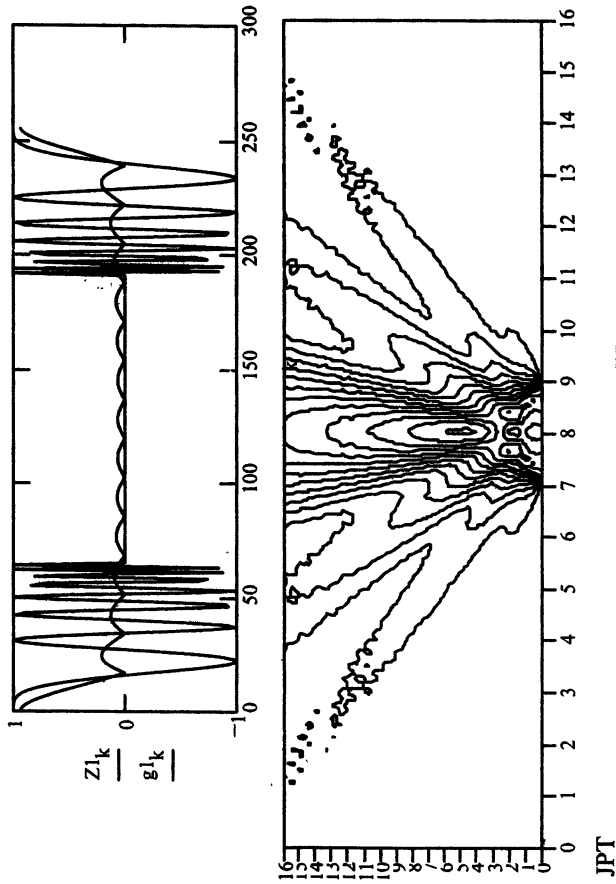


Рис. 2П

## Фокусировка

Переменные различного диапазона изменений

$$L := 256 \quad M := 64 \quad q := 0..M - 1 \quad k := 0..L - 1 \quad n := 0..0.5 \cdot L - 1$$

Пространственная частота

$$R := 10 \quad z_q := R + 0.25 \cdot (q - 0.5 \cdot M) \quad p := 4 \quad F := \frac{1}{8} \cdot L \quad u_k := 2 \cdot \frac{\pi}{L} \cdot (k - 0.5 \cdot L) \cdot p$$

Щель ширины F

$$O(x) := \begin{cases} 1 & \text{if } 0.5 \cdot L - F < x < 0.5 \cdot L + F \\ 0 & \text{otherwise} \end{cases}$$

Поле за щелью

$$x_k := \frac{k - 0.5 \cdot L}{p} \quad r_k := \sqrt{R^2 + (x_k)^2}$$

Сферическая волна

$$PT_k := \frac{1}{\sqrt{r_k}} \cdot \exp(-i \cdot r_k \cdot 2 \cdot \pi) \cdot O(k)$$

Волновое поле

$$PFT^{\langle q \rangle} := icfft(GT^{\langle q \rangle})$$

ЧХ свободного пространства

$$f_{k,q} := \exp \left[ i \cdot z_q \cdot \sqrt{(2 \cdot \pi)^2 - (u_k)^2} \right]$$

Приведение ЧХ к виду пространственного спектра

$$ff_{(n,q)} := f_{(0.5 \cdot L + n, q)} \quad \text{ff}_{(0.5 \cdot L + n, q)} := f_{n, q}$$

Спектры

$$gt := cfft(PT) \quad GT_{k,q} := gt_k \cdot ff_{k,q}$$

## Данные для построения графиков

Спектр поля на щели

$$g1_k := |g1_k| \cdot 2$$

Реальная часть ЧХ

$$Z1_k := \text{Re}(ff_k, 16)$$

Поле за щелью

$$B4_{n,q} := |PFT_{n+0.25 \cdot L, q}|$$

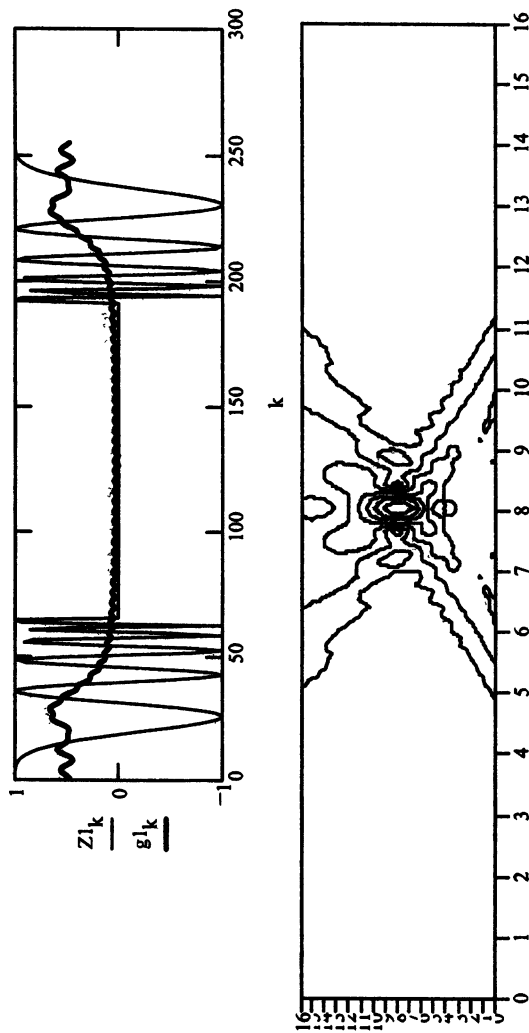


Рис. 3П

## Рассеяние волны неровной поверхностью

Переменные различного диапазона изменений  
 $L := 512$   $M := 32$   $q := 0..M - 1$   $k := 0..L - 1$   $F := 20$   $n := 0..0.5 \cdot L - 1$   $tr := 0..80 - 1$

Параметры задачи

$er1 := 50$   $R := 0$   $z_q := R + 0.5 \cdot q$   $p := 4$   $t_k := \frac{3.95}{p \cdot 2} \cdot \pi \cdot k$

Пространственная частота

$$u_k := 2 \cdot \frac{\pi}{L} \cdot (k - 0.5 \cdot L) \cdot p$$

Щель ширины F

ЧХ свободного пространства

$$O(x) := \begin{cases} 1 & \text{if } er1 < x < er1 + F \\ 0 & \text{otherwise} \end{cases}$$

$$f_{k,q} := \exp[i \cdot z_q \cdot \sqrt{(2 \cdot \pi)^2 - (u_k)^2}]$$

Поле за щелью

$BY_k := md(1)$  Случайные числа

$$B_k := \exp(i \cdot t_k) \cdot O(k) \cdot BY_k$$

Спектры

$$gt := cfft(B) \quad GT_{k,q} := gt \cdot ff_{k,q}$$

Приведение ЧХ к виду пространственного спектра

$$ff_{(n,q)} := f_{(0.5 \cdot L + n, q)} \quad ff_{(0.5 \cdot L + n, q)} := f_{n, q}$$

Волновое поле

$$pfft^{<q>} := icfft(GT^{<q>})$$

Данные для построения графиков

Спектр поля      Реальная часть ЧХ      Рассеянное поле

Негатив

$$g^1_k := |g^1_k| \cdot 2 \quad Z1_k := \text{Re}(ff(k, 16)) \quad B4_{tr,q} := |PFT_{tr+20,q}| \quad U := \max(B4) \quad B5_{tr,q} := U - B4_{tr,q}$$

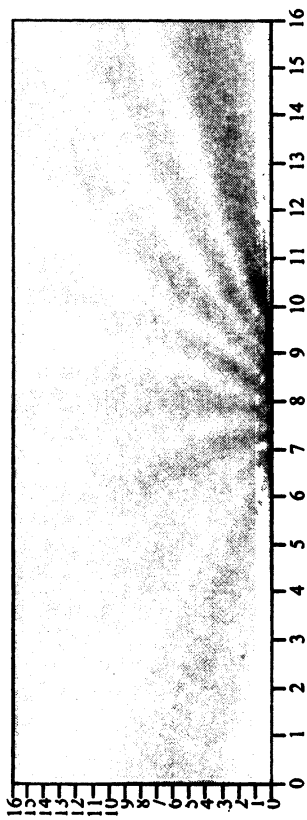
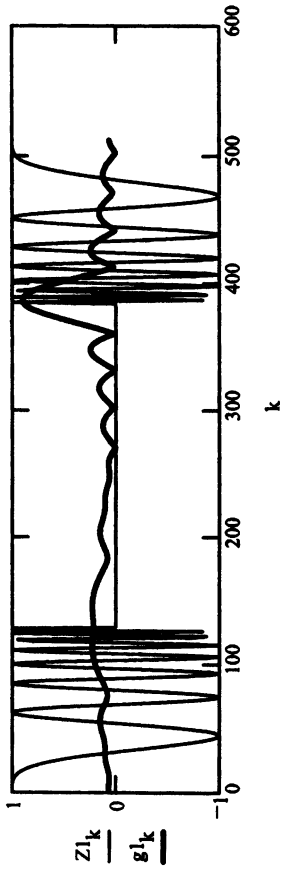


Рис. 4П

## Получение частотной характеристики свободного пространства на основе его отклика

Массив данных и их дискретность

$L := 256$      $k := 0..L-1$      $n := 0..L-1$      $v := 0..0.5 \cdot L-1$      $u := 0..0.5 \cdot L-1$

$z := 2$     Расстояние в длинах волн     $p := 4$     Длина волны

$$x_k := \frac{k - 0.5 \cdot L}{p} \qquad y_n := \frac{n - 0.5 \cdot L}{p}$$

Координаты

$$R_{k,n} := \sqrt{z^2 + (x_k)^2 + (y_n)^2}$$

$$U1_k := \frac{k - 0.5 \cdot L}{L} \cdot 2 \cdot \pi \cdot p \qquad U2_n := \frac{n - 0.5 \cdot L}{L} \cdot 2 \cdot \pi \cdot p$$

$$H_{k,n} := -i \cdot \frac{\exp(i \cdot 2 \cdot \pi \cdot R_{k,n}) \cdot z \cdot p^2}{R_{k,n}} + \frac{1}{R_{k,n}} \cdot \frac{p^2}{2 \cdot \pi \cdot (R_{k,n})^2} \cdot \exp(i \cdot 2 \cdot \pi \cdot R_{k,n}) \cdot \frac{z}{R_{k,n}}$$

Точное выражение для отклика

$$E_{v,u+0.5L} := H_{v+0.5L,u}$$

$$E_{v+0.5L,u} := H_{v,u+0.5L}$$

Нормализация матрицы

$$E_{v,u} := H_{v+0.5L,u+0.5L}$$

$$E_{v+0.5L,u+0.5L} := H_{v,u}$$

$$UR_k := \frac{U1_k}{2 \cdot \pi}$$

$$T := \text{cffft}(E)$$

$$HG_{k,n} := \text{Re}(T_{k,n})$$

$$E1_{v,u+0.5L} := HG_{v+0.5L,u}$$

$$E1_{v+0.5L,u} := HG_{v,u+0.5L}$$

$$E1_{v,u} := HG_{v+0.5L,u+0.5L}$$

$$E1_{v+0.5L,u+0.5L} := HG_{v,u}$$

$$N1_k := E1_{k,0.5L}$$

$$TW_{k,n} := \exp \left[ i \cdot z \cdot \sqrt{(2 \cdot \pi)^2 - (U1_k)^2 - (U2_n)^2} \right]$$

ЧХ свободного пространства

$$NW_k := \text{Re}(TW_{k,0.5L})$$

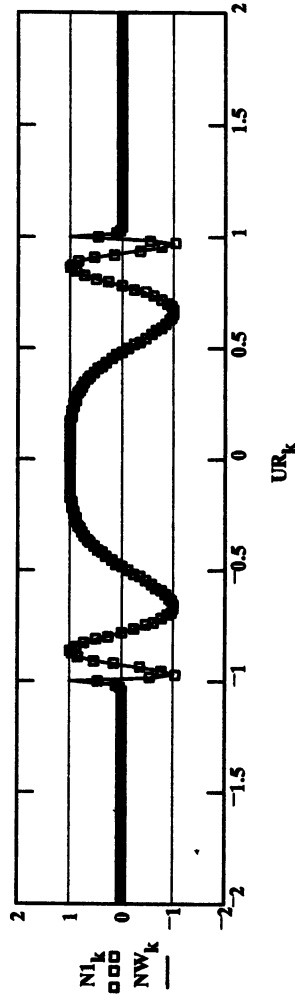


Рис. 5П



## Обращение волнового фронта

```

L := 2048      k := 0..L-1      m := 0..0.5·L-1      p := 16      r := 10
Пространственная частота
u_k := 2·π·(k - 0.5·L)·p

Источник сферической волны      Ширина щели      ЧХ свободного пространства
M := L/8
O(x) := | L if x=0.5·L
        | 0 otherwise
Запись ЧХ в виде пространственного спектра
p_k := O(k)      z1_m := z_0.5·L+m      z1_0.5·L+m := z_m
Спектр поля на расстоянии r
g := cfft(p)      gr_k := g_k·z1_k      КА волны на расстоянии r
a := icfft(gr)

Формирование поля щелью ширины M
O1(x) := | a if 0.5·L - 0.5·M < x < 0.5·L + 0.5·M
         | 0 otherwise
Вычисление поля без ОВФ
b_k := O1(k)      gr1_k := g1_k·z1_k      br := icfft(gr1)
Вычисление поля без ОВФ
b1_k := Re(b_k)
    
```

Вычисление поля после ОВФ

$$gr3_k := gr2_k \cdot z1_k \quad bro := icfft(gr3)$$

$$B_k := Re(bro_k) + 200$$

$$br1_k := Re(br_k) + 100$$

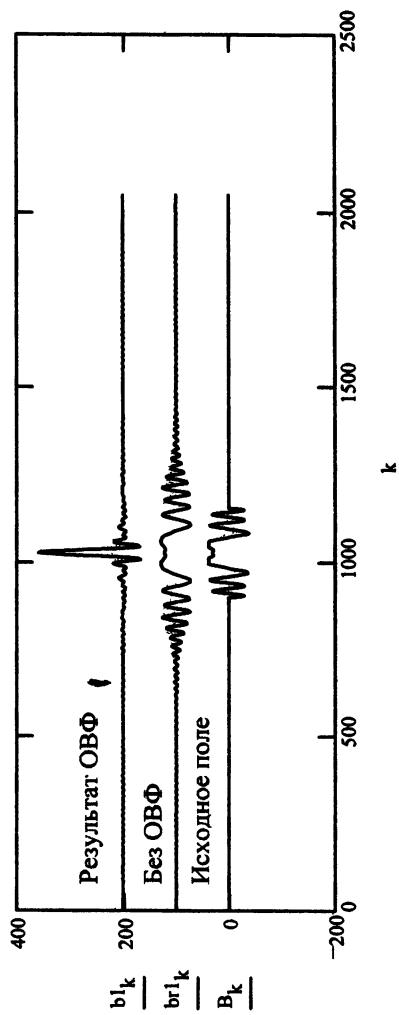


Рис. 6П

## Голограмма Габора

$L := 2048$     $k := 0..L-1$     $m := 0..0.5 \cdot L-1$     $p := 16$     $r := 10$    Пространственная частота  
 $u_k := 2 \cdot \frac{\pi}{L} \cdot (k - 0.5 \cdot L) \cdot p$   
 Источник сферической волны    $M := \frac{L}{8}$    ширина щели   ЧХ свободного пространства  
 $O(x) := \begin{cases} L & \text{if } x = 0.5 \cdot L \\ 0 & \text{otherwise} \end{cases}$     $z_k := \exp\left[i \cdot r \cdot \sqrt{(2 \cdot \pi)^2 - (u_k)^2}\right]$   
 $p_k := O(k)$    Спектр поля на расстоянии  $r$     $z1_m := z_{0.5 \cdot L + m}$     $z1_m := z_m$   
 $g := \text{cff}(p)$     $gr_k := g_k \cdot z1_k$    КА волны на расстоянии  $r$     $a := \text{icfft}(gr)$   
 Формирование поля щелью ширины  $M$   
 $O1(x) := \begin{cases} a & \text{if } 0.5 \cdot L - 0.5 \cdot M < x < 0.5 \cdot L + 0.5 \cdot M \\ 0 & \text{otherwise} \end{cases}$   
 $b_k := O1(k)$

Запись ЧХ в виде пространственного спектра

Запись голограммы Габора

$$A_k := \left( \left| b_k \cdot 10^{-4} + 1 \right| \right)^2$$

Восстановление сигнала с голограммы Габора

$g1 := \text{cfft}(A)$

$gr1_k := g1_k \cdot z1_k$

$br := \text{icfft}(gr1)$

$B_k := \text{Re}(br_k) + 0.015$

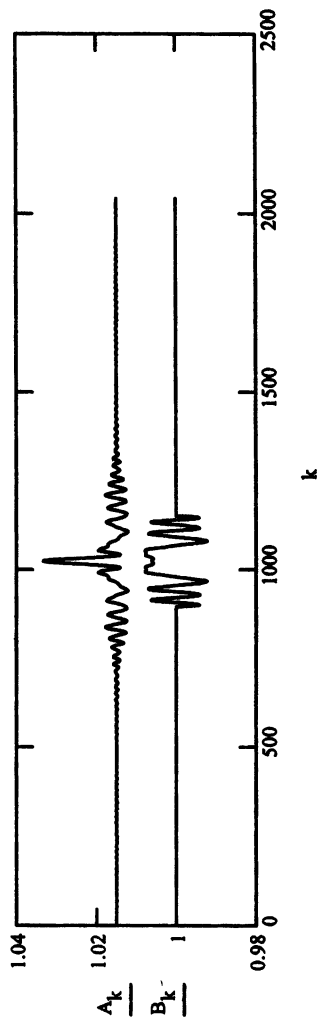


Рис. 7П

## Голограмма Лейта

$L := 2048$     $k := 0..L - 1$     $m := 0..0.5 \cdot L - 1$     $p := 16$     $r := 10$

Пространственная частота

$$u_k := 2 \cdot \frac{\pi}{L} \cdot (k - 0.5 \cdot L) \cdot p$$

$M := \frac{L}{8}$    Ширина щели

ЧХ свободного пространства

$$z_k := \exp \left[ i \cdot r \cdot \sqrt{(2 \cdot \pi)^2 - (u_k)^2} \right]$$

Источник сферической волны

$$O(x) := \begin{cases} L & \text{if } x = 0.5 \cdot L \\ 0 & \text{otherwise} \end{cases}$$

Запись ЧХ в виде пространственного спектра

$P_k := O(k)$

Спектр поля на расстоянии  $r$

$$z1_m := z_{0.5 \cdot L + m} \quad z1_{1+m} := z_m$$

$g := \text{cffit}(P)$

$$g r_k := g_k \cdot z1_k$$

КА волны на расстоянии  $r$

Формирование поля щелью ширины  $M$

$$a := \text{isfft}(g r)$$

$$O1(x) := \begin{cases} a_x & \text{if } 0.5 \cdot L - 0.5 \cdot M < x < 0.5 \cdot L + 0.5 \cdot M \\ 0 & \text{otherwise} \end{cases}$$

$$f := \frac{0.2}{p}$$

Угол опорной волны

Опорная волна

$b_k := O1(k)$

$$A2_k := \exp(i \cdot u_k \cdot 0.5 \cdot L \cdot f)$$

Запись голограммы Лейта

$$A_k := \left( b_k \cdot 10^{-2} + A2_k \right)^2$$

Считывание голограммы опорной волной  
к.с. без к.с.

$$A5_k := A_k \cdot A2_k \quad A6_k := A_k \cdot A2_k$$

Восстановление сигнала с голограммы  
к.с. опорной волной

$$g1 := \text{cfft}(A5) \quad gr1_k := g1_k \cdot z1_k$$

$$br := \text{icfft}(gr1)$$

$$B_k := \text{Re}(br_k) + 3.5$$

Опорной волной без к.с.

$$g2 := \text{cfft}(A6) \quad gr2_k := g2_k \cdot z1_k$$

$$br2 := \text{icfft}(gr2)$$

$$B2_k := \text{Re}(br2_k) + 2.5$$

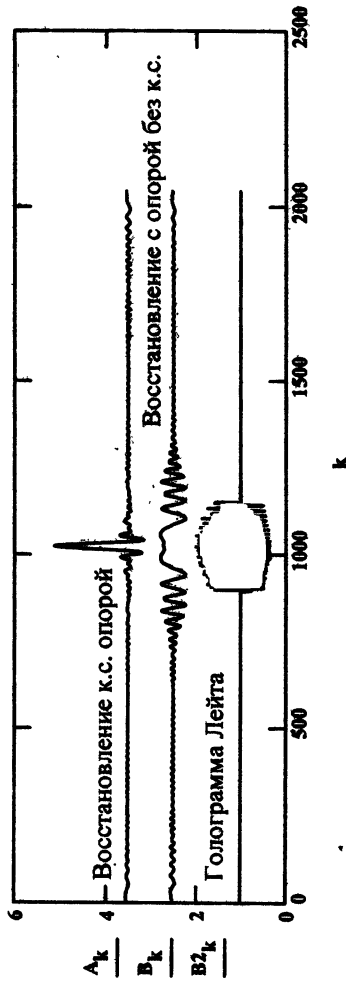


Рис. 8П

## Спекл-шум

Массив данных

$L := 256$      $k := 0 .. L - 1$

Глубина фазовой модуляции

$M := 0.2$

Случайная последовательность

$U := \text{rngnm}(L, 0, 1)$

Фильтр с жесткими границами

$$F(x) := \begin{cases} 1 & \text{if } x < 0.25 \cdot L \\ 1 & \text{if } L - 0.25 \cdot L \leq x < L \\ 0 & \text{otherwise} \end{cases}$$

Сигнал, модулированный только по фазе

$$W_k := \exp(i \cdot 2 \cdot \pi \cdot M \cdot U_k)$$

Преобразование Фурье

$$G := \text{cfft}(W)$$

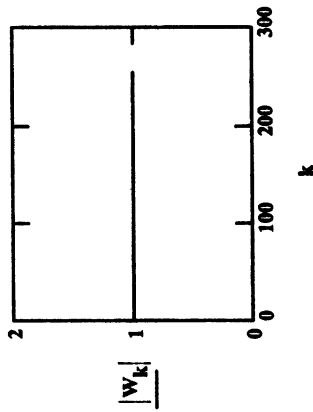
Ограничение полосы частот сигнала

$$G1_k := G_k \cdot F(k)$$

Обратное преобразование Фурье от сигнала с ограниченной полосой частот (выход фильтра)

$$S := \text{icfft}(G1)$$

Модуль сигнала, модулируемого только по фазе



Модуль сигнала, модулированного только по фазе, с ограничением полосы частот (стекл-шум)

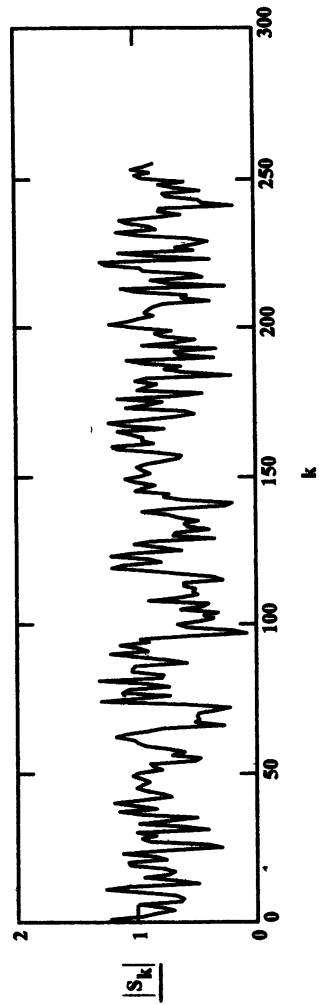


Рис. 9П



## Механизм возникновения спекла

```

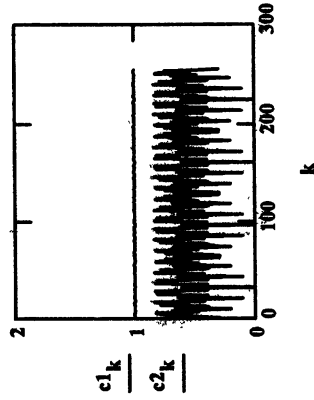
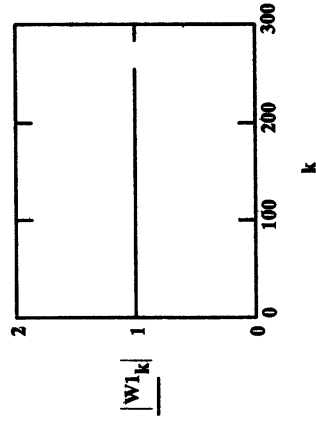
L := 256      k := 0..L-1      n := 0..0.5·L-1      f1 := 10      f2 := 70
U1k := cos(2· $\frac{\pi}{L}$ ·k·f1)      W1k := exp(i·2· $\pi$ ·0.5·U1k)      F(x) :=  $\begin{cases} 1 & \text{if } x < 0.25 \cdot L \\ 1 & \text{if } L - 0.25 \cdot L \leq x < L \\ 0 & \text{otherwise} \end{cases}$ 
U2k := cos(2· $\frac{\pi}{L}$ ·k·f2)      W2k := exp(i·2· $\pi$ ·0.5·U2k)

```

```

G1 := cfft(W1)      ak := |W1k|      bk := |W2k|
G2 := cfft(W2)      T1k := G1k·F(k)      S1 := left(T1)      c1k := |S1k|
T2k := G2k·F(k)      S2 := left(T2)      c2k := |S2k|

```



$$H_k := F(k) \cdot 5$$

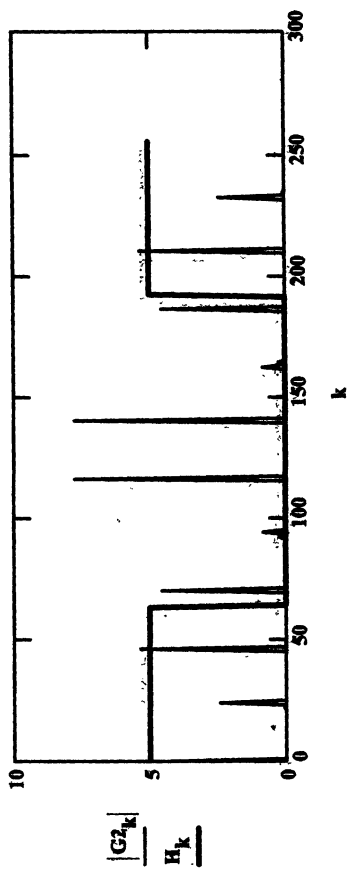
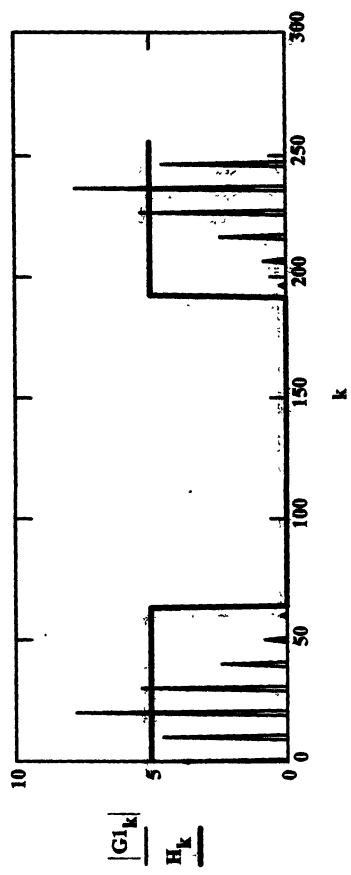


Рис. 10П

## Спектр функции (1.40)

Диапазонные переменные	Параметры
$L := 512$	$k := 0..L - 1$
$n := 0..L - 1$	$n := 0..0.5 \cdot L - 1$
$M := \frac{L}{8}$	$M := \frac{L}{8}$
$f := 100$	$f := 100$
$A := 1$	$A := 1$
$\pi := 0.05$	$\pi := 0.05$
$R := 400$	$R := 400$
$u := 2 \cdot \frac{\pi}{L} \cdot f$	$u := 2 \cdot \frac{\pi}{L} \cdot f$

$$O(x) := \begin{cases} 1 & \text{if } 0.5 \cdot L - M \leq x < 0.5 \cdot L + M \\ 0 & \text{otherwise} \end{cases}$$

Сферическая волна

$$r_k := \sqrt{R^2 + (k - 0.5 \cdot L)^2}$$

$$p_k := \exp\left(i \cdot 2 \cdot \frac{\pi}{2} \cdot r_k\right) \cdot O(k) \cdot \frac{R}{r_k}$$

Функция (1.42)

$$E_k := \overline{p_k} \cdot p_k + A^2 + A \cdot \exp(i \cdot u \cdot k) \cdot p_k + A \cdot \exp(i \cdot u \cdot (-k)) \cdot \overline{p_k}$$

Спектр

$$G := \text{cfft}(E)$$

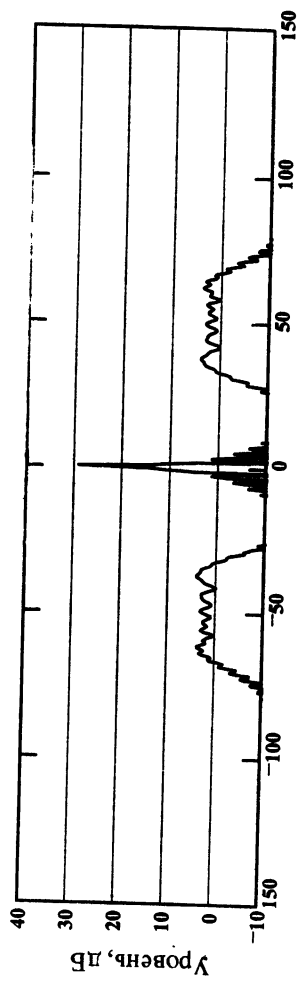
Представление результата

$$N_n := G_{0.5 \cdot L + n}$$

$$N_{0.5 \cdot L + n} := G_n$$

$$K_k := 20 \cdot \log(|N_k|)$$

$$w_k := \frac{k - 0.5 \cdot L}{2}$$



Пространственная частота

Рис. 11П

## Угловое сканирование антенны

```

Частота          Уровень шума
L := 512          f := 40.567      t := 1      U := gnorm(L, 0, 1)
k := 0..L - 1    x_k := 2 * pi / L * (k - 0.5 * L)  V := gnorm(L, 0, 1)
M := L / 16      Aпертура      tt := 10    Z(u) := 20 * log(|u| + 10^-4)
SH_k := U_k + i * V_k      Шум
n := 0..M - 1    A_k := exp(i * x_k * f) * tt + t * SH_k
m := 0..0.5 * L - M * 0.5 - 1
e := 0..0.5 * L - 1

```

Плоская волна и шум

Комплексная амплитуда волнового поля в плоскости раскрыва антенны

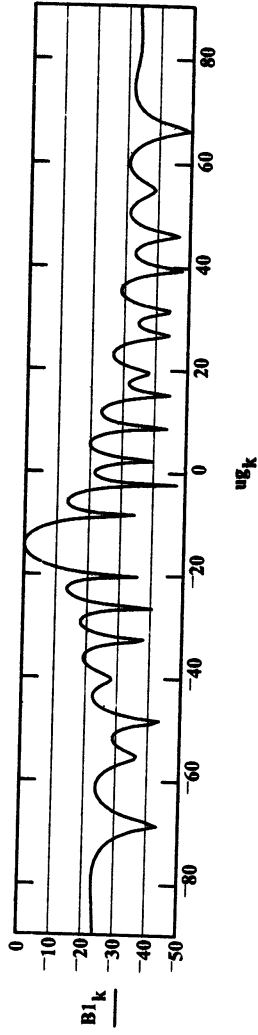
```

P_m := 0          P_0.5 * L - M * 0.5 + n := A_0.5 * L - M * 0.5 + n      P_0.5 * L + 0.5 * M + m := 0
G = cfft(P)      Сканирование по углам или пространственным частотам
B_e := Z(G_0.5 * L + e)      B_0.5 * L + e := Z(G_e)
hm := max(B)      B1_k := B_k - hm

```

Переход к градусам

$$ug_k := \text{asin}(0.5 \cdot x_k) \cdot \frac{180}{\pi}$$



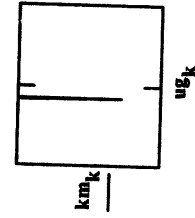
Результат углового сканирования

Процедура определения координаты максимума при угловом сканировании антенны

$$km_k = ug_k \cdot \Phi(B_k - h_{km})$$

Координата максимума в градусах

$$\max(km) + \min(km) = -14.571$$



$$ky_k := (k - 0.5 \cdot L) \cdot \Phi(B_k - \max(B))$$

Координата максимума в отсчетах от центра

$$\max(ky) + \min(ky) = -41$$

Рис. 12П

## Точное определение положения максимума углового отклика антенны

Заданное положение  
максимума отклика

$$L := 512$$

$$f := 40.4812$$

$$t := 5$$

$$U := \text{gnorm}(L, 0, 1)$$

$$k = 0..L - 1$$

$$x_k := 2 \cdot \frac{\pi}{L} \cdot (k - 0.5 \cdot L)$$

$$V := \text{gnorm}(L, 0, 1)$$

$$M := \frac{L}{16}$$

$$\text{SN}_k := U_k + i \cdot V_k$$

$$tt := 10$$

$$Z(u) := 20 \cdot \log(|u| + 10^{-4})$$

$$n := 0..M - 1$$

$$A_k := \exp(i \cdot x_k \cdot f) \cdot tt + t \cdot \text{SN}_k$$

$$m := 0..0.5 \cdot L - M \cdot 0.5 - 1$$

$$e := 0..0.5 \cdot L - 1$$

$$P_m := 0$$

$$P_{0.5 \cdot L - M \cdot 0.5 + n} := A_{0.5 \cdot L - M \cdot 0.5 + n}$$

$$P_{0.5 \cdot L + 0.5 \cdot M + m} := 0$$

$$\text{mm} := 0..4 \cdot M - 1$$

$$h_{\text{mm}} := \frac{\text{mm}}{64} - f - \frac{M}{32}$$

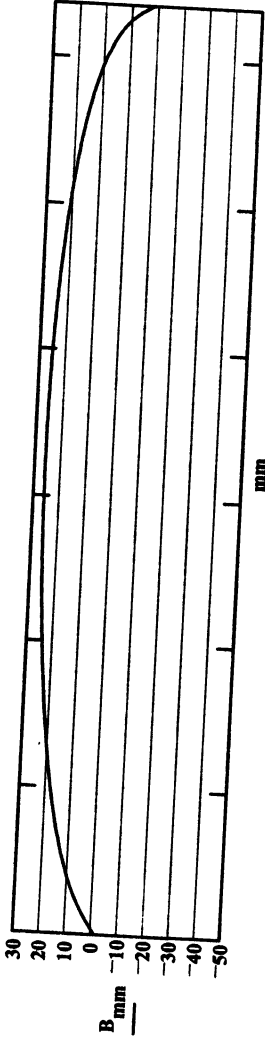
Шаг сканирования

$$G_{mm} := \frac{2}{L} \sum_k A_k \cdot \exp(i \cdot x_k \cdot h_{mm})$$

Угловое сканирование антенны

$$B_{mm} := Z(G_{mm})$$

$$ug_k := \text{asin}(0.5 \cdot x_k) \cdot \frac{180}{\pi}$$



$$ky_{mm} := h_{mm} \cdot \Phi(B_{mm} - \max(B))$$

Измеренная координата максимума

$$fm := \max(ky) - \min(ky)$$

$$f = 40.481$$

$$fm - f = 0$$

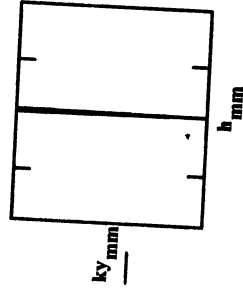


Рис. 13П





КА волнового поля в плоскости раскрыва антенны

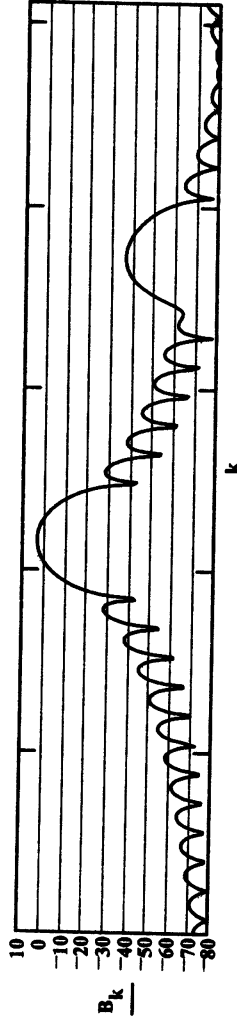
$$P_m := 0 \quad P_{0.5L - M \cdot 0.5 + n} := A_{0.5L - M \cdot 0.5 + n} \cdot U P_n \quad P_{0.5L + 0.5 \cdot M + m} := 0$$

$$G := \text{cfft}(P)$$

$$B_e := Z(G_{0.5L + e})$$

$$B_{0.5L + e} := Z(G_e)$$

$$ug_k := \text{asin}(0.5 \cdot x_k) \cdot \frac{180}{\pi}$$



Результат углового сканирования

Рис. 14П

## Слабая волна на фоне сильной без обработки совокупного сигнала

	Положение основной волны	Доля слабой волны	Положение слабой волны
$L := 512$	$f := 40.5$	$t := 0.01$	$f1 := 395.5$

$k := 0..L - 1$

$$x_k := 2 \cdot \frac{\pi}{L} \cdot (k - 0.5 \cdot L)$$

$M := \frac{L}{16}$

$$Z(u) := 20 \cdot \log(|u| + 10^{-4})$$

$n := 0..M - 1$

$$SS_k := \exp(i \cdot x_k \cdot f1)$$

Сумма двух волн

$$A_k := \exp(i \cdot x_k \cdot f) + t \cdot (SS_k)$$

$m := 0..0.5 \cdot L - M \cdot 0.5 - 1$

$e := 0..0.5 \cdot L - 1$

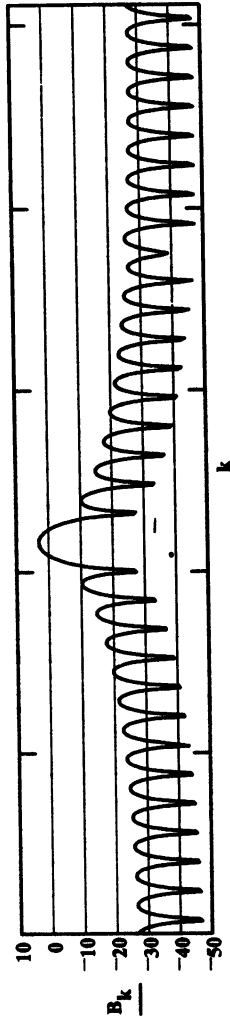
КА волнового поля в плоскости раскрыва антенны

$$P_m := 0 \quad P_{0.5.L-M.0.5+n} := A_{0.5.L-M.0.5+n} \quad P_{0.5.L+0.5.M+m} := 0$$

$$G := \text{cfft}(P)$$

$$B_e := Z(G_{0.5.L+e}) \quad B_{0.5.L+e} := Z(G_e)$$

$$u_{g_k} := \text{asin}(0.5 \cdot x_k) \cdot \frac{180}{\pi}$$



Результат углового сканирования

Рис. 15II

## Влияние шумов или ошибок считывания на диаграмме направленности антенны

Интенсивность  
шума

```

L := 512
k := 0..L-1
M :=  $\frac{L}{4}$ 

f := 40.5
t := 0.01
i :=  $\sqrt{-1}$ 
f1 := 395.5

xk :=  $2 \cdot \frac{\pi}{L} \cdot (k - 0.5 \cdot L)$ 
Z(u) := 20 · log(|u| + 10-4)

```

Шум или ошибки считывания

```

SHk := rnd(1) + i · rnd(1)
n := 0..M-1
UPn := 1 + cos  $\left[ 2 \cdot \frac{\pi}{M} \cdot (n - 0.5 \cdot L + M \cdot 0.5) \right]$ 

```

$$A_k := \exp(i \cdot x_k \cdot f) + t \cdot (SH_k - \text{mean}(SH))$$

$$m := 0..0.5 \cdot L - M \cdot 0.5 - 1$$

$$e := 0..0.5 \cdot L - 1$$

$$P_m := 0 \quad P_{0.5 \cdot L - M \cdot 0.5 + n} := A_{0.5 \cdot L - M \cdot 0.5 + n} \cdot UP_n$$

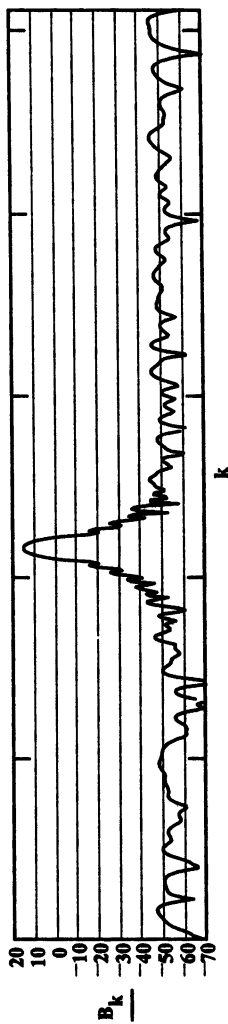
$$P_{0.5 \cdot L + 0.5 \cdot M + m} := 0$$

$$G := \text{cfft}(P)$$

$$B_e := Z(G_{(0.5 \cdot L + e)})$$

$$B_{0.5 \cdot L + e} := Z(G_e)$$

$$ug_k := \text{asin}(0.5 \cdot x_k) \cdot \frac{180}{\pi}$$



Результат углового сканирования

Рис. 16П

## Адаптивная антенна

Доля слабой  
волны

$\pi := 395.5$

$f := 40.5$

$i := \sqrt{-1}$

$L := 512$

$k := 0..L-1$

$x_k := 2 \cdot \frac{\pi}{L} \cdot (k - 0.5 \cdot L)$

$Z(u) := 20 \cdot \log(|u| + 10^{-5})$

$M := \frac{L}{16}$

$SS_k := \exp(i \cdot x_k \cdot \pi)$

$n := 0..M-1$

$A_k := \exp(i \cdot x_k \cdot f) + t \cdot (SS_k)$

$s := 0.5 \cdot L - 0.5 \cdot M$

$m := 0..0.5 \cdot L - M \cdot 0.5 - 1$

$fa := f$  Положение удаляемого сигнала

Комплексный множитель

$$T := \frac{1}{M} \sum_n A_{n+s} \cdot \exp[i \cdot x_{s+n} \cdot fa \cdot (-1)]$$

Устранение сигнала с частотой  $f_a$

$$A2_{s+n} := A_{s+n} - T \cdot \exp(i \cdot x_{s+n} \cdot f_a)$$

$$e := 0.05 \cdot L - 1$$

$$P_m := 0 \quad P_{0.5 \cdot L - M \cdot 0.5 + n} := A2_{s+n}$$

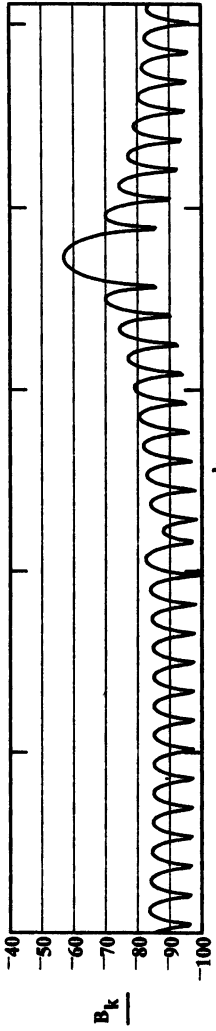
$$P_{0.5 \cdot L + 0.5 \cdot M + m} := 0$$

$$G = \text{cfft}(P)$$

$$B_e := Z(G_{0.5 \cdot L + e})$$

$$B_{0.5 \cdot L + e} := Z(G_e)$$

$$u g_k := \text{asin}(0.5 \cdot x_k) \cdot \frac{180}{\pi}$$



Результат углового сканирования

Рис. 17П



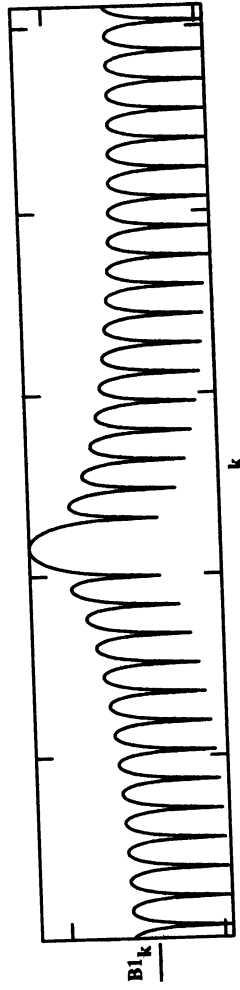
## Антенна с адаптивным исключением самого сильного сигнала

Положение максимума

$$\begin{aligned}
 L &:= 512 & f &:= 40.812 & t &:= 1 & i &:= \sqrt{-1} & fl &:= 395.5 \\
 k &:= 0..L - 1 & x_k &:= 2 \cdot \frac{\pi}{L} \cdot (k - 0.5 \cdot L) & tt &:= 0.001 & Z(u) &:= 20 \cdot \log(|u| + 10^{-4}) \\
 M &:= \frac{L}{16} & SN_k &:= \exp[i \cdot x_k \cdot f \cdot (1 + 0.003)] & SS_k &:= \exp(i \cdot x_k \cdot fl) & A_k &:= \exp(i \cdot x_k \cdot f) + t \cdot (SN_k) + tt \cdot SS_k \\
 n &:= 0..M - 1 & & & & & & & & \\
 m &:= 0..0.5 \cdot L - M \cdot 0.5 - 1 & & & & & & & & \\
 e &:= 0..0.5 \cdot L - 1 & & & & & & & & \\
 P_m &:= 0 & P_{0.5 \cdot L - M \cdot 0.5 + n} &:= A_{0.5 \cdot L - M \cdot 0.5 + n} & & & & & & P_{0.5 \cdot L + 0.5 \cdot M + m} := 0
 \end{aligned}$$

$G1 = \text{cfft}(P)$  Сканирование по углам или пространственным частотам

$$B1_k = Z(G1_{0.5 \cdot L + e}) \quad B1_{0.5 \cdot L + e} := Z(G1_e)$$



Результат углового сканирования

Процедура определена; координаты максимума при угловом сканировании антенны

Координата максимума в отсчетах от центра

$$ky1_k := (k - 0.5 \cdot L) \cdot \Phi(B1_k - \max(B1))$$

$$fm1 := \max(ky1) - \min(ky1)$$

$$f = 40.812$$

$$fm1 = 41$$

Рис. 18II (начало)

Уточнение положения максимума

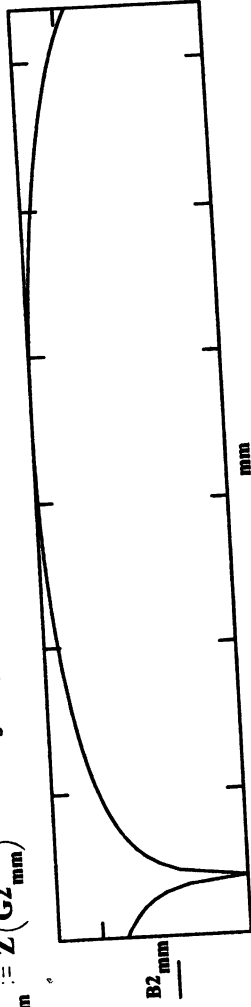
$$mm := 0.4 \cdot M - 1$$

$$h_{mm} := \left( fm1 + \frac{M}{32} \right) - \frac{mm}{64}$$

Угловое сканирование антенны

$$G2_{mm} := \frac{2}{L} \cdot \sum_k A_k \cdot \exp[i \cdot (-1) \cdot x_k \cdot h_{mm}]$$

$$B2_{mm} := Z(G2_{mm}) \quad s := 0.5 \cdot L - 0.5 \cdot M$$



$$ky2_{mm} := h_{mm} \cdot \Phi(B2_{mm} - \max(B2))$$

Измеренная координата максимума

$$fm2 := \max(ky2) - \min(ky2) \quad f = 40.812 \quad fm2 = 40.875$$

**fa** := fm2

Положение устрояемого сигнала

Комплексный множитель

$$T := \frac{1}{M} \cdot \sum_{n+s} A_{n+s} \cdot \exp[i \cdot x_{s+n} \cdot fa \cdot (-1)]$$

Устранение сигнала с частотой fa

$$A2_{s+n} := A_{s+n} - T \cdot \exp(i \cdot x_{s+n} \cdot fa)$$

$$e := 0 .. 0.5 \cdot L - 1$$

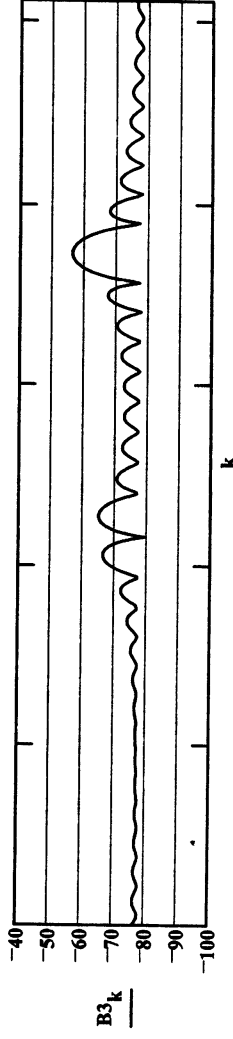
$$P_m := 0 \quad P_{0.5 \cdot L - M \cdot 0.5 + n} := A2_{s+n}$$

$$P_{0.5 \cdot L + 0.5 \cdot M + m} := 0$$

$$G3 := \text{cfft}(P)$$

$$B3_e := Z(G3_{0.5 \cdot L + e})$$

$$B3_{0.5 \cdot L + e} := Z(G3_e)$$



Результат углового сканирования

Рис. 18П (окончание)



Фильтр

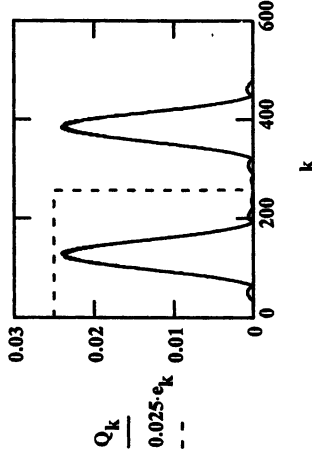
$$Q_{nn} := |G3_{0.5 \cdot LM + nn}| \cdot 10 \quad Q_{0.5 \cdot LM + nn} := |G3_{nn}| \cdot 10$$

$$e_k := \Phi(0.5 \cdot LM \cdot 1 - k)$$

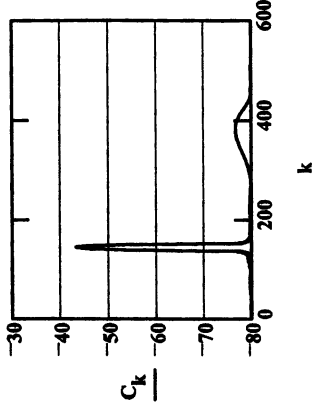
Оптическая обработка действительной части выхода антенны

$$G4_k := G3_k \cdot ZU_k \cdot e_k \quad G5 := \text{icfft}(G4)$$

$$C_k := ZW(G5_k) \quad d = 12$$



Спектр изображения  
и фильтр



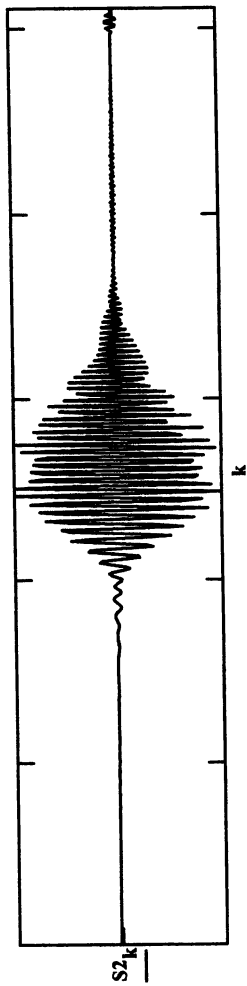
Изображение

Положение изображения

$$CM_k := k \cdot \Phi(C_k - \max(C)) \quad \max(CM) + 0.5 \cdot L = 153$$

Рис. 19П (начало)

Сигнал, считанный антенной при ее пространственном сканировании  
(изображение источника без апертурного синтеза)



Пространственное распределение сигнала в раскрыве антенны

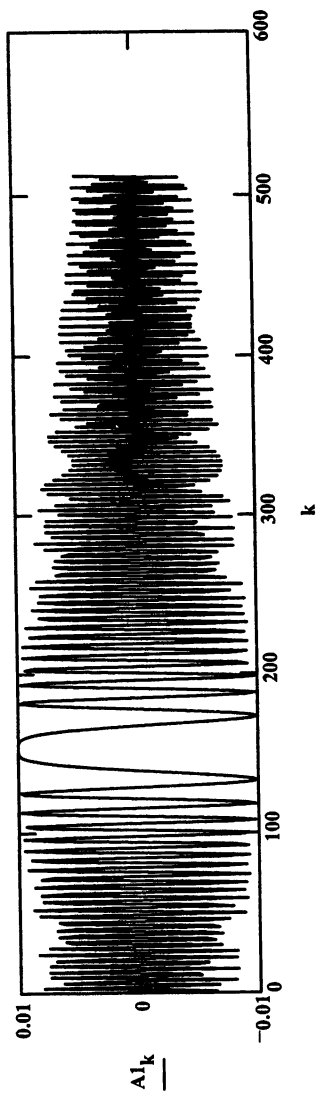


Диаграмма направленности антенны

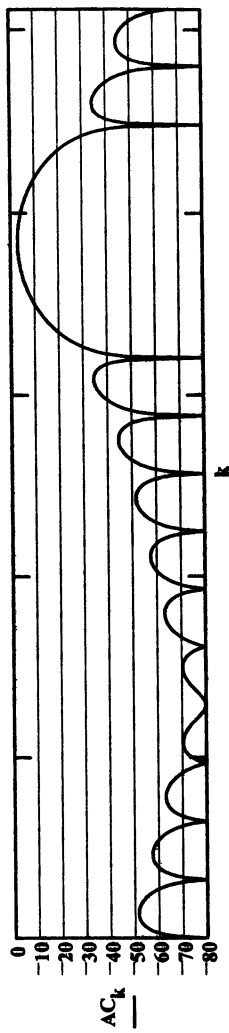


Рис. 19II (окончание)



## Преобразование полей оптической системой

```

L := 2048  k := 0..L - 1  M :=  $\frac{L}{8}$   n := 0..M - 1  m := 0..0.5·M - 1
p :=  $\frac{1}{2}$   z := 10·M  F := z  f := 150
H(u) := exp(i ·  $\frac{\pi}{z} \cdot u^2$ )  x_k := k - 0.5·L  W_k := 1 + cos(2 ·  $\frac{\pi}{L} \cdot k \cdot f$ )
O(x) :=  $\begin{cases} 1 & \text{if } 0.5 \cdot L - 0.5 \cdot M < x < 0.5 \cdot L + 0.5 \cdot M \\ 0 & \text{otherwise} \end{cases}$ 
D_k := O(k)  R(u) := exp(-i ·  $\frac{\pi}{p \cdot F} \cdot u^2$ )  t_n := 0.5·M - n

```

M – апертура

z – расстояние

F – фокусное расстояние

H – отклик пространства

R – объектив

p – длина волны

O(x) – формирование

конечной апертуры

W – входной сигнал

L – массив данных

k, n, m – диапазонные

переменные

Формула (3.1)

$$P_k := D_k \cdot R(x_k) \cdot W_k \cdot P1_{0.5M+m} := \sum_k P_k \cdot H(x_k - x_{m+0.5L}) \quad P1_{0.5M-m} := \sum_k P_k \cdot H(x_k - x_{0.5L-m})$$

$$R1(u) := \exp\left(-i \cdot \frac{\pi \cdot 2}{p \cdot F} \cdot u^2\right)$$

Основной объектив оптической системы с фокусным расстоянием F/2

$$O1(x) := \begin{cases} P1_{x-0.5L+0.5M} & \text{if } 0.5L - 0.5M < x < 0.5L + 0.5M \\ 0 & \text{otherwise} \end{cases}$$

$$SR_k := O1(k)$$

P2 – поле в плоскости объектива

$$P2_k := SR_k \cdot R1(x_k)$$

Формула (3.1)

$$P3_{0.5M+m} := \sum_k P2_k \cdot H(x_k - x_{m+0.5L}) \quad P3_{0.5M-m} := \sum_k P2_k \cdot H(x_k - x_{0.5L-m})$$

Представление полей в исходной плоскости, плоскости объектива и в выходной плоскости

$$A_n := |P3_n|$$

$$A2_n := W_{0.5L-0.5M+n} \cdot 10^2 \cdot 4 \cdot \pi$$

$$A1_n := |SR_{0.5L-0.5M+n}|$$

Рис. 20П (начало)

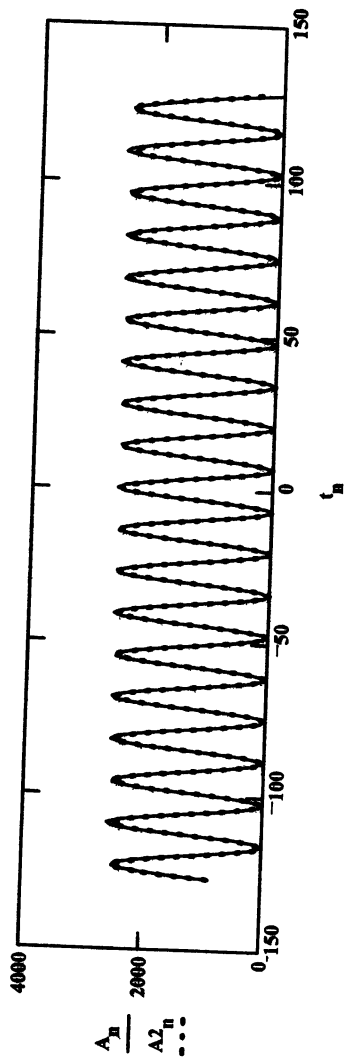
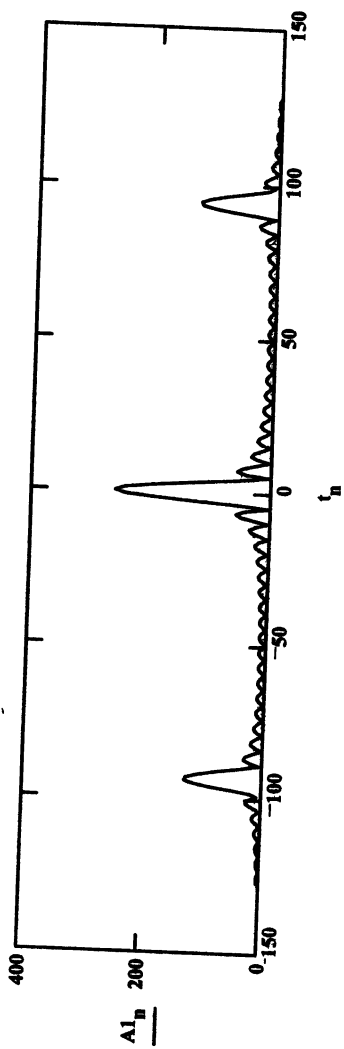


Рис. 20П (окончание)

## Преобразование полей оптической системой

Быстрая программа

```

L := 2048   k := 0..L - 1   M :=  $\frac{L}{8}$    n := 0..M - 1   m := 0..0.5·M - 1   rr := 0..0.5·L - 1
p :=  $\frac{1}{2}$      z := 10·M       F := z     tn := 0.5·M - n     f := 150

```

$$H(u) := \exp\left(i \cdot \frac{\pi}{z \cdot p} \cdot u^2\right) \quad x_k := k - 0.5 \cdot L$$

$$W_k := 1 + \cos\left(2 \cdot \frac{\pi}{L} \cdot k \cdot f\right)$$

$$W1_n := |W_{0.5 \cdot L - 0.5 \cdot M + n}|$$

$$O(x) := \begin{cases} 1 & \text{if } 0.5 \cdot L - 0.5 \cdot M < x < 0.5 \cdot L + 0.5 \cdot M \\ 0 & \text{otherwise} \end{cases}$$

Условие правильной свертки  
комплексных функций

$$HH_k := H(x_k)$$

$$HH1_k := \overline{HH_k}$$

$$D_k := O(k) \quad R(u) := \exp\left(-i \cdot \frac{\pi}{p \cdot F} \cdot u^2\right)$$

$$GH := \text{cfft}(HH1)$$

$$U3 := \max(W1)$$

Рис. 21П (начало)

$$\begin{aligned}
P_k &:= D_k \cdot R(x_k) \cdot W_k & GP &:= \text{cfft}(P) & A_k &:= GP_k \cdot \overline{GH_k} & T &:= \text{icfft}(A) \\
T_{rr} &:= T_{0.5L+rr} & T_{rr} &:= T_{rr} & R1(u) &:= \exp\left(-i \cdot \frac{\pi \cdot 2 \cdot u^2}{p \cdot F}\right) \\
O1(x) &:= \begin{cases} T1_x & \text{if } 0.5 \cdot L - 0.5 \cdot M < x < 0.5 \cdot L + 0.5 \cdot M \\ 0 & \text{otherwise} \end{cases} \\
S_k &:= O1(k) & P2_k &:= S_k \cdot R1(x_k) & GP2 &:= \text{cfft}(P2) & A2_k &:= GP2_k \cdot \overline{GH_k} \\
T2 &:= \text{icfft}(A2) & T3_{rr} &:= T2_{0.5L+rr} & T3_{0.5L+rr} &:= T2_{rr} & U2 &:= \max(T5) \\
T5_n &:= \left| T3_{0.5L-0.5M+n} \right| & T4_n &:= \left| T1_{0.5L-0.5M+n} \right| & U1 &:= \max(T4) \\
JP1_n &:= \frac{T4_n}{U1} & JP2_n &:= \frac{T5_n}{U2} & JP3_n &:= \frac{W1_n}{U3}
\end{aligned}$$

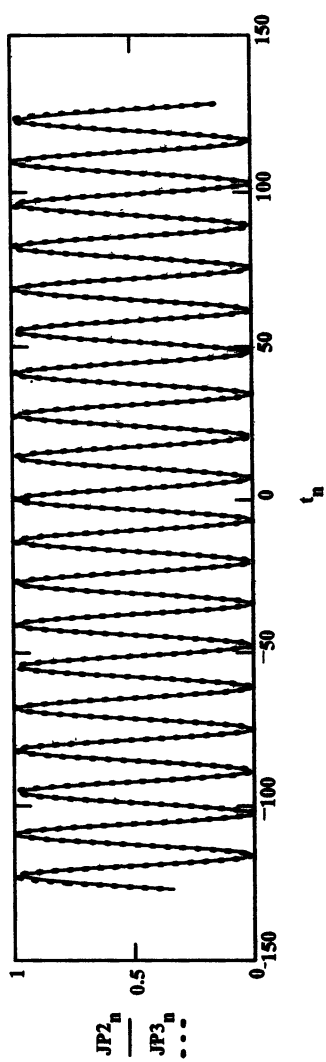
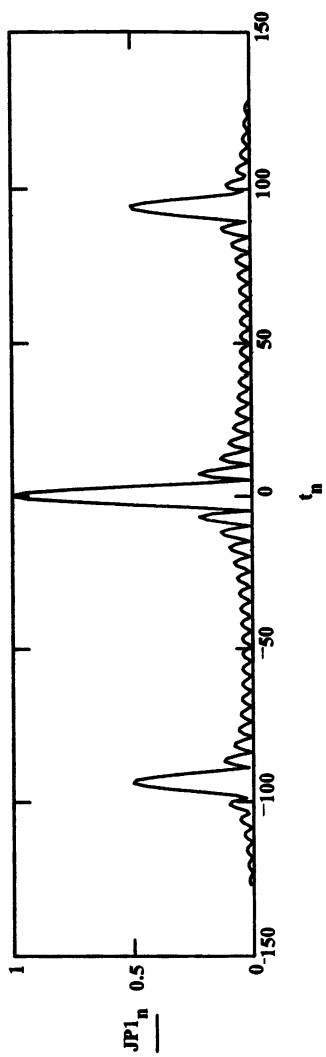


Рис. 21П (окончание)

## Согласованная фильтрация

Ввод данных

$L := 1024$     $k := 0..L-1$     $n := 0..0.5 \cdot L-1$     $U := \text{morm}(L, 0, 1)$     $V := \text{morm}(L, 0, 1)$

Формирование входных сигналов

$x_k := U_k$     $y_k := V_k$

Формирование ЧХ фильтра, согласованного с сигналом  $x(k)$

$g := \text{cfft}(x)$     $G_k := \overline{g_k}$

Отклик фильтра на "свой" сигнал

Отклик фильтра  $G$  на "свой" сигнал

$gg_k := g_k \cdot G_k$     $X := \text{icfft}(gg)$

Смесь "своего" и "чужого" сигналов

$e_k := g_k + \gamma \cdot gy_k$

Отклик фильтра на "чужой" сигнал

$gy := \text{cfft}(y)$

$gggy_k := gy_k \cdot G_k$

$Y := \text{icfft}(gggy)$

$\gamma := 3$

Отклик фильтра на смесь сигналов

$$g_k := e_k \cdot G_k \quad E := \text{icfft}(ge)$$

Преобразования матриц к виду, удобному для наблюдения формы отклика

$$Y1_n := Y_{n+0.5 \cdot L} \quad Y1_{n+0.5 \cdot L} := Y_n \quad t_k := k - 0.5 \cdot L$$

$$X1_n := X_{n+0.5 \cdot L} \quad X1_{n+0.5 \cdot L} := X_n$$

$$E1_n := E_{n+0.5 \cdot L} \quad E1_{n+0.5 \cdot L} := E_n$$

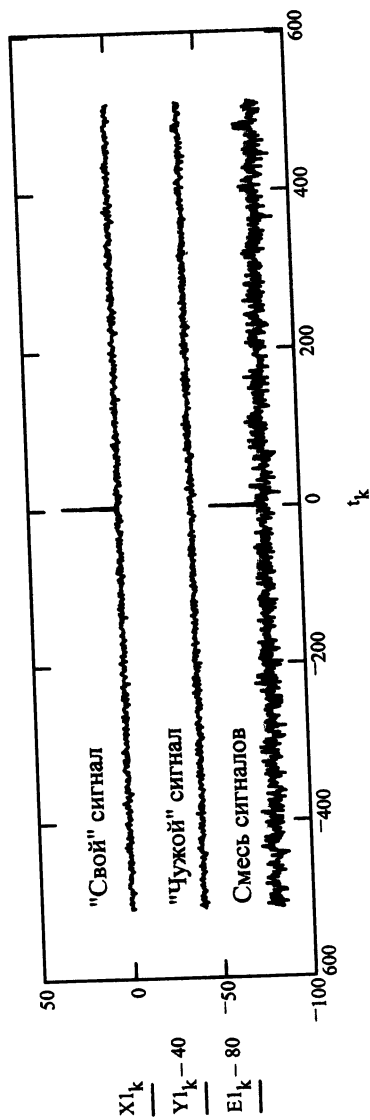


Рис. 22II



## Литература

1. *Зверев В. А.* Радиооптика. М.: Сов. радио, 1975.
2. *Зверев В. А., Орлов Е. Ф.* Оптические анализаторы. М.: Сов. радио, 1971.
3. *Зверев В. А., Славинский М. М.* Метод расчета акустического поля вблизи неровной поверхности // Акустический журнал. 1997. Т. 43, № 1. С. 67–72.
4. *Зверев В. А., Матвеев А. Л., Славинский М. М., Стромков А. А.* Фокусируемая антенна темного поля // Акустический журнал. 1997. Т. 43, № 4. С. 501–507.
5. *Zverev V. A., Matveyev A. L., Slavinsky M. N., Stromkov A. A.* Acoustic Array Noises in a shallow Water // Proc. I. O. A. 1996. V. 18, part 5. P. 85–91.
6. *Мандельштам Л. И.* Лекции по избранным вопросам оптики // Полное собрание трудов / Под ред. М. А. Леонтовича. М.: Изд-во АН СССР, 1950. Т. 5.
7. *Рытов С. М.* Введение в статистическую радиофизику. М.: Наука, 1966.
8. *Зверев В. А.* Исследование рассеяния акустических волн в присутствии интенсивных помех // Акустический журнал. 1994. Т. 40, вып.2. С. 263–266.
9. *Зверев В. А.* Антенна темного поля // Акустический журнал. 1994. Т. 40, № 3. С. 401–404.
10. *Зверев В. А.* Нелинейная фильтрация: Доклад на юбилейной научной конференции радиофака // Современные проблемы радиофизики: Сборник научных трудов / Под ред. А. В. Якимова. Н. Новгород: Изд-во ННГУ, 1966. 192 с.
11. *Зверев В. А., Хилько А. И., Шишарин А. В.* К вопросу об использовании источников некогерентного света в схемах пространственной фильтрации изображений // Автометрия. 1978. 108 с.

12. Mathcad 6.0 plus. Финансовые, инженерные и научные расчеты в среде Windows 95 / Пер. с англ. М.: Информационно-издательский дом "Филинь", 1996. 712 с.

13. Дьяконов В. П. Справочник по алгоритмам на языке Бейсик для персональных ЭВМ. М.: Наука, 1987. 239 с.

14. Монзинго Р. А., Миллер Т. У. Адаптивные антенные решетки. М.: Радио и связь, 1986.

15. Борн М., Вольф Э. Основы оптики. М.: Наука, 1970.

16. Гудман Дж. Введение в Фурье оптику. М.: Мир, 1970.

17. Сороко Л. М. Основы голографии и когерентной оптики. М.: Наука, 1971.

18. Строук Дж. Введение в когерентную оптику и голографию. М.: Мир, 1967.

19. Lugt A. V. Operational notation for the analysis and synthesis of optical data-processing systems // Proc. IEEE. 1966. V. 54, № 8. P. 1055-1063.

20. Петров М. П. Фоточувствительные электрооптические среды в голографии и оптической обработке информации. Л.: Наука, Ленинградское отделение, 1983.

21. Новаковский С. В., Котельников А. В. Новые системы телевидения. Цифровые методы обработки видеосигналов. М.: Радио и связь, 1992. 89 с.

22. ПЗС и системы на их основе. V научно-техническая конференция ПЗС-95 // Оптическая техника. 1996, № 1. С. 45.

23. Lugt A. V. Signal detection by complex spatian filtering // IEEE Trans. Inform. Theory. 1964. V. IT-10. P. 139-145.

24. Leith E. N. Quasi-Holographic techniques in the microwave region // Proc. of the IEEE. 1971. V. 59, № 9. P. 1305-1318.

25. Малахов А. Н. Флуктуации в автоколебательных системах. М.: Наука. Физматгиз. 1968.

26. Нейл О. Введение в статистическую оптику. М.: Мир, 1966.

27. Ryle M. The new Cambridge radio telescope // Nature. 1962. V. 194. P. 517-518.

28. Рытов С. М. Введение в статистическую радиофизику. Часть 1. Случайные процессы. М.: Наука, 1976. С. 496.

29. Жогликов В. А., Кияшко Б. В. Многоканальный анализатор спектра. Авторское свидетельство 530 623 // Бюллетень изобретений. 1976. № 36. Авторское свидетельство 589 623 // Бюллетень изобретений. 1978. № 3.

30. Cherry E. C., Sayers B. A. Human correlator // J. Acoust. Soc. Am. 1956. V. 28, № 5. P. 889.

31. *Жестяников Л. А.* Исследование стереофонического звучания // Экспериментальная радиооптика / Под ред. В. А. Зверева и Н. С. Степанова. М.: Наука, 1979.

32. *Schroeder M. R., Gotlob D., Stebrasse K. F.* Comparative study of European concert halls: correlation of subjektive preference with geometric and acoustic parametrs // J. Acoust. Soc. Am. 1974. V. 56, № 4. P. 1195–1201.

33. *Schroeder M. R.* Improved quasy-stereophony and "colorless" artificial reverberation // J. Acoust. Soc. Am. 1961. V. 33, № 8, P. 1061–1064.

34. *Kozma A., Cristensen Ch. R.* Effects of speckle on resolution // JOSA. 1976. V. 66, № 11. P. 1257.

35. *Gerschman A. B., Turchin V. I., Zverev V. A.* Experimental results for location of moving underwater signal using adaptive beamforming // IEEE Trans. on Signal Processing. 1995. VSP-42, № 10. P. 2249–2257.

36. *Зверев В. А.* Шумы акустической антенны в волноводе // Акустический журнал (в печати).

37. *Зверев В. А., Литвак Н. В.* Численное моделирование случайных акустических сигналов // Акустический журнал (в печати).



## СОДЕРЖАНИЕ

<i>Введение</i> . . . . .	3
<i>Часть первая. Регулярные поля</i> . . . . .	7
1. Поля в свободном пространстве . . . . .	7
1.1. Параметры волнового поля, существенные для получения изображений . . . . .	8
1.2. Частотная характеристика и отклик свободного пространства . . . . .	14
1.3. В каких случаях и почему необходимо малоугловое приближение? . . . . .	18
1.4. Упрощения при малоугловом приближении . . . . .	23
1.5. Голография . . . . .	28
1.6. Зернистость, или спекл-шум . . . . .	38
1.7. Замечания о численном нахождении поля . . . . .	50
1.8. Примеры математических программ, иллюстрирующих формирование полей при их распространении в свободном пространстве . . . . .	60
2. Антенны . . . . .	70
2.1. Параметры антенны как системы, формирующей изображение . . . . .	70
2.2. Антенна с угловым сканированием . . . . .	76
2.3. Фокусируемая антенна . . . . .	82
2.4. Синтез апертуры . . . . .	88
2.5. Радиоголография . . . . .	88
2.6. Проблемы теории и техники антенн . . . . .	100
2.7. Математические программы по антенной тематике . . . . .	105
3. Изображения в оптических системах . . . . .	110
3.1. Оптическая система . . . . .	110
3.2. Преобразование полей оптической системой . . . . .	111
3.3. Идеальная оптическая система . . . . .	113
3.4. Изображения, формируемые идеальной оптической системой . . . . .	115
3.5. Реальная оптическая система – это идеальная система + фильтр . . . . .	118
3.6. Как оптическая система преобразует изображение . . . . .	120
<i>Часть вторая. Изображения в случайных полях</i> . . . . .	129
4. Оптическая система в случайном поле . . . . .	129
4.1. Какие поля считаются случайными . . . . .	129
4.2. Какие параметры случайного поля влияют на формирование изображения . . . . .	130
4.3. Виды функций когерентности . . . . .	132
4.4. Частотно-контрастная характеристика оптической системы . . . . .	134
4.5. Роль фокальной плоскости объектива в случайном поле . . . . .	138

4.6. Как вводится частотно-контрастная характеристика оптической системы в некогерентной оптике . . . . .	139
5. Апертурный синтез в случайном поле . . . . .	141
5.1. Функция когерентности в свободном пространстве . . . . .	141
5.2. Некогерентная голография . . . . .	142
5.3. Некогерентный апертурный синтез . . . . .	143
<i>Часть третья. Основы совершенствования</i>	
методов формирования изображений . . . . .	148
6. Апертурный синтез и бинауральный слух . . . . .	148
6.1. Характерные особенности различных методов формирования изображений . . . . .	148
6.2. Физическая модель бинаурального слуха . . . . .	150
6.3. Естественная и художественная стереофония . . . . .	153
6.4. Физические основы естественной стереофонии . . . . .	155
6.5. Художественная стереофония . . . . .	158
6.6. Физические основы художественной стереофонии . . . . .	160
6.7. Бинауральная голография . . . . .	162
6.8. Объемное и стереообъемное звучание . . . . .	165
6.9. Чем отличается естественное звучание . . . . .	167
6.10. Проблемы и сомнения . . . . .	168
7. Формирование изображения на основе его численной модели . . . . .	169
7.1. Общее описание метода . . . . .	169
7.2. Формирование изображения на входе антенны . . . . .	171
7.3. Формирование изображения на выходе антенны . . . . .	183
Приложение . . . . .	191
Литература . . . . .	243

## CONTENTS

<i>Introduction</i> . . . . .	3
<i>Part One. Regular fields</i> . . . . .	7
1. Fields in free space . . . . .	7
1.1. Wave fields parameters which are important for getting images . . . . .	8
1.2. Frequency characteristics and the free space response . . . . .	14
1.3. In what cases and why the small-angle approximation is used . . . . .	18
1.4. Simplification used in the small-angle approximation . . . . .	23
1.5. Holography . . . . .	28
1.6. Granularity or speckle noise . . . . .	38
1.7. Remarks on the field finding . . . . .	50
1.8. Examples of mathematical programs illustrating the field formation due to their propagation in free space . . . . .	60
2. Antennas . . . . .	70
2.1. Parameters of the antenna as the system forming images . . . . .	70
2.2. Antenna with angular scanning . . . . .	76
2.3. Focusing antenna . . . . .	82
2.4. Aperture synthesis . . . . .	88
2.5. Radioholography . . . . .	88
2.6. Problems of antenna theory and engineering . . . . .	100
2.7. Mathematical programs on antennas . . . . .	105
3. Images in optic systems . . . . .	110
3.1. Optic system . . . . .	110
3.2. Transformation of fields by an optic system . . . . .	111
3.3. Ideal optic system . . . . .	113
3.4. Images formed by an ideal optic system . . . . .	115
3.5. A real optic system is an ideal optic system + a filter . . . . .	118
3.6. How the optic system transforms the image . . . . .	120
<i>Part Two. Images in random fields</i> . . . . .	129
4. Optic system in the random field . . . . .	129
4.1. What kind of fields are considered random . . . . .	129
4.2. What parameters of the random field affect the image formation . . . . .	130
4.3. Forms of coherent functions . . . . .	132
4.4. Frequency-contrast characteristic of the optic system . . . . .	134
4.5. The role of the lens focal plane in the random field . . . . .	138
4.6. How the frequency-contrast characteristic of the optic system is introduced in noncoherent optics . . . . .	139
5. Aperture synthesis in the random field . . . . .	141

5.1. Coherence function in free space . . . . .	141
5.2. Noncoherent holography . . . . .	142
5.3. Noncoherent aperture analysis . . . . .	143
<i>Part Three. Foundations of the improvement of the image formation methods . . . . .</i>	<i>148</i>
6. Aperture synthesis and binaural hearing . . . . .	148
6.1. Characteristic features of various methods of forming images which allow to recognize them . . . . .	148
6.2. Physical model of binaural hearing . . . . .	150
6.3. Natural and soft-focus stereophonics . . . . .	153
6.4. Physical basis of natural stereophonics . . . . .	155
6.5. Soft-focus stereophonics . . . . .	158
6.6. Physical basis of soft-focus stereophonics . . . . .	160
6.7. Binaural holography . . . . .	162
6.8. 3-D and stereophonic sound . . . . .	165
6.9. What is the peculiar feature of the natural sound . . . . .	167
6.10. Problems and doubts Conclusion . . . . .	168
7. Formation of the image on the basis of its numerical model . . . . .	169
7.1. General description of the method . . . . .	169
7.2. Image formation at the antenna input . . . . .	171
7.3. Image formation at the antenna output . . . . .	183
Appendix . . . . .	191
Reference . . . . .	243



# **PHYSICAL FOUNDATION OF IMAGE FORMATION BY WAVE FIELDS**

**V. A. Zverev**

From the unified viewpoint - the signal transformation by a chain of filters - wave diffraction both in free space and due to wave fields transformation using devices forming images (antennas, optic lenses and correlators) is considered for coherent and noncoherent illumination. Physical foundation of holography, coherent optics, aperture synthesis are illustrated by concrete examples obtained by mathematical simulation of relations introduced in the book which may serve as a practical manual on mathematical simulation of the problems of the formation of wave fields images.

The book is intended for students, graduates and research fellows of corresponding professions.