Е.С. КУХАРКИН

ЭЛЕКТРОФИЗИКА ИНФОРМАЦИОННЫХ СИСТЕМ

Допущено
Советом учебно-методического
объединения по образованию
в области энергетики и электротехники
в качестве учебного пособия
для студентов высших учебных заведений,
обучающихся по направлению
"Электроника и микроэлектроника",
а также по специальностям
"Информационно-измерительная техника и технологии",
"Электрические и электронные аппараты",
"Промышленная электроника и автоматизация
технологических процессов и производств"

Москва
«Высшая школа»
2001

УДК 681.51
ББК 32.81
К 88

Федеральная целевая программа книгоиздания России

Рецензенты:
кафедра теоретических основ оптоэлектроники Московского государственного института радиотехники, электроники и автоматики (зав. кафедрой — чл.-корр. РАН, д-р техн. наук, проф. Л.Д. Бахрах), д-р техн. наук, проф. В.Н. Рудаков

Кухаркин Е.С.
К 88 Электрофизика информационных систем: Учеб. пособие для вузов. — М.: Высш. шк., 2001.— 671 с.: ил.

ISBN 5-06-003986-2

Книга объединяет и обобщает в одном издании основные работы по главным направлениям развития нетрадиционной элементной базы информатики. Особенностью книги является также сочетание теории взаимодействия электромагнитного поля и вещества с изложением практического применения и передачи информации, имея ввиду как современные устройства, так и находящиеся в стадии исследования и разработки.

Для студентов высших учебных заведений. Книга может быть полезна инженерам, научным работникам, аспирантам и студентам, специализирующимся на применении, проектировании и исследовании нетрадиционной элементной базы информатики.

УДК 681.51
ББК 32.81

ISBN 5-06-003986-2

© ГУП «Издательство «Высшая школа», 2001

Оригинал-макет данного издания является собственностью издательства «Высшая школа», и его репродуцирование (воспроизведение) любым способом без согласия издательства запрещается.

Предисловие

Технический арсенал средств современной информатики включает, кроме традиционной элементной базы на основе полупроводников и другие разнообразные приборы и устройства, например, оптроны, светодиоды, фотодиоды, фототранзисторы, транзисторы, трансфазеры, квантроны, криотроны, лазеры, магнитодиоды, магниторезисторы.

Для понимания процессов, происходящих в подобных приборах и элементах, необходимо изучить их природу, зависящую в значительной мере от электрических, магнитных и электромагнитных полей, и в результате иметь представление о возможностях их применения и разработки в вычислительной, информационно-измерительной технике и технике связи. Всем этим вопросам и посвящена настоящая книга. Она завершила многолетнюю работу, начатую на кафедре инженерной электрофизики Московского энергетического института. У ее истоков стояли профессора П. А. Ионкин, А. А. Соколов, Ю. М. Шамаев, М. К. Поливанов и А. И. Пирогов.

В настоящее время опубликовано много книг и статей по отдельным вопросам, освещенным в этом издании. Особенностью настоящей книги является то, что она обобщает и объединяет в одном издании основные работы по новым направлениям развития элементной базы информатики. По содержанию и объему книга вышла за рамки учебника для какой-либо одной специальности высших учебных заведений, приобрела черты монографии по одному из научных направлений в области электрофизических основ информационных средств и технологий.

Другой особенностью книги является сочетание теории взаимодействия электромагнитного поля и вещества с изложением технического применения явлений в твердом теле для управления информацией, имея ввиду как современные устройства, так и находящиеся в стадии исследования и разработки. Значительное место, отводимое теоретическим положениям, отвечает требованиям современной практики, поскольку, как заметил Больцман, «нет ничего более полезного для практики, чем хорошая теория».

Книга написана в результате значительной переработки и дополнения учебника автора «Инженерная электрофизика. Техническая электродинамика» (М.: Высшая школа, 1982). За время, прошедшее после издания, существенно расширился круг физических явлений, открывающих перспективы создания новых более эффективных приборов и устройств для ввода, хранения, переработки и передачи информации. К новым материалам в настоящем издании относятся в частности спекл-интерфе-

рометрия, трехмерные голограммы, формирование Гауссова пучка в оптическом резонаторе, интегральные микроволновые и оптические схемы, теория передачи сигналов в оптических волокнах, элементы машинного анализа и проектирования волноводных и оптических устройств, интегральные оптические корреляторы, электроуправляемые и микроволноводные элементы, магнитостатические и спиновые волны, солитоны, акустоэлектронные устройства на основе концентрационной и решеточной нелинейности, электродинамика джозефсоновского контакта, интегральные микросхемы на сверхпроводниках, синхронизация мод, самофокусировка, генерация ультракоротких импульсов, обращение волнового фронта, бистабильность в инжекционных лазерах, интегральные схемы квантовой электроники в оптических преобразователях и системах связи.

Автор с благодарностью принял и учел критические замечания и советы д-ра техн. наук, проф. В. Н. Рудакова. Автор признателен также за помощь в работе над книгой зав. кафедрой электрофизики МЭИ проф., акад. Международной академии информатизации Ю. А. Казанцеву и д-ру техн. наук, проф. П. Н. Матханову.

Автор

Введение

Наше время — время научно-технической революции и информационного взрыва — поставило перед человечеством ряд фундаментальных проблем. Одна из них связана с поиском и созданием новых технических средств обработки, передачи и хранения информации. Современная электроника достигла в этом направлении выдающихся достижений. Дальнейшее ее развитие идет по пути широкого использования сверхбольших интегральных схем и непрерывного продвижения в область все более коротких длин волн, коротких и ультракоротких длительностей сигналов при одновременном повышении пропускной способности информационных систем, их производительности.

Развитие информационно-измерительной и вычислительной техники привело все к расширяющемуся применению техники сверхвысоких частот (СВЧ). Совершенствование электроники СВЧ характеризуется непрерывным усложнением устройств, повышением степени интеграции, расширением видов линий передач и их функциональных узлов, предназначенных для формирования, обработки и передачи сигналов. Так при переходе к диапазону сантиметровых волн двухпроводная и коаксиальная линии были заменены полыми волноводами. Необходимость миниатюризации и микроминиатюризации привела к освоению сложных и микрополосковых линий передачи различных типов, техники гибридных и объемных интегральных схем, а также функциональных элементов на их основе.

Проблема создания многофункциональных электронных узлов не может быть решена только путем увеличения количества элементов и повышением их интеграции. Поэтому возникла необходимость в поиске решений, основанных на новых физических явлениях, позволяющих значительно увеличить функциональные возможности единицы объема устройства, что является одним из основных направлений современной электроники. Исследования, имеющие целью разработку функциональных интегральных узлов, работа которых основывается на применении сложных электронных эффектов в одном элементе, приобрели особую актуальность. Один из возможных путей реализации функциональных элементов — использование различных физических явлений, наблюдаемых в нелинейных средах. Такие элементы являются основой для нового интенсивно развивающегося направления микроэлектроники — создания функциональных устройств.

Примерами перспективных многофункциональных приборов могут служить диоды с междолинным электронным переносом, акустоэлек-

тронные приборы, принцип работы которых основан на нелинейных взаимодействиях акустических и электрических волн. Благодаря своей многофункциональности, адаптивности к виду обрабатываемых сигналов, миниатюрности, низкому энергопотреблению и высокой надежности они получают все большее распространение. Новые пути в развитии функциональных интегральных схем открывает использование устройств на магнитостатических и спиновых волнах.

Новым этапом развития микроэлектроники стало применение сверхпроводимости. В пленочных сверхпроводниковых структурах и контактах один за другим были обнаружены неизвестные ранее эффекты. Эффекты Джозефсона, нелинейность вольт-амперной характеристики (ВАХ) контактов или магнитные вихри позволяют получить интегральные микросхемы с уникально низким энергопотреблением. Криогенная электроника — относительно молодое и быстроразвивающееся направление, использующее макроскопические квантовые эффекты в измерительных приборах и элементах вычислительной техники с высокой чувствительностью и быстродействием и малым энергопотреблением. Прорыв в высокотемпературную сверхпроводимость создал условия для использования этих явлений в ближайшие годы.

Освоение все более коротковолновых диапазонов, интенсивное развитие техники СВЧ стали предпосылками для создания квантовых генераторов сначала сантиметрового, а потом и оптического диапазона длин волн. Появление и быстрое развитие лазеров дало возможность широко применить в оптике большое количество хорошо разработанных методов и средств электроники СВЧ — техники когерентных электромагнитных колебаний. В канализирующих системах диэлектрические волноводы и излучатели оказали решающее воздействие на развитие оптической связи на основе световодов. Для создания элементной базы в оптике вполне пригодны методы микроэлектронной схемотехники.

Среди квантовых генераторов следует выделить полупроводниковые лазеры, которые составляют фундамент современной оптоэлектроники. После быстрого развития волоконных линий полупроводниковые лазеры и волоконно-оптические линии связи разрабатываются в тесной взаимосвязи.

Использование когерентно-оптических методов открывает новые перспективы преобразования, хранения и обработки информации. Их основные достоинства — хранение информации с огромной плотностью (до 10^6 бит/мм^2), одновременное преобразование двумерных массивов информации в виде картин, изображений, таблиц. Производительность оптических методов может быть на несколько порядков выше, чем электронных. В динамическом режиме работы когерентно-оптических систем носителем информации является модулируемый во времени и пространстве лазерный луч. Одними из ключевых элементов таких систем явля-

ются пространственно-временные модуляторы света. Формирование и преобразование двумерных массивов информации такими модуляторами позволяет реализовывать оптические запоминающие и периферийные устройства, когерентные оптические процессоры и другие важнейшие узлы информационных и вычислительных систем. Их физической основой служат электро-, акусто- и магнитооптические эффекты в различных материалах, в том числе в жидких кристаллах, полупроводниковых материалах, магнитных пленках и керамике.

Предполагается, что оптоэлектронные средства передачи, обработки и хранения информации будут способны обеспечить скорости передачи информации по каналам связи $10^8...10^{11}$ бит/с, скорости обработки информации для процессоров — $10^{10}...10^{11}$ бит/с, объем памяти запоминающих устройств — $10^{12}...10^{14}$ бит. Если сопоставить этот объем с объемом памяти человеческого мозга (порядка 10^{13} бит), то нетрудно видеть, что оптоэлектроника вплотную приближается к «биологическому» уровню развития информационно-измерительной техники. Решение конструктивно-технологических проблем создания элементной базы оптоэлектроники позволит реализовать чисто оптическую (фотонную) вычислительную машину.

Материал настоящей книги имеет целью познакомить читателя с принципами функционирования устройств информационно-измерительной техники в изложенных направлениях. Более глубокому их изучению поможет литература, приведенная в конце книги.

Раздел первый

ЭЛЕКТРОМАГНИТНЫЕ ЯВЛЕНИЯ В ИЗОТРОПНЫХ ЛИНЕЙНЫХ СРЕДАХ

Глава 1
Отражение, преломление и интерференция электромагнитных волн

Электромагнитные явления на границе двух сред, а также при взаимодействии различных волн наиболее отчетливо проявляются в монохроматических плоских волнах настолько большой протяженности, что их можно считать обладающими волновыми фронтами бесконечной протяженности и их пространственной дисперсией можно пренебречь. В связи с этим полезно вспомнить некоторые свойства плоских волн.

§ 1.1. Однородные и неоднородные плоские волны

В плоской волне зависимость напряженностей электромагнитного поля от координат определяется экспонентой

$$e^{j\underline{\mathbf{k}}\mathbf{R}} = e^{j(\pm\underline{k}_x x \pm \underline{k}_y y \pm \underline{k}_z z)}, \tag{1.1}$$

где $\mathbf{R} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z$ — радиус-вектор рассматриваемой точки;

$$\underline{\mathbf{k}} = \pm\underline{k}_x\mathbf{e}_x \pm \underline{k}_y\mathbf{e}_y \pm \underline{k}_z\mathbf{e}_z = \underline{\mathbf{k}}' - j\underline{\mathbf{k}}'' \tag{1.2}$$

— волновой вектор.

Например, комплексная амплитуда вектора напряженности электрического поля

$$\dot{\mathbf{E}} = \dot{\mathbf{E}}_0 e^{\pm j\underline{\mathbf{k}}\mathbf{R}}, \tag{1.3}$$

где

$$\dot{\mathbf{E}}_0 = \mathbf{E}' + j\mathbf{E}'' \tag{1.4}$$

— начальная комплексная векторная амплитуда (при $R = 0$), не зависящая от координат и времени. Поэтому для любой комплексной компоненты плоской волны оператор набла

$$\nabla = \mathbf{e}_x \frac{\partial}{\partial x} + \mathbf{e}_y \frac{\partial}{\partial y} + \mathbf{e}_z \frac{\partial}{\partial z} = \pm j(\underline{k}_x \mathbf{e}_x + \underline{k}_y \mathbf{e}_y + \underline{k}_z \mathbf{e}_z) = \pm j\underline{\mathbf{k}}.$$

Разные знаки в приведенных выражениях соответствуют двум волнам, отличающимся только направлениями распространения. Дальнейшее изложение, если не будет дополнительных оговорок, посвящено *прямым* волнам, распространяющимся вдоль положительных направлений координатных осей (например, x, y или z) и соответствующим отрицательному знаку.

При этом первое уравнение Максвелла имеет вид

$$\operatorname{rot}\dot{\mathbf{H}} \equiv \nabla \times \dot{\mathbf{H}} \equiv -j\underline{k} \times \dot{\mathbf{H}} = j\omega\dot{\mathbf{D}} \equiv j\omega\underline{\varepsilon}_a\dot{\mathbf{E}},$$

второе уравнение Максвелла

$$\operatorname{rot}\dot{\mathbf{E}} \equiv \nabla \times \dot{\mathbf{E}} \equiv -j\underline{k} \times \dot{\mathbf{E}} = -j\omega\dot{\mathbf{B}} \equiv j\omega\underline{\mu}_a\dot{\mathbf{H}},$$

где $\underline{\varepsilon}_a$, $\underline{\mu}_a$ — абсолютные диэлектрическая и магнитная проницаемости*.

В общем случае — для анизотропных сред с потерями (проводимостями) — эти проницаемости являются комплексными тензорами. Таким образом, уравнения Максвелла для плоских волн приобретают вид

$$\omega\dot{\mathbf{D}} = \dot{\mathbf{H}} \times \mathbf{k}, \quad \omega\dot{\mathbf{B}} = \mathbf{k} \times \dot{\mathbf{E}} \tag{1.5а}$$

или

$$\omega(\underline{\varepsilon})\dot{\mathbf{E}} = \dot{\mathbf{H}} \times \mathbf{k}, \quad \omega(\underline{\mu})\dot{\mathbf{H}} = \mathbf{k} \times \dot{\mathbf{E}}. \tag{1.5б}$$

Волны в анизотропных средах будут рассмотрены во втором разделе.

Дальнейшее изложение в данном разделе относится только к изотропным средам, для которых совместное решение уравнений (1.5а) и (1.5б) приводит** к *дисперсионному уравнению* (уравнению нормалей)

$$\underline{k}^2 = \frac{\omega^2}{c_0^2}\underline{\varepsilon}_r\underline{\mu}_r = \frac{\omega^2}{c_0^2}(\varepsilon' - j\varepsilon'')(\mu' - j\mu'') \tag{1.6}$$

или

$$\underline{k} = \frac{\omega}{c_0}\sqrt{\underline{\varepsilon}_r\underline{\mu}_r} = \frac{\omega}{c_0}\underline{n}, \tag{1.7}$$

где

$$\underline{n} = \sqrt{\underline{\varepsilon}_r\underline{\mu}_r} = n' - jn'' \tag{1.8}$$

— показатель преломления.

---

* В дальнейшем индекс a опускается.

** С учетом того, что $\mathbf{kH} = 0$, $\mathbf{kE} = 0$.

Во многих расчетах удобно вместо волнового вектора (1.2) пользоваться *вектором рефракции*

$$\underline{\mathbf{n}} = \underline{n}\,\mathbf{e}_n = \frac{c_0}{\omega}\underline{\mathbf{k}}, \tag{1.9}$$

длина которого равна показателю преломления (1.8), а направление совпадает с направлением волнового вектора $\mathbf{k}$ и *волновой нормали* $\mathbf{e}_n$. После замены волнового вектора вектором рефракции уравнения (1.5а) и (1.5б) примут вид

$$(\underline{\varepsilon}_r)\dot{\mathbf{E}} = c\mu_0\dot{\mathbf{H}} \times \underline{\mathbf{n}},$$
$$(\underline{\mu}_r)\dot{\mathbf{H}} = c\varepsilon_0\underline{\mathbf{n}} \times \dot{\mathbf{E}}. \tag{1.10}$$

Волны с вещественным волновым вектором ($\underline{\mathbf{k}}'' = 0$) называют *однородными*. Вещественным значением волнового вектора обладают волны в однородных прозрачных средах, где отсутствует поглощение (диссипация) и проницаемости имеют вещественные значения ($\varepsilon'' = 0$, $\mu'' = 0$). В этом случае из уравнений Максвелла

$$\omega\varepsilon\dot{\mathbf{E}} = \dot{\mathbf{H}} \times \mathbf{k}, \quad \omega\mu\dot{\mathbf{H}} = \mathbf{k} \times \dot{\mathbf{E}} \tag{1.11}$$

следует, что взаимно ортогональные векторы $\mathbf{E}$ и $\mathbf{H}$ лежат в одной плоскости, перпендикулярной волновому вектору $\mathbf{k} = k\mathbf{e}_n$ и являющейся плоскостью равных фаз:

$$\mathbf{kR} = \text{const} \;\text{ и }\; \mathbf{e}_n\mathbf{R} = \text{const}. \tag{1.12}$$

Кроме того,

$$\dot{\mathbf{H}} = \frac{\mathbf{e}_n \times \dot{\mathbf{E}}}{\sqrt{\mu/\varepsilon}} = \frac{\mathbf{e}_n \times \dot{\mathbf{E}}}{Z_c} = \dot{\mathbf{H}}_0\, e^{-j\mathbf{kR}}, \tag{1.13}$$

где $Z_c = \sqrt{\mu/\varepsilon}$ — волновое сопротивление, которое в прозрачных средах является вещественным, т. е. $\mathbf{E}$ и $\mathbf{H}$ совпадают по фазе и их амплитуды не меняются на фронте волны. Иначе говоря, в однородной волне плоскость равных амплитуд совпадает с плоскостью равных фаз.

Неоднородными называют волны, у которых плоскости равных фаз и равных амплитуд непараллельны. Поскольку, строго говоря, любая среда является в некоторой мере поглощающей, то неоднородные волны представляют собой распространенное явление. Одним из их признаков является комплексный характер волнового вектора (1.2). В этом случае согласно (1.3),

$$\dot{\mathbf{E}} = \dot{\mathbf{E}}_0\, e^{-\mathbf{k}''\mathbf{R}} e^{-j\mathbf{k}'\mathbf{R}}. \tag{1.14}$$

Если представить компоненты волнового вектора в виде

$$\mathbf{k}' = k'\mathbf{e}_n, \qquad \mathbf{k}'' = k''\mathbf{e}_s, \tag{1.15}$$

то уравнения

$$\mathbf{e}_n\mathbf{R} = \text{const}, \qquad \mathbf{e}_s\mathbf{R} = \text{const} \tag{1.16}$$

будут определять плоскости равной фазы и равной амплитуды (*фазовая* и *амплитудная плоскости*). Единичный вектор $\mathbf{e}_s$ называют *амплитудной нормалью*. Поскольку в неоднородной волне поверхности равных фаз и амплитуд не совпадают, составляя угол $\psi = \hat{\mathbf{e}_n\mathbf{e}_s}$, амплитуда на поверхности равных фаз (по фронту волны) меняется.

Связь между компонентами волнового вектора и составляющими проницаемостей следует из дисперсионного уравнения (1.6):

$$\mathbf{n}'^2 - \mathbf{n}''^2 = \frac{c_0^2}{\omega^2}(\mathbf{k}'^2 - \mathbf{k}''^2) = \varepsilon_r'\mu_r' - \varepsilon_r''\mu_r''; \tag{1.17}$$

$$2\mathbf{n}'\mathbf{n}'' = 2\frac{c_0^2}{\omega^2}\mathbf{k}'\mathbf{k}'' = \varepsilon_r'\mu_r' + \varepsilon_r''\mu_r''. \tag{1.18}$$

Отметим, что неоднородные волны могут распространяться и в прозрачных средах (без потерь). Это следует из того, что в случае вещественной правой части дисперсионное уравнение (1.6)

$$\mathbf{k}^2 = \mathbf{k}'^2 - \mathbf{k}''^2 - j2\mathbf{k}'\mathbf{k}'' = \frac{\omega^2}{c_0^2}\varepsilon_r\mu_r \tag{1.19}$$

соответствует такому комплексному волновому вектору, у которого

$$\mathbf{k}'^2 - \mathbf{k}''^2 = \frac{\omega^2}{c_0^2}\varepsilon_r\mu_r; \tag{1.20}$$

$$\mathbf{k}'\mathbf{k}'' = 0. \tag{1.21}$$

Следовательно, неоднородные волны в изотропных прозрачных диэлектриках возникают в тех случаях, когда вещественная и мнимая части волнового вектора взаимно ортогональны, т. е. в изотропных диэлектриках без потерь затухающая волна всегда неоднородна, причем фазовая и амплитудная плоскости взаимно перпендикулярны.

§ 1.2. Поляризация электромагнитной волны

Одним из существенных параметров электромагнитной волны является направление колебаний, называемое *поляризацией волны*. Будем называть *направлением поляризации* направление вектора напряженности электрического поля (1.3), а *плоскостью поляризации* — плоскость, в ко-

торой лежат вектор напряженности электрического поля и направление распространения.

Характер поляризации плоских волн определяется видом замкнутой кривой, которую описывает за период электромагнитных колебаний конец вектора напряженности электрического поля

$$\mathbf{E} = \mathbf{E}(t) = \mathrm{Re}(\dot{\mathbf{E}} e^{j\omega t}) = \mathrm{Re}(\dot{\mathbf{E}}_0 e^{j\psi}) = \mathbf{E}' \cos\psi - \mathbf{E}'' \sin\psi, \qquad (1.22)$$

где $\psi = \omega t - \mathbf{kR}$.

Уравнение этой кривой в инвариантной (относительно координат) форме можно получить, если исключить из уравнения (1.22) фазу ψ*:

$$[\mathbf{E}\mathbf{E}']^2 + [\mathbf{E}\mathbf{E}'']^2 = [\mathbf{E}'\mathbf{E}'']^2. \qquad (1.23)$$

Можно убедиться, что это уравнение эллипса. Таким образом, плоская волна в общем случае обладает *эллиптической поляризацией*.

В дальнейшем будет встречаться простейший случай, когда $\mathbf{E}'$ и $\mathbf{E}''$ взаимно перпендикулярны ($\mathbf{E}'\mathbf{E}'' = 0$), но не равны по длине ($\mathbf{E}' \neq \mathbf{E}''$). При этом вектор (1.22) описывает канонический эллипс, причем $\mathbf{E}'$ и $\mathbf{E}''$ по значению и направлению дают полуоси эллипса. Действительно, если выбрать направление $\mathbf{E}'$ за ось x, а $\mathbf{E}''$ — за ось y, то получим

$$[\mathbf{E}\mathbf{E}']^2 = \mathbf{E}'^2 E^2 \sin^2(\hat{\mathbf{E}\mathbf{E}'}) = \mathbf{E}'^2 E^2 \cos^2(\hat{\mathbf{E}\mathbf{e}_y}) = \mathbf{E}'^2 y^2,$$

$$[\mathbf{E}\mathbf{E}'']^2 = \mathbf{E}''^2 x^2, \quad [\mathbf{E}'\mathbf{E}'']^2 = \mathbf{E}'^2 \mathbf{E}''^2.$$

Тогда выражение (1.23) примет вид

$$\frac{x^2}{E'^2} + \frac{y^2}{E''^2} = 1. \qquad (1.24)$$

Другие виды поляризации можно рассматривать как частные случаи эллиптической поляризации. Например, при равенстве ортогональных полуосей

$$\mathbf{E}'\mathbf{E}'' = 0, \qquad \mathbf{E}'^2 - \mathbf{E}''^2 = 0 \qquad (1.25)$$

и эллипс (1.24) превращается в окружность радиусом $\mathbf{E}' = \mathbf{E}''$, что характеризует *круговую поляризацию*.

Круговая поляризация делится на правую и левую в зависимости от направления вращения вектора $\mathbf{E}$. Будем называть поляризацию *правой*, когда конец вектора напряженности электрического поля вращается по

---

* Для этого достаточно векторно умножить уравнение (1.22) сначала на $\mathbf{E}'$, затем на $\mathbf{E}''$, далее полученные результаты возвести в квадрат и сложить. Здесь квадратные скобки обозначают векторное произведение. В дальнейшем это обозначение используется наряду с общепринятым (×).

часовой стрелке, если смотреть со стороны положительного направления движения волны. При *левой* поляризации наблюдатель, смотрящий навстречу лучу, будет видеть вектор напряженности вращающимся против часовой стрелки. Направление вращения определяется вектором

$$\frac{\mathrm{d}\mathbf{E}}{\mathrm{d}\psi} = -\mathbf{E}'\sin\psi - \mathbf{E}''\cos\psi. \tag{1.26}$$

Поэтому критерием правой поляризации будет

$$-\left[\mathbf{E}\frac{\mathrm{d}\mathbf{E}}{\mathrm{d}\psi}\right]\mathbf{e}_n = [\mathbf{E}'\mathbf{E}'']\mathbf{e}_n > 0, \tag{1.27}$$

а левой поляризации —

$$[\mathbf{E}'\mathbf{E}'']\mathbf{e}_n < 0. \tag{1.28}$$

Если полуоси эллипса (1.24) подчиним условиям (1.25) и (1.27), то получим соотношение между компонентами комплексной амплитуды ($\mathbf{E}_0 = \mathbf{E}_x + \mathbf{E}_y$) правополяризованной волны

$$\dot{E}_y = j\dot{E}_x. \tag{1.29}$$

Исходя из (1.25) и (1.28), для левополяризованной волны найдем

$$\dot{E}_y = -j\dot{E}_x. \tag{1.30}$$

Другой частный случай эллиптической поляризации соответствует вырождению эллипса в прямую линию, когда гармонически изменяющийся вектор напряженности (1.22) все время сохраняет свое направление. В такой *линейно поляризованной* волне изменение вектора (1.26) должно быть параллельно самому вектору напряженности (1.22), условие линейной поляризации приобретает вид

$$\left[\mathbf{E}\frac{\mathrm{d}\mathbf{E}}{\mathrm{d}\psi}\right] = -[\mathbf{E}'\mathbf{E}''] = 0. \tag{1.31}$$

Следовательно, вектор в любой момент времени направлен вдоль вектора своей комплексной амплитуды $\mathbf{E}_0 \parallel \mathbf{E}' \parallel \mathbf{E}''$, которая и определяет направление плоскости поляризации. Только в этом случае комплексная амплитуда волны (1.4) может быть представлена в виде

$$\dot{\mathbf{E}}_0 = \mathbf{E}_0 e^{j\varphi}, \tag{1.32}$$

соответствующем *постоянству пространственного направления* вектора (1.22). Соотношение между компонентами E' и E'' влияет здесь только на начальную фазу $\varphi = \mathrm{arctg}(E''/E')$.

§ 1.3. Сложение волн с различной поляризацией

При рассмотрении электромагнитных явлений приходится сталкиваться с необходимостью сложения различных волн и получения новой результирующей волны, а также разложения заданной волны на две другие. Используемое при этом геометрическое суммирование гармонических волн в линейных средах допустимо вследствие линейности уравнений Максвелла, поскольку суммарное решение позволяет получить волну, удовлетворяющую этим же уравнениям.

Рассмотрим несколько типичных случаев сложения гармонических волн, обладающих одинаковой частотой, скоростью и направлением распространения. Результирующая волна имеет такие же значения перечисленных величин. В уравнениях этих волн фазовые множители $e^{j\psi}$ одинаковы, и волны различаются за счет комплексных векторных амплитуд $\dot{\mathbf{E}}_0$, т. е. за счет поляризации и интенсивности. Сложение комплексных векторных амплитуд происходит в одной плоскости.

При сложении двух линейно поляризованных волн $\dot{\mathbf{E}}_1 = \mathbf{E}_1 e^{j\psi_1}$ и $\dot{\mathbf{E}}_2 = \mathbf{E}_2 e^{j\psi_2}$ *линейная поляризация* результирующей волны имеет место, что следует из (1.31) или (1.32), в двух случаях:
когда комплексные амплитуды параллельны:

$$\mathbf{E}_1 \parallel \mathbf{E}_2, \tag{1.33}$$

либо когда отношения длин вещественной и мнимой частей одинаковы:

$$\psi_1 = \psi_2. \tag{1.34}$$

Следовательно, линейно поляризованная волна возникает, если у исходных линейно поляризованных волн совпадают плоскости поляризации или фазы.

Из рассмотренных свойств линейно поляризованной волны также следует, что компоненты $\mathbf{E}'$ и $\mathbf{E}''$ комплексной векторной амплитуды $\dot{\mathbf{E}}_0$ определяют две линейно поляризованные волны, сдвинутые по фазе на 90°. Поэтому равенства (1.25) являются условиями того, что при сложении таких двух линейно поляризованных волн образуется волна с *круговой поляризацией*. Согласно этим условиям, волна с круговой поляризацией возникает из двух линейно поляризованных волн, если они имеют одинаковые амплитуды, сдвинутые в пространстве и по фазе на 90°. Для правополяризованной волны (1.29) амплитуды

$$\dot{\mathbf{E}}_+ = \dot{\mathbf{E}}_x + \dot{\mathbf{E}}_y = A(\mathbf{e}_x + j\mathbf{e}_y), \tag{1.35}$$

а для левополяризованной волны (1.30)

$$\dot{\mathbf{E}}_- = A(\mathbf{e}_x - j\mathbf{e}_y). \tag{1.36}$$

Из выражений (1.35) и (1.36) видно, что при сложении волн с круговой поляризацией с левым и правым вращением с одинаковой амплитудой A возникает линейно поляризованная волна с удвоенной амплитудой

$$\dot{\mathbf{E}}_0 = \dot{\mathbf{E}}_+ + \dot{\mathbf{E}}_- = 2A\mathbf{e}_x = 2\mathbf{E}'. \tag{1.37}$$

Отсюда следует и обратный вывод: любую линейно поляризованную волну можно разложить на две волны с круговой поляризацией с противоположными направлениями вращения и одинаковыми амплитудами.

Из условия (1.34) следует, что любую линейно поляризованную волну можно также разложить на две другие линейно поляризованные волны с ортогональными амплитудами при условии, что их фазы будут совпадать. Например, $\dot{\mathbf{E}} = E_x\mathbf{e}_x + E_y\mathbf{e}_y$.

Если рассмотренные условия суммирования волн с линейной и круговой поляризацией не выполняются, то результирующая волна обычно имеет эллиптическую поляризацию.

§ 1.4. Законы отражения и преломления

Параметры неоднородных сред зависят от пространственных координат. Эта зависимость может иметь различный характер. При наличии резкой границы между двумя однородными средами проницаемости испытывают скачкообразное изменение. Такое изменение представляет собой широко распространенный и практически важный случай, рассмотрению которого и посвящен настоящий параграф. Законы отражения и преломления в феноменологической макроскопической линейной электродинамике обычно получают, рассматривая падение неограниченной плоской монохроматической волны на гладкую неподвижную границу раздела двух полубесконечных сред. Предполагается, что обе среды идеально однородны, а граница раздела — геометрическая плоскость, определяемая уравнением

$$\mathbf{r}\mathbf{e}_q = 0, \tag{1.38}$$

где $\mathbf{r}$ — радиус-вектор на поверхности раздела; $\mathbf{e}_q$ — единичный вектор нормали к поверхности раздела, направленный из первой среды (с падающей волной) во вторую.

На этой поверхности должны быть выполнены граничные условия

$$\mathbf{E}_1 \times \mathbf{e}_q = \mathbf{E}_2 \times \mathbf{e}_q, \tag{1.39}$$

$$\mathbf{H}_1 \times \mathbf{e}_q = \mathbf{H}_2 \times \mathbf{e}_q. \tag{1.40}$$

При падении на границу волны $\dot{\mathbf{E}}^{+} = \dot{\mathbf{E}}_1^{+} e^{j\psi_1^{+}}$ граничные условия могут быть удовлетворены в изотропной среде только при наличии двух новых волн: отраженной ($\dot{\mathbf{E}}^{-} = \dot{\mathbf{E}}_1^{-} e^{j\psi_1^{-}}$) и преломленной ($\dot{\mathbf{E}}' = \dot{\mathbf{E}}_2^{-} e^{j\psi_2^{-}}$), где

$$\begin{aligned} \psi_1^{+} &= \omega_1^{+} t - \mathbf{k}_1^{+}\mathbf{r}; \\ \psi_1^{-} &= \omega_1^{-} t - \mathbf{k}_1^{-}\mathbf{r}; \\ \psi_2^{-} &= \omega_2 t - \mathbf{k}_2\mathbf{r}. \end{aligned} \tag{1.41}$$

При этом из (1.39) следует, что

$$\dot{\mathbf{E}}_1^{+} \times \mathbf{e}_q e^{j\varphi_1^{+}} + \dot{\mathbf{E}}_1^{-} \times \mathbf{e}_q e^{j\varphi_1^{-}} - \dot{\mathbf{E}}_2^{-} \times \mathbf{e}_q e^{j\varphi_2^{-}} = 0. \tag{1.42}$$

Векторные произведения постоянны, поэтому уравнение (1.42) дает линейное соотношение между экспоненциальными функциями с различными независимыми показателями, причем это уравнение должно удовлетворяться в любой точке плоскости раздела и в любой момент времени. Это возможно в том случае, когда либо все фазы (1.41) совпадают, либо все векторные коэффициенты равны нулю. Последняя возможность отпадает, поскольку для падающей волны $\dot{\mathbf{E}}_1^{+}$ отлична от нуля. Поэтому

$$\psi_1^{+} = \psi_1^{-} = \psi_2^{-}. \tag{1.43}$$

Отсюда следует, во-первых, равенство частот:

$$\omega_1^{+} = \omega_1^{-} = \omega_2 = \omega$$

и, во-вторых, равенство произведений:

$$\mathbf{k}_1^{+}\mathbf{r} = \mathbf{k}_1^{-}\mathbf{r} = \mathbf{k}_2\mathbf{r}; \tag{1.44}$$

$$(\mathbf{k}_1^{-} - \mathbf{k}_1^{+})\mathbf{r} = (\mathbf{k}_2 - \mathbf{k}_1^{+})\mathbf{r} = 0;$$

$$\mathbf{k}_1^{+} \times \mathbf{e}_q = \mathbf{k}_1^{-} \times \mathbf{e}_q = \mathbf{k}_2 \times \mathbf{e}_q = \mathbf{a} = a\mathbf{e}_a. \tag{1.45}$$

Уравнение (1.45) выражает общий закон отражения и преломления электромагнитных волн на границе любых двух сред в инвариантной форме: векторное произведение волнового вектора на вектор нормали к поверхности раздела дает вектор, постоянный для всех волн — падающих, отраженных и преломленных. Он направлен вдоль нормали $\mathbf{e}_a$ к плоскости падения.

Первым следствием из общего закона является равенство проекций a всех волновых векторов падающих, отраженных и преломленных волн на плоскость раздела и *копланарность волновых векторов* этих волн.

Из сравнения абсолютных значений векторных произведений (1.45) следует

$$k_1^+ \sin\theta_1^+ = k_1^- \sin\theta_1^- = k_2 \sin\theta_2, \qquad (1.46)$$

где θ_1^+, θ_1^- и θ_2 — углы падения, отражения и преломления.

Поскольку $k_1^+ = k_1^- = k_1$, то вторым следствием из общего закона, согласно (1.46), является *равенство углов падения и отражения*: $\theta_1^+ = \theta_1^- = \theta_1$.

Соотношение между углами падения и преломления (*закон Снеллиуса*) можно рассматривать как третье следствие из общего закона (1.45). Согласно (1.46),

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{k_2}{k_1} = \frac{n_2}{n_1} = N. \qquad (1.47)$$

При выводе равенства (1.45) не предполагалась вещественность волновых векторов. Поэтому сформулированный закон применим как для прозрачных, так и для поглощающих сред. Однако для неоднородных волн следствия из общего закона должны относиться отдельно к вещественным и мнимым компонентам волновых векторов, определяя их пространственное расположение отдельно для фазовой и амплитудной нормалей. Иногда удобно в поглощающей среде формально использовать следствие (1.47) и для комплексных показателей преломления, применяя «комплексный угол преломления» θ_2.

§ 1.5. Коэффициенты отражения и преломления

Амплитуды отраженной и преломленной волн зависят не только от свойств среды по обе стороны от границы раздела и угла падения, но и от поляризации падающей волны. Поскольку волну с произвольной поляризацией можно разложить на две ортогонально поляризованные волны, то целесообразно рассмотреть отдельно волну, поляризованную параллельно плоскости падения, и волну, поляризованную перпендикулярно плоскости падения.

Нормально поляризованные волны имеют векторы напряженности электрического поля, направленные перпендикулярно плоскости падения и вектору нормали $\mathbf{e}_q$. В этом случае согласно (1.39) и (1.40),

$$\dot{\mathbf{E}}_1^+ + \dot{\mathbf{E}}_1^- = \dot{\mathbf{E}}_2^-;$$

$$(\dot{\mathbf{H}}_1^+ - \dot{\mathbf{H}}_1^-)\cos\theta_1 = \dot{\mathbf{H}}_2^- \cos\theta_2,$$

или с учетом (1.13)

$$\frac{\dot{\mathbf{E}}_1^+ - \dot{\mathbf{E}}_1^-}{Z_1}\cos\theta_1 = \frac{\dot{\mathbf{E}}_2^-}{Z_2}\cos\theta_2,$$

где Z_1, Z_2 — волновые сопротивления первой и второй сред.

Тогда амплитуды отраженной и преломленной волн

$$\dot{\mathbf{E}}_1^- = \frac{Z_2\cos\theta_1 - Z_1\cos\theta_2}{Z_2\cos\theta_1 + Z_1\cos\theta_2}\dot{\mathbf{E}}_1^+,$$

$$\dot{\mathbf{E}}_2^- = \frac{2Z_2\cos\theta_1}{Z_2\cos\theta_1 + Z_1\cos\theta_2}\dot{\mathbf{E}}_1^+$$

оказываются прямо пропорциональными амплитуде падающей волны. Поэтому отражающие и преломляющие свойства границы удобнее характеризовать относительными величинами, называемыми *коэффициентами отражения и преломления*:

$$\rho_{11}^{\perp} = \frac{\dot{\mathbf{E}}_1^-}{\dot{\mathbf{E}}_1^+} = \frac{Z_2\cos\theta_1 - Z_1\cos\theta_2}{Z_2\cos\theta_1 + Z_1\cos\theta_2}; \qquad (1.48)$$

$$\rho_{21}^{\perp} = \frac{\dot{\mathbf{E}}_2^-}{\dot{\mathbf{E}}_1^+} = \frac{2Z_2\cos\theta_1}{Z_2\cos\theta_1 + Z_1\cos\theta_2}. \qquad (1.49)$$

Параллельно поляризованные волны имеют векторы напряженности магнитного поля, параллельные границе раздела и перпендикулярные нормали $\mathbf{e}_q$. Поэтому по (1.39) и (1.40)

$$(\dot{\mathbf{E}}_1^+ + \dot{\mathbf{E}}_1^-)\cos\theta_1 = \dot{\mathbf{E}}_2^-\cos\theta_2; \quad \dot{\mathbf{H}}_1^+ - \mathbf{H}_1^- = \dot{\mathbf{H}}_2^-;$$

$$\frac{\dot{\mathbf{E}}_1^+ - \dot{\mathbf{E}}_1^-}{Z_1} = \frac{\dot{\mathbf{E}}_2^-}{Z_2}.$$

При этом

$$\rho_{11}^{\parallel} = \frac{Z_2\cos\theta_2 - Z_1\cos\theta_1}{Z_2\cos\theta_2 + Z_1\cos\theta_1}; \qquad (1.50)$$

$$\rho_{21}^{\parallel} = \frac{2Z_2\cos\theta_1}{Z_2\cos\theta_2 + Z_1\cos\theta_1}. \qquad (1.51)$$

Уравнения, связывающие амплитуды отраженных и преломленных волн с амплитудой и углом падения падающей волны, а также с параметрами граничных сред, как и выражения для коэффициентов (1.48)—(1.51), называют *формулами Френеля*. С помощью этих формул далее рассмотрено несколько частных случаев, встречающихся на практике (см. §§ 1.7, 1.8). Предварительно полезно обратить внимание на то, что вся совокупность волновых процессов на границах раздела и в различных технических устройствах может быть представлена в компактной и наглядной форме с помощью волновой матрицы и ориентированных графов.

§ 1.6. Волновая матрица

Преимуществом топологического способа анализа и расчета электродинамических устройств является, как известно, не только наглядность графического изображения, но и быстрота получения конечного результата. Анализ сложного устройства *методом графов* позволяет отвлечься от решения граничной электродинамической задачи, избежать громоздких математических преобразований, не требует составления системы алгебраических уравнений.

На границе раздела двух сред подобное положение складывается при прохождении через нее прямой $\dot{\mathbf{E}}_1^+$ и обратной $\dot{\mathbf{E}}_2^+$ волн. Для последней коэффициенты преломления ρ_{12} и отражения ρ_{22} могут быть получены из уравнений Френеля (1.48)—(1.51) после перестановки индексов. Коэффициенты преломления и отражения для прямой и обратной волн определяют передачи путей ориентированного графа, соответствующего прохождению волн через границу раздела (рис. 1.1).

Граф позволяет выразить результирующие амплитуды выходящих волн сразу через амплитуды входящих волн

$$\dot{\mathbf{E}}_1^- = \rho_{11}\dot{\mathbf{E}}_1^+ + \rho_{12}\dot{\mathbf{E}}_2^+;$$

$$\dot{\mathbf{E}}_2^- = \rho_{21}\dot{\mathbf{E}}_1^+ + \rho_{22}\dot{\mathbf{E}}_2^+$$

или в матричной форме

$$\begin{bmatrix}\dot{\mathbf{E}}_1^-\\ \dot{\mathbf{E}}_2^-\end{bmatrix} = \begin{bmatrix}\rho_{11} & \rho_{12}\\ \rho_{21} & \rho_{22}\end{bmatrix}\begin{bmatrix}\dot{\mathbf{E}}_1^+\\ \dot{\mathbf{E}}_2^+\end{bmatrix}.$$

В сокращенной записи

$$\dot{\mathbf{E}}^- = \mathbf{S}\dot{\mathbf{E}}^+. \qquad (1.52)$$

Матрицу, объединяющую коэффициенты связи между параметрами падающих и отраженных волн, называют *волновой*. Известно несколько ее видов. Например, волновая *матрица передачи* Т связывает амплитуды падающей и отраженных волн в первой среде (на входе) с амплитудами волн во второй среде (на выходе):

$$\begin{bmatrix}\dot{\mathbf{E}}_1^-\\ \dot{\mathbf{E}}_1^+\end{bmatrix} = \begin{bmatrix}T_{11} & T_{12}\\ T_{21} & T_{22}\end{bmatrix}\begin{bmatrix}\dot{\mathbf{E}}_2^+\\ \dot{\mathbf{E}}_2^-\end{bmatrix}.$$

Волновая матрица

$$\mathbf{S} = \begin{bmatrix}\rho_{11} & \rho_{12}\\ \rho_{21} & \rho_{22}\end{bmatrix}, \qquad (1.52а)$$

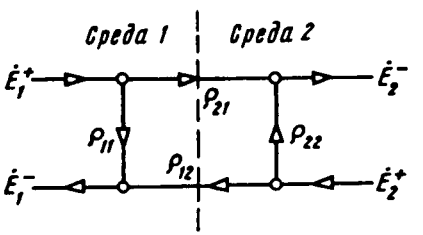


Рис. 1.1. Граф прямых и обратных волн на границе раздела двух сред

использованная в (1.52), носит название волновой *матрицы рассеяния*. Она связывает амплитуды всех уходящих волн с амплитудами приходящих.

Элементы волновой матрицы (1.52а) в виде (1.48)—(1.51) справедливы только для двух частных случаев поляризации. В общем случае, когда у приходящей волны плоскость поляризации составляет угол θ с плоскостью падения, ее представляют в виде суммы двух волн: $\dot{\mathbf{E}}^+ = \dot{\mathbf{E}}_p^+ + \dot{\mathbf{E}}_s^+$ одна из них $\dot{E}_p^+ = \dot{E}^+ \cos\theta$ поляризована параллельно плоскости падения, вторая $\dot{E}_s^+ = \dot{E}^+ \sin\theta$ — нормально к плоскости падения.

В *изотропной среде* соответствующие компоненты отраженных и преломленных волн сохраняют поляризацию компонент первичной волны:

$$\dot{\mathbf{E}}_p^- = \rho_{11}^{\parallel}\dot{\mathbf{E}}_p^+, \quad \dot{\mathbf{E}}_{2p}^- = \rho_{21}^{\parallel}\dot{\mathbf{E}}_p^+;$$
$$\dot{\mathbf{E}}_s^- = \rho_{11}^{\perp}\dot{\mathbf{E}}_s^+, \quad \dot{\mathbf{E}}_{2s}^- = \rho_{11}^{\perp}\dot{\mathbf{E}}_s^+.$$

Поэтому отраженная волна

$$\dot{\mathbf{E}}_1^- = \dot{\mathbf{E}}_p^- + \dot{\mathbf{E}}_s^- = \rho_{11}^{\parallel}\dot{\mathbf{E}}_p^+ + \rho_{11}^{\perp}\dot{\mathbf{E}}_s^+$$

и суммарный коэффициент отражения для прозрачных сред

$$\rho_{11} = \frac{\dot{\mathbf{E}}_1^-}{\dot{\mathbf{E}}_1^+} = \sqrt{(\rho_{11}^{\parallel})^2 \cos\theta + (\rho_{11}^{\perp})^2 \sin^2\theta}.$$

В *анизотропной среде* соотношения между компонентами падающих и отраженных волн приобретают тензорный характер:

$$\dot{E}_s^- = \rho_{ss}\dot{E}_s^+ + \rho_{sp}\dot{E}_p^+;$$
$$\dot{E}_p^- = \rho_{ps}\dot{E}_s^+ + \rho_{pp}\dot{E}_p^+, \qquad (1.53)$$

т. е.

$$\begin{bmatrix}\dot{E}_s^- \\ \dot{E}_p^-\end{bmatrix} = \begin{bmatrix}\rho_{ss} & \rho_{sp} \\ \rho_{ps} & \rho_{pp}\end{bmatrix}\begin{bmatrix}\dot{E}_s^+ \\ \dot{E}_p^+\end{bmatrix}$$

или $\dot{\mathbf{E}}^- = \mathbf{S}_{sp}\dot{\mathbf{E}}^+$, где *матрица отражения*

$$\mathbf{S}_{sp} = \begin{bmatrix}\rho_{ss} & \rho_{sp} \\ \rho_{ps} & \rho_{pp}\end{bmatrix}, \qquad (1.52б)$$

причем в изотропной среде $\rho_{ss} = \rho_{11}^{\perp}$, $\rho_{pp} = \rho_{11}^{\parallel}$, $\rho_{ps} = \rho_{sp} = 0$.

Выражение, аналогичное (1.52б), можно получить для *матрицы преломления*, определяющей соотношение между компонентами преломленной и падающей волн.

§ 1.7. Отражение и преломление волн на границе раздела металла и диэлектрика

Познакомимся с некоторыми особенностями электромагнитного поля на границе *прозрачного* диэлектрика ($\varepsilon'' = \mu'' = 0$) и *идеального* проводника, в котором токами смещения можно пренебречь по сравнению с токами проводимости [$\varepsilon_r = -j\sigma/(\omega\varepsilon_0)$] и можно не учитывать магнитные потери ($\mu'' = 0$). При этом существенно, что вплоть до сверхвысоких частот (оптический диапазон требует дополнительного исследования) соотношение между параметрами диэлектриков и проводников обычно таково, что волновые сопротивления и числа связаны неравенствами

$$z_1 \gg z_2, \qquad k_2 \gg k_1. \tag{1.54}$$

Согласно закону преломления (1.45) $\underline{\mathbf{k}}_1^+ \times \mathbf{e}_q = \underline{\mathbf{k}}_2 \times \mathbf{e}_q$ или

$$\mathbf{k}_1^+ \times \mathbf{e}_q = \mathbf{k}_2' \times \mathbf{e}_q - j\mathbf{k}_2'' \times \mathbf{e}_q. \tag{1.55}$$

Отсюда следует, что, во-первых, $\mathbf{k}_2'' \times \mathbf{e}_q = 0$, т. е. $\mathbf{k}_2'' \| \mathbf{e}_s \| \mathbf{e}_q$, и, во-вторых, $k_1 \sin\theta_1 = k_2' \sin\theta_2'$, т. е. $\sin\theta_2' = \dfrac{k_1}{k_2'}\sin\theta_1 \approx 0$, или $\theta_2' \approx 0$, и приблизительно $\mathbf{k}_2' \| \mathbf{e}_q$.

Таким образом, структура волны в м е т а л л е такова, что ее волновая и амплитудная нормали практически (с высокой степенью точности) совпадают с нормалью к поверхности металла — волна распространяется как однородная перпендикулярно поверхности практически не зависимо от угла падения θ_1. При этом векторы напряженностей поля преломленной волны параллельны поверхности:

$$\mathbf{E}_2\mathbf{e}_q = 0, \qquad \mathbf{H}_2\mathbf{e}_q = 0, \tag{1.56}$$

а соотношение между ними то же, что и в плоской волне (1.13).

Поскольку выражение (1.13) справедливо для любой точки во второй среде, где нормальные составляющие поля отсутствуют, то это позволяет определить связь и между тангенциальными составляющими поля границы раздела металла и диэлектрика:

$$\dot{\mathbf{E}}_t = Z_2\dot{\mathbf{H}}_t \times \mathbf{e}_q. \tag{1.57а}$$

Если использовать понятие *поверхностного сопротивления* $Z_s = \dot{\mathbf{E}}_t/\dot{\mathbf{H}}_t$, которое в рассматриваемом случае оказывается равным волновому сопротивлению Z_2, то граничное условие (1.57а) можно переписать в виде

$$\dot{\mathbf{E}}_t = Z_s \dot{\mathbf{H}}_t \times \mathbf{e}_q. \qquad (1.57б)$$

Соотношения (1.57а) и (1.57б) называют *приближенным граничным условием* Щукина—Рытова—Леонтовича. Приближение тем точнее, чем меньше нормальные составляющие поля во второй среде. В большинстве практических задач связанная с этим погрешность пренебрежимо мала. Это же относится и к применимости (1.57) для криволинейных границ и неоднородных сред, если радиус кривизны намного превышает длину волны (во второй среде и на ее протяжении изменение параметров невелико).

Преломленная волна распространяется в глубь металла как затухающая однородная волна. Затухание определяется мнимой компонентой волнового числа

$$\underline{k}_2 = k_2' - jk_2'' = \omega\sqrt{\underline{\varepsilon}_2\underline{\mu}_2} = \sqrt{-j\frac{\sigma}{\omega}\mu_2} = \frac{1-j}{\Delta},$$

где

$$\Delta = \sqrt{2/(\omega\mu\sigma)} \qquad (1.58)$$

— *глубина проникновения*, или *толщина скин-слоя*, на протяжении которой амплитуда уменьшается в e раз, а мощность — в e^2 раз. Глубина проникновения уменьшается по мере увеличения частоты электромагнитных колебаний.

На высоких частотах поле $\mathbf{E}$ и ток $\mathbf{J} = \sigma\mathbf{E}$ практически сосредоточены в тонком слое у поверхности проводника, что позволяет ввести понятие *поверхностного тока* проводника. Плотностью поверхностного тока $\mathbf{A}$ называют поток вектора плотности тока $\mathbf{J}$ через поперечное сечение S в виде прямоугольника, одна из сторон которого единичной длины лежит на поверхности проводника, а другая бесконечно простирается в глубь металла. Если, например, $\mathbf{e}_q = \mathbf{e}_z$ и начало координат совмещено с поверхностью проводника, то

$$\dot{\mathbf{A}} = \int_S \dot{\mathbf{J}}\,\mathrm{d}S = \sigma\int_S \dot{\mathbf{E}}_2^-\,\mathrm{d}S = \sigma\dot{\mathbf{E}}_t\int_0^\infty \mathrm{e}^{-j\underline{k}_2 z}\,\mathrm{d}z =$$

$$= \frac{\sigma\dot{\mathbf{E}}_t}{j\underline{k}_2} = \frac{-j\dfrac{\sigma}{\omega}}{\sqrt{\underline{\varepsilon}_2\underline{\mu}_2}}\dot{\mathbf{E}}_t = \frac{\dot{\mathbf{E}}_t}{\sqrt{\underline{\mu}_2/\underline{\varepsilon}_2}} = \frac{\dot{\mathbf{E}}_t}{Z_s} = \dot{\mathbf{H}}_t \times \mathbf{e}_q. \qquad (1.59)$$

Плотность поверхностного тока совпадает по значению, фазе и размерности с напряженностью магнитного поля на поверхности, а по направлению — с направлением вектора напряженности электрического поля (тангенциального).

Из выражения (1.59) следует, что поверхностное сопротивление можно определить также из соотношения

$$Z_s = \frac{\dot{E}_t}{A} = \sqrt{\frac{\underline{\mu}_2}{\underline{\varepsilon}_2}} = \sqrt{\frac{\underline{\mu}_2}{-j\sigma/\omega}} = \frac{1+j}{\sigma\Delta} = R_s + jx_s.$$

При достаточной глубине металла вся энергия, проникающая извне вместе с преломленной волной, целиком рассеивается в проводящей среде. Поглощаемая энергия, отнесенная к единице времени и поверхности металла, может быть найдена из выражения комплексного вектора Пойнтинга на границе

$$\underline{\mathbf{\Pi}}_s = \frac{1}{2}\dot{E}_t\dot{H}_t = \frac{1}{2}Z_s H_t^2 = \frac{1+j}{2\sigma\Delta}H_t^2.$$

Потери, или *активная мощность на единицу поверхности*,

$$P_s = \mathrm{Re}\,\underline{\mathbf{\Pi}}_s = H_t^2/(2\sigma\Delta) \qquad (1.60)$$

могут быть подсчитаны по известной на границе напряженности магнитного поля в диэлектрике ($\mathbf{H}_1 = \mathbf{H}_t = \mathbf{H}_2$).

Поле в диэлектрике на поверхности металла определяется суммой падающей и отраженной волн. На границе раздела прозрачного диэлектрика и хорошего проводника из формул Френеля (1.48)—(1.51) и (1.54) следует, что с малой погрешностью коэффициенты отражения и преломления

$$\rho_{11} = -1, \quad \rho_{21} = 0. \qquad (1.61)$$

Следовательно, при рассмотрении поля в диэлектрике можно считать, что амплитуда отраженной волны равна амплитуде падающей волны ($E_1^- = E_1^+$), поле в проводнике можно не учитывать ($E_2^- = 0$). В этом приближении согласно граничному условию (1.39) $\mathbf{E}_1 \times \mathbf{e}_q = 0$, т. е. $\mathbf{E}_1 \| \mathbf{e}_q$, или

$$E_{1t} = 0 \qquad (1.62)$$

— вектор напряженности электрического поля в диэлектрике на поверхности металла имеет только нормальную составляющую.

Из граничного условия для нормальных составляющих магнитной индукции $\mathbf{B}_1\mathbf{e}_q = \mathbf{B}_2\mathbf{e}_q$ с учетом (1.56) следует, что $\mu_1\mathbf{H}_1\mathbf{e}_q = \mu_2\mathbf{H}_2\mathbf{e}_q = 0$, т. е. $\mathbf{H}_1 \perp \mathbf{e}_q$ или

$$H_{1n} = 0 \qquad (1.63)$$

— вектор напряженности магнитного поля на поверхности металла и в диэлектрике направлен вдоль его границы и не имеет нормальной составляющей. Следовательно, при $\mathbf{re}_q = 0$

$$\mathbf{H}_1 = \mathbf{H}_t = \mathbf{H}_2. \tag{1.64}$$

§ 1.8. Отражение и преломление волн на границе раздела прозрачных диэлектриков

В немагнитных средах без потерь углы падения и преломления согласно (1.47)

$$\frac{\sin\theta_1}{\sin\theta_2} = \sqrt{\frac{\varepsilon_2}{\varepsilon_1}} = \frac{v_1}{v_2} = \frac{z_1}{z_2} = N. \tag{1.65}$$

Это упрощает формулы Френеля (1.48)—(1.51):

$$\rho_{11}^{\perp} = \frac{\sin\theta_2\cos\theta_1 - \sin\theta_1\cos\theta_2}{\sin\theta_2\cos\theta_1 + \sin\theta_1\cos\theta_2} = \frac{\sin(\theta_2 - \theta_1)}{\sin(\theta_2 + \theta_1)}; \tag{1.66}$$

$$\rho_{11}^{\parallel} = \frac{\sin\theta_2\cos\theta_2 - \sin\theta_1\cos\theta_1}{\sin\theta_2\cos\theta_2 + \sin\theta_1\cos\theta_1} = \frac{\mathrm{tg}(\theta_2 - \theta_1)}{\mathrm{tg}(\theta_2 + \theta_1)};$$

$$\rho_{21}^{\perp} = \frac{2\sin\theta_2\cos\theta_1}{\sin(\theta_1 + \theta_2)}; \qquad \rho_{21}^{\parallel} = \frac{2\sin\theta_2\cos\theta_1}{\sin(\theta_1 + \theta_2)\cos(\theta_1 - \theta_2)}. \tag{1.67}$$

Из полученных выражений следует, что при

$$\theta_1 + \theta_2 = 90° \tag{1.68}$$

в отраженной волне исчезает параллельно поляризованная компонента ($H_{\parallel}^{-} = 0$, $E_{\parallel}^{-} = 0$). Угол $\theta_1 = \theta_Б$, определяемый условием (1.68), называют *поляризующим углом* или *углом Брюстера*. Он может быть найден из совместного решения уравнений (1.65) и (1.68):

$$\mathrm{tg}\,\theta_Б = \sqrt{\varepsilon_2/\varepsilon_1} = n_2/n_1 = N. \tag{1.69}$$

При таком угле падения в отраженной волне может остаться лишь одна нормально поляризованная компонента. Это свойство отражающей поверхности может использоваться для поляризации неполяризованных волн, например солнечных лучей, или излучения тепловых источников света.

Отсутствие отражения у параллельно поляризованной волны означает, что вся энергия, приходящая с падающей волной, поступает во вторую среду вместе с преломленной волной. В этом заключается явление *полного прохождения* через границу двух сред. Пластинка, наклоненная под поляризующим углом к параллельно поляризованной волне, ока-

зывается полностью прозрачной, тогда как прохождение нормально поляризованной волны сопровождается потерями энергии с отраженными волнами. Стопу таких пластин используют как простейший, хотя и несовершенный, *поляризатор*. В проходящем через стопу пластин луче параллельно поляризованные компоненты проходят практически без изменения, а нормально поляризованные ослабляются тем больше, чем больше пластин в стопе.

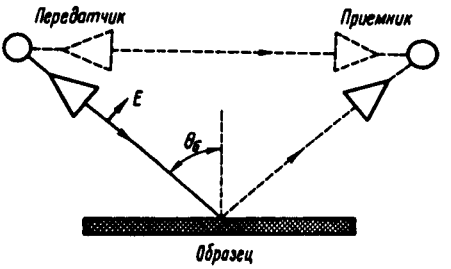


Рис. 1.2. Измерение поляризующего угла

На измерении поляризующего угла основан один из методов определения диэлектрической проницаемости ($\varepsilon_{r2} = \mathrm{tg}^2\,\theta_Б$ при $\varepsilon_{r1} = 1$). Схема измерений приведена на рис 1.2, причем передающую и принимающую антенны крепят на поворачивающихся одновременно и одинаково плечах гониометра, а необходимое направление поляризации устанавливают соответствующей ориентацией этих антенн. Приемник не нуждается в калибровке, поскольку он используется в качестве индикатора, обнаруживающего исчезновение отражения при регистрируемом угле. Измерение поляризующего угла позволяет найти действительную часть диэлектрической проницаемости с ошибкой, меньшей 2%, и при небольших потерях в диэлектрике ($\mathrm{tg}\,\delta \leqslant 0{,}2$), так как в этом случае можно считать, что при данном угле коэффициент отражения минимален.

Явление, в некотором отношении противоположное полному прохождению, возникает на границе двух сред при $\varepsilon_1 > \varepsilon_2$ и угле падения

$$\theta_1 > \theta_k = \arcsin N. \tag{1.70}$$

В этом случае угол преломления имеет комплексный характер, так как $\sin\theta_2 = \sin\theta_1 / N > 1$; $\cos\theta_2 = -\sqrt{1-\sin^2\theta_2} = -j\sqrt{\sin^2\theta_1/N^2 - 1}$. Второй возможный знак перед радикалом здесь отброшен, так как в рассматриваемых условиях он не имеет физического смысла (противоречит *принципу излучения* — требованию, чтобы волна была уходящей). Тогда в соответствии с (1.48), (1.50) и (1.65)

$$\rho_{11}^{\perp} = \frac{\cos\theta_1 + j\sqrt{\sin^2\theta_1 - N^2}}{\cos\theta_1 - j\sqrt{\sin^2\theta_1 - N^2}} = e^{j\varphi_\perp};$$

$$\rho_{11}^{\parallel} = \frac{N^2\cos\theta_1 + j\sqrt{\sin^2\theta_1 - N^2}}{N^2\cos\theta_1 - j\sqrt{\sin^2\theta_1 - N^2}} = e^{j\varphi_\parallel}. \tag{1.71}$$

Равенство единице модулей коэффициентов отражения свидетельствует о *полном отражении*, при котором $E_1^- = E_1^+$ и энергия отраженной волны равна энергии падающей волны. Средняя мощность преломленной волны в этом случае равна нулю, однако преломленная волна не исчезает, она принимает новую и неожиданную форму.

Структура поля преломленной волны может быть получена из анализа волнового вектора этой волны.

Предположим, что граница раздела совмещена с плоскостью xy ($\mathbf{e}_q = \mathbf{e}_z$), а плоскость падения — с плоскостью xz ($\mathbf{e}_n = \mathbf{e}_y$). Тогда проекции волнового вектора преломленной волны

$$k_x = k_2 \sin\theta_2 = k_2 \sin\theta_1 / N = k_2', \quad k_y = 0;$$

$$k_z = k_2 \cos\theta_2 = -jk_2\sqrt{\sin^2\theta_1/N^2 - 1} = -j/\Delta_0 = -jk_2'',$$

где $k_2 = \omega\sqrt{\varepsilon_2\mu_2}$; $\Delta_0 = N(k_2\sqrt{\sin^2\theta_1 - N^2})^{-1}$.

Следовательно, преломленная волна оказалась неоднородной с ортогональными направлениями мнимой и вещественных компонент волнового вектора (1.21), причем амплитудная нормаль ($\mathbf{e}_s = \mathbf{e}_z = \mathbf{e}_q$) направлена в глубь проводника, а фазовая нормаль ($\mathbf{e}_n = \mathbf{e}_x \perp \mathbf{e}_q$) — вдоль поверхности раздела диэлектриков.

Мгновенные значения напряженности электрического поля преломленной волны (1.14) в данном случае выражаются уравнением

$$E_2(t) = E_2^- \, e^{-z/\Delta_0} \sin(\omega t - k_x x).$$

Полученное выражение описывает тот тип неоднородных волн, которые получили название *поверхностных*. Такие волны распространяются вдоль границы раздела диэлектриков с фазовой скоростью

$$v_2 = \frac{dx}{dt} = \frac{\omega}{k_x} = \frac{v_1}{\sin\theta_1} > v_1$$

и экспоненциально затухают по мере удаления от поверхности в соответствии с глубиной проникновения Δ_0. Только при критическом угле падения $\theta_{кр}$ (1.70) затухание во второй среде исчезает и поверхностная волна становится однородной.

В случае полного отражения при высоких частотах возникает поверхностная волна, амплитуда которой практически отлична от нуля только в тонком слое, прилегающем к границе раздела диэлектриков. Граница раздела служит в качестве направляющей для распространения поверхностной волны, что лежит в основе применения *диэлектрических волноводов*.

Полное отражение от диэлектрика в части энергетического баланса в первой среде аналогично отражению от хорошо проводящей металлической поверхности. Однако фазовые соотношения (1.71) совершенно иные. При отражении от идеального проводника сдвиг фаз между отраженной и падающей волнами равен 180°, постоянен при любых поляризациях волны и не зависит от угла падения. При полном отражении от диэлектрика сдвиг фаз для нормально и параллельно поляризованных волн различен и зависит от угла падения. Таким образом, в общем случае отраженная волна оказывается эллиптически поляризованной, хотя, подбирая соответствующим образом направление поляризации, угол падения и материал диэлектрика, отмеченное явление можно использовать для получения из линейно поляризованной падающей волны отраженной волны с круговой поляризацией.

§ 1.9. Интерферометр

Отражение электромагнитных волн от поверхности диэлектриков и проводников используется в различных технических устройствах (например, в волноводах и антеннах) и в измерительной технике. Однако в ряде случаев отражение волн становится нежелательным явлением. Потеря энергии, связанная о образованием отраженных волн, уменьшает передаваемую мощность в оптических устройствах, волноводах, диэлектрических обтекателях и герметизирующих колпачках высотных антенн. Если не принять мер по радиотехнической маскировке, то отражение от металлических поверхностей самолетов, ракет и наземных предметов позволяет радиолокаторам обнаружить расположение объектов оборонного значения. Один из способов создания неотражающих покрытий состоит в использовании слоев диэлектриков и проводников. Многослойные структуры применяются также для исследования их электромагнитных свойств. Во всех случаях, подобных перечисленным, происходит взаимодействие или интерференция падающих и отраженных волн. Типичным примером может служить четырехслойная структура, изображенная на рис. 1.3.

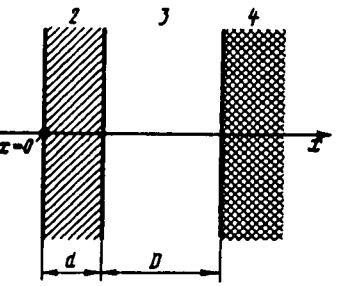


Рис. 1.3. Расположение диэлектрика *2* в устройстве для исследования его электромагнитных свойств

Пусть падающая из среды *1* плоская волна распространяется нормально к границам раздела в направлении оси x. Необходимо определить тот результирующий коэффициент отражения ρ от первой границы ($x = 0$), который определяется интерференцией падающих и

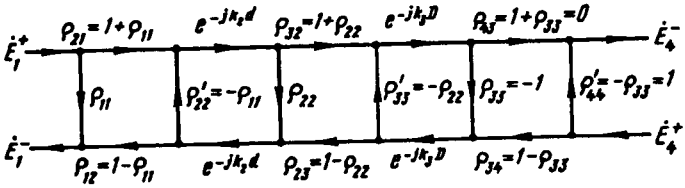


Рис. 1.4. Ориентированный граф прямых и обратных волн в четырехслойной структуре рис. 1.3

отраженных волн в средах *2*, *3* и *4*. При этом предполагают, что среда *3* хороший диэлектрик, а среда *4* обладает хорошей проводимостью (металл).

Расчет проведем, используя ориентированный граф многослойной системы (рис. 1.4). Передачи соответствующих ветвей найдем по формулам Френеля ($\theta_1 = 0$). На первой границе раздела для прямой волны

$$\rho_{11} = \frac{Z_2 - Z_1}{Z_2 + Z_1}; \qquad \rho_{21} = \frac{2Z_2}{Z_2 + Z_1} = 1 + \rho_{11};$$

для обратной волны

$$\rho'_{22} = \frac{Z_1 - Z_2}{Z_1 + Z_2} = -\rho_{11}; \qquad \rho_{12} = \frac{2Z_1}{Z_1 + Z_2} = 1 - \rho_{11}.$$

Аналогично на второй границе

$$\rho_{22} = \frac{Z_3 - Z_2}{Z_3 + Z_2}, \quad \rho_{32} = 1 + \rho_{22}, \quad \rho_{23} = 1 - \rho_{22}, \quad \rho'_{33} = -\rho_{22}.$$

На поверхности металла $\rho_{33} = -1$, $\rho_{43} = 1 + \rho_{33} = 0$.

По правилу некасающихся контуров (Мезона) результирующий коэффициент отражения от первой границы

$$\rho = \frac{\dot{E}_1^-}{\dot{E}_1^+} = \frac{T_1[1 - L_1^{(1)} - L_2^{(1)} - L_3^{(1)} + L_1^{(2)}] + T_2[-1 - L_2^{(1)}] + T_3}{1 - L_1^{(1)} - L_2^{(1)} - L_3^{(1)} + L_1^{(2)}}, \qquad (1.72)$$

где $T_1 = \rho_{11}$, $T_2 = \rho_{21} e^{-jk_2d} \rho_{22} e^{-jk_2d} \rho_{12}$, $T_3 = \rho_{21} e^{-jk_2d} \rho_{32} e^{-jk_3D} \rho_{33} e^{-jk_3D} \times$ $\times \rho_{23} e^{-jk_2d} \rho_{12}$ — передачи путей; $L_1^{(1)} = \rho_{11}\rho_{22} e^{-j2k_2d}$, $L_2^{(1)} = \rho_{22} e^{-j2k_3D}$, $L_3^{(1)} = \rho_{11}(1-\rho_{22}^2) e^{-j2(k_2d+k_3D)}$ — передачи контуров первого рода; $L_1^{(2)} = L_1^{(1)} L_2^{(1)} = -\rho_{11}\rho_{22}^2 e^{-j2(k_2d+k_3D)}$ — передача контура второго рода.

Если проводящая среда отсутствует или удалена достаточно далеко ($D = \infty$), то коэффициент отражения по (1.72)

$$\rho_{\infty} = \frac{\rho_{11} + \rho_{22}\, e^{-2j\underline{k}_2 d}}{1 + \rho_{11}\rho_{22}\, e^{-j2\underline{k}_2 d}}. \qquad (1.73)$$

Если металлический слой вплотную придвинут ко второй среде ($D = 0$), то

$$\rho_0 = \frac{\rho_{11} - e^{-2j\underline{k}_2 d}}{1 - \rho_{11}\, e^{-2j\underline{k}_2 d}}. \qquad (1.74)$$

Если третий слой образован прозрачным диэлектриком ($k_3 = 2\pi/\lambda_3$) четвертьволновой толщины ($D = \lambda_3/4$), то

$$\rho_{\lambda/4} = \frac{\rho_{11} + e^{-2j\underline{k}_2 d}}{1 + \rho_{11}\, e^{-2j\underline{k}_2 d}}. \qquad (1.75)$$

Комплексный коэффициент отражения $\rho = |\rho| e^{j\psi}$ может быть найден по измеренным в первой среде (воздухе) коэффициенту стоячей волны (КСВН) и положению (x_0) первого минимума стоячей волны:

$$|\rho| = \frac{\text{КСВН} - 1}{\text{КСВН} + 1}, \quad \psi = \pi\left(4\frac{x_0}{\lambda_1} - 1\right).$$

Поэтому измерения коэффициентов (1.74) и (1.75) в режимах, эквивалентных режимам *короткого замыкания* и *холостого хода*, дают возможность получить два уравнения, из которых могут быть найдены сначала волновое число $\underline{k}_2$ и волновое сопротивление Z_2, а затем, если необходимо, диэлектрическая и магнитная проницаемости: $\underline{\varepsilon}_2 = \underline{k}_2/(\omega Z_2)$, $\underline{\mu}_2 = \underline{k}_2 Z_2/\omega$.

Таким образом, система сред (см. рис. 1.3) может выполнять функции *интерферометра* — устройства для измерения параметров (второй) среды, основанного на интерференции падающих и отраженных волн.

Плоскопараллельная пластина представляет собой частный случай интерферометра, изобретенного и использованного в 1896 г. Ч. Фабри и А. Перо для измерения длин волн света. В общем случае *интерферометр Фабри-Перо* состоит из двух параллельных полупрозрачных плоских зеркал, разделенных каким-либо прозрачным веществом. Для пластины, находящейся в воздухе ($n_1 = n_3 = 1$) и имеющей показатель преломления $n_2 = n$, коэффициент отражения на границе диэлектриков без потерь

$$\rho_{11} = \frac{\dot{E}_1^-}{\dot{E}_1^+} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{1-n}{1+n}.$$

Энергетический коэффициент отражения на границе диэлектриков

$$R = I_1^- / I_1^+ = (E_1^- / E_1^+)^2 = \rho_{11}^2,$$

где I — интенсивность света.

Согласно (1.73), энергетический коэффициент отражения от поверхности пластины

$$R_\infty = |\rho_\infty|^2 = R\left|\frac{1 - e^{-j4\pi d/\lambda_2}}{1 - Re^{-j4\pi d/\lambda_2}}\right|^2 = R\left|\frac{e^{j\pi p} - e^{-j\pi p}}{e^{j\pi p} - Re^{-j\pi p}}\right|^2 =$$

$$= R\left|\frac{2j\sin\pi p}{(1-R)\cos\pi p + j(1+R)\sin\pi p}\right|^2,$$

где $2\pi d/\lambda_2 = 2\pi nd/\lambda = \pi p$; λ — длина волны в воздухе; $p = 2nd/\lambda$ — нормированная оптическая длина пластины (nd), называемая *порядком интерференции*. Энергетический коэффициент пропускания (пропускание) пластины, равный отношению интенсивностей светового потока на выходе и входе интерферометра, зависит от оптической длины интерферометра и описывается *функцией Эйри*:

$$T = \frac{I_2}{I_1} = 1 - R_\infty = \frac{(1-R)^2}{(1-R)^2 + 4R\sin^2\pi p}.$$

График функции Эйри (рис. 1.5) имеет вид периодических максимумов, соответствующих целым значениям порядка интерференции, когда оптическая длина становится кратна половине длины волны падающего светового потока и его интенсивность на выходе (I_2) равна его интенсивности на входе (I_1), т. е. $T = 1$. В этом случае фазы прямых и обратных световых пучков в интерферометре совпадают. Минимальная интенсивность светового потока на выходе наблюдается при полуцелых значениях порядка интерференции, когда интерферирующие пучки находятся в противофазе.

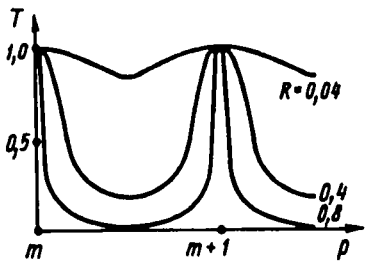


Рис. 1.5. Зависимость коэффициента пропускания плоскопараллельной пластинки от порядка интерференции

Ширина пика на уровне $T = 0{,}5$

$$\Delta p \approx (1-R)/\pi\sqrt{R}.$$

Отношение расстояния между двумя соседними максимумами к ширине пика $F = \pi\sqrt{R}/(1-R)$ называют *критерием резкости*.

Чем больше коэффициенты отражения R, тем острее становятся максимумы кривой Эйри. При достаточно большом значении R ($R > 0{,}8$) функция Эйри описывает последовательность острых пиков пропускания, между ко-

торыми значение этой функции мало, и функция изменяется плавно. В этом случае при постепенном увеличении оптической длины пропускание будет меняться медленно до определенного порогового значения, после чего произойдет резкий скачок пропускания. Поэтому длина волны падающего светового потока может быть определена с большой точностью.

Сумма интенсивностей прямых и обратных световых пучков определяет интенсивность светового потока в интерферометре:

$$\Phi = T\frac{1+R}{1-R}I_1.$$

Поэтому при достаточно больших коэффициентах отражения ($R > 0{,}8$) полупрозрачных зеркал и совпадении фаз световых пучков ($T = 1$) интенсивность светового потока внутри интерферометра превышает интенсивность падающего светового потока не менее, чем на порядок.

Интерференция волн в четырехслойной структуре (см. рис. 1.3) используется также в простейших неотражающих слоях.

§ 1.10. Неотражающие покрытия

Во многих устройствах для разделения областей с разными давлениями газа на пути электромагнитной волны приходится располагать пластины диэлектрика толщиной d. В частности, это имеет место в обтекателях антенн. Даже в том случае, когда материал пластины (обтекателя) прозрачен ($k_2 = 2\pi/\lambda_2$), ее поверхность частично ослабляет проходящую волну при образовании отраженных волн. Этих потерь можно избежать правильным выбором толщины пластины.

Если второй диэлектрик на рис. 1.3 расположен между средами с одинаковыми электромагнитными свойствами ($Z_1 = Z_3$), то

$$\rho_{11} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = -\frac{Z_3 - Z_2}{Z_3 + Z_2} = -\rho_{22}$$

и согласно (1.73)

$$\rho_\infty = \rho_{11}\frac{1 - e^{-j4\pi d/\lambda_2}}{1 - \rho_{11}^2 e^{-j4\pi d/\lambda_2}}.$$

Из полученного выражения следует, что отсутствие отражения от обтекателя и пластинки будет обеспечено, если их толщина равна целому числу полуволн: $d = m\lambda_2/2$ при $m = 1, 2, 3, \ldots$. При этом $\rho_\infty = 0$ независимо от ρ_{11}, так как $1 - \exp(-j2k_2 d) = 1 - \exp(j2\pi m) = 0$. Потери в материале несколько ослабляют проходящую волну, однако полу-

волновая пластина по-прежнему создает оптимальные условия прохождения волны.

Аналогичная задача возникает при создании неотражающего покрытия на поверхности двух сред с разными электромагнитными свойствами ($Z_1 \neq Z_3$). Полное «просветление» требует выполнения двух условий: $\rho_{11} = \rho_{22}$ и $d = m\lambda_2/4$ при $m = 1, 3, 5, ...$, так как из (1.73)

$$\rho_\infty = \rho_{11} \frac{1 + e^{-j\pi m}}{1 + \rho_{11}^2 e^{-j\pi m}} = 0.$$

Следовательно, параметры «просветляющего слоя» четвертьволновой толщины должны быть такими, чтобы

$$Z_2/Z_1 = Z_3/Z_2 \text{ или } Z_2 = \sqrt{Z_1 Z_3}. \qquad (1.76)$$

Для прозрачных немагнитных диэлектриков $\varepsilon_2 = \sqrt{\varepsilon_1 \varepsilon_3}$.

Уравнение (1.76) напоминает аналогичное соотношение для волнового сопротивления трансформатора сопротивлений или четвертьволновой линии, служащей для согласования на определенной частоте двух линий с разными волновыми сопротивлениями. Здесь волновое сопротивление согласующего слоя равно среднему геометрическому от волновых сопротивлений согласуемых сред.

Устранение отражений от поверхности проводника является более сложной задачей ввиду разного характера комплексных волновых сопротивлений металла и диэлектрика. Она может быть решена, например, расположением тонкого проводника *2* на расстоянии $D = \lambda_3/4$ от поверхности металла *4* (см. рис. 1.3). Как следует из (1.75), параметры проводящего покрытия должны в этом случае удовлетворять условию

$$\frac{Z_2 - Z_1}{Z_2 + Z_1} + e^{-j2\underline{k}_2 d} = 0$$

или, если учесть, что в тонком слое $\text{th}\, j\underline{k}_2 d \approx j\underline{k}_2 d$ и $\underline{\varepsilon}_2 = -j\sigma/\omega$,

$$Z_1 = Z_2 \frac{1 + e^{-j2\underline{k}_2 d}}{1 - e^{-j2\underline{k}_2 d}} = \frac{Z_2}{\text{th}\, j\underline{k}_2 d} \approx \frac{Z_2}{j\underline{k}_2 d} = \frac{1}{j\omega\varepsilon_2 d} = \frac{1}{\sigma d}.$$

§ 1.11. Интерференционные зеркала и фильтры

Отражательная способность (1.73) однослойного покрытия зависит от его толщины и частоты электромагнитных колебаний. Их удобно характеризовать порядком интерференции $p = 2d/\lambda_2 = n_2 2d/\lambda_0$, т. е. нор-

мированным запаздыванием фазы волны, проникшей в слой и вернувшейся к его передней поверхности.

Коэффициент отражения оказывается осциллирующей функцией порядка интерференции. Например, для прозрачных диэлектриков, когда $n_1 = \sqrt{\varepsilon_{r1}}$, $n_2 = \sqrt{\varepsilon_{r2}}$, $n_3 = \sqrt{\varepsilon_{r3}}$ согласно (1.73)

$$\rho_\infty = \frac{(n_1 - n_2)(n_2 + n_3)\,e^{j\pi p} + (n_1 + n_2)(n_2 - n_3)\,e^{-j\pi p}}{(n_1 + n_2)(n_2 + n_3)\,e^{j\pi p} + (n_1 - n_2)(n_2 - n_3)\,e^{-j\pi p}} =$$

$$= \frac{n_2(n_1 - n_3)\cos\pi p + j(n_1 n_3 - n_2^2)\sin\pi p}{n_2(n_1 + n_3)\cos\pi p + j(n_1 n_3 + n_2^2)\sin\pi p}. \qquad (1.77)$$

Экстремум (1.77) $\rho_1 = (n_1 - n_3)(n_1 + n_3)^{-1}$ соответствует целым значениям $p = m$, когда поверхностный слой *никак* не сказывается на оптических свойствах покрытой им поверхности.

Наибольшее влияние на отражение электромагнитной волны поверхностный слой n_2 оказывает при получисленных значениях порядка интерференции $p = m + 1/2$ (m = 0, 1, 2, 3, ...), когда коэффициент отражения (1.77) достигает другого экстремума $\rho_2 = (n_1 n_3 - n_2^2)(n_1 n_3 + n_2^2)^{-1}$. Отражательная способность достигает максимума, если показатель преломления слоя больше или меньше показателей преломления прилегающих сред ($n_1 < n_2 > n_3$ или $n_1 > n_2 < n_3$). Такой слой называют *антипросветляющим покрытием.*

Нанесение на поверхность раздела сред *1* и *3* слоя *2* с такими свойствами влечет за собой увеличение отражательной способности первоначальной границы. Обусловлено это тем, что здесь возвращаемая слоем волна оказывается синфазной с волной от передней границы слоя. Коэффициент отражения тем больше, чем больше $|n_1 - n_2|$ и $|n_2 - n_3|$, т. е. чем больше амплитуды интерферирующих волн.

Верхняя граница коэффициента отражения антипросветляющего покрытия лимитируется только возможностью подбора веществ с достаточно большим или достаточно малым показателем преломления. Например, покрывая стекло (n_3 = 1,5) пленкой с n_2 = 3,0, можно получить $\rho_2^2 = 0{,}52$ вместо $\rho_{33}^2 = 0{,}04$ без пленки.

Однослойные диэлектрические покрытия обладают рядом недостатков: резко выраженной хроматичностью; трудностью, а иногда и невозможностью подбора материала, удовлетворяющего оптимальным требованиям; сильной зависимостью оптических свойств от угла падения, невозможностью получения покрытия с отражательной способностью, заметно превышающей 0,5. Перечисленные недостатки преодолены в мно-

гослойных покрытиях. Современная технология позволяет получать достаточно однородные слои, оптическая толщина которых на всем протяжении отклоняется от заданной всего на доли процента, что вполне достаточно для большинства практических целей.

Многослойные *интерференционные зеркала* или, как их чаще называют, *диэлектрические зеркала* имеют следующие преимущества перед металлическими. Во-первых, они обладают меньшими потерями на поглощение, что имеет особенно существенное значение и для порога генерации, и для излучаемой мощности квантовых генераторов. Во-вторых, применение диэлектрических зеркал позволяет преодолеть ограничения, налагаемые природой металлического отражения на диапазон эффективно отражаемых волн (возможно осуществление полноценных интерференционных систем вплоть до микроволнового диапазона). В-третьих, имеется возможность изготовления коррозионно-устойчивых интерференционных зеркал с любым заданным коэффициентом отражения, в том числе и незначительно отличающимся от единицы (недостижимо при использовании металлических зеркал). В-четвертых, диэлектрические зеркала оказываются более стойкими, чем металлические, при воздействии больших плотностей светового потока. В-пятых, многослойные интерференционные зеркала могут быть изготовлены таким образом, что, обладая крайне высоким коэффициентом отражения в одной области спектра, они оказываются почти прозрачными для других длин волн.

Последнее свойство может быть применено при разработке теплозащитных фильтров. Одной из областей их использования являются мощные кинематографические установки, мощность которых ограничивается нагреванием объективов и пленки обильным инфракрасным излучением источников света. Мощность проекторов может быть существенно увеличена, если применить «холодное зеркало», обладающее высоким коэффициентом отражения в видимой области и просветленным для инфракрасного излучения. Аналогичная задача не менее актуальна и при разработке квантовых генераторов с оптической накачкой, поскольку используемые при этом газоразрядные лампы лишь небольшую часть энергии выделяют в нужном диапазоне частот. При этом бóльшая часть энергии, не расходуемой на возбуждение, затрачивается на перегрев активного материала (например, рубина). Это увеличивает расход энергии на охлаждение или снижает выходную энергию и частоту следования импульсов когерентного излучения.

К простейшим светофильтрам относятся абсорбционные, в которых выделение нужной спектральной области осуществляется за счет поглощения (абсорбции) света средой в других участках спектра. Наряду с ними все большее распространение получают неабсорбционные светофильтры, выделяющие более узкие спектральные области путем использования интерференции, полного внутреннего отражения, рассеяния и т. п.

В отличие от абсорбционных фильтров подобные системы обладают значительно меньшими потерями света, возможностями перемещать по спектру полосу пропускания и работать в таких областях спектра, для которых подбор абсорбционных фильтров невозможен (например, в инфракрасном диапазоне). В отличие от монохроматоров неабсорбционные фильтры позволяют получать световые пучки с большим сечением и угловой апертурой.

Среди неабсорбционных фильтров наиболее простыми и удобными оказались *интерференционные фильтры*. Обычно они имеют вид стеклянной пластинки, одну из сторон которой последовательно покрывают полупрозрачной пленкой серебра, диэлектрической пленкой и вторым полупрозрачным металлическим слоем, после чего закрывают второй стеклянной пластиной и широкополосным абсорбционным светофильтром.

Интерференционный светофильтр можно рассматривать как видоизмененный вариант слоистого интерферометра, физические явления в котором аналогичны явлениям в многослойном покрытии со слоями с чередующимися значениями показателя преломления. Используется, в частности, то, что влияние диэлектрического слоя на коэффициент отражения (1.77) сводится к нулю при целочисленном значении порядка интерференции. Для одного слоя это соответствует максимальному значению коэффициента пропускания $T_1 = 1 - \rho_1$. Целочисленным значениям порядка интерференции этого слоя $p_1 = 2n_1 d/\lambda_m = m$ соответствует ряд длин волн λ_m, определяющих серию максимумов пропускания (рис. 1.6, *а*). Бóльшую часть этих максимумов можно погасить, если совместить два слоя, порядки интерференции которых находятся в рациональном отношении $k_1 p_1 = k_2 p_2$ (k_1 и k_2 — целые числа).

Непогашенными останутся только те максимумы первого слоя, которые соответствуют значениям p_1, кратным k_2. Например, при $k_1 = 4$ и $k_2 = 5$ (рис. 1.6, *б*) в проходящем луче будут заметны только максимумы, соответствующие $p_1 = 0, 5, 10, 15, \ldots$. Число максимумов в световом потоке, проходящем через два слоя, уменьшится, а расстояние между линиями пропускания увеличится в 5 раз (рис. 1.6, *в*).

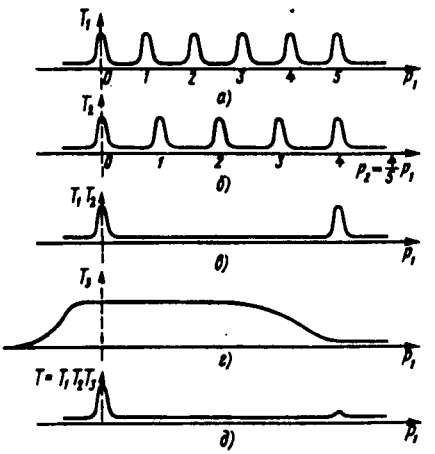


Рис. 1.6. Зависимость коэффициентов пропускания слоев интерференционного фильтра от порядка интерференции

Подбором третьего слоя, обладающего свойствами абсорбционного фильтра (рис. 1.6, *г*), можно выделить лишь один необходимый максимум пропускания (рис. 1.6, *д*).

В настоящее время изготовляют светофильтры с максимумом пропускания 90% и шириной полосы пропускания, достигающей долей ангстрема.

§ 1.12. Интерференция некогерентных волн

Две электромагнитные волны одинаковой частоты с амплитудами $\dot{\mathbf{E}}_1$ и $\dot{\mathbf{E}}_2$, приходя в одну точку с координатой **R**, образуют результирующее поле $\dot{\mathbf{E}} = \dot{\mathbf{E}}_1 + \dot{\mathbf{E}}_2$. В области пересечения двух волн в момент t возникает интерференция, при которой поля двух волн складываются либо со взаимным усилением, либо со взаимным ослаблением в зависимости от **R**, t и разности фаз $\varphi = \varphi_1 - \varphi_2$, образуя поле напряженностью $E(t) = E\sin(\omega t + \varphi)$. Для двух одинаково поляризованных волн, распространяющихся в одном направлении, квадрат амплитуды результирующего поля $E^2 = E_1^2 + E_2^2 + 2E_1E_2\cos\varphi$, а интенсивность излучения $J = J_1 + J_2 + 2\sqrt{J_1J_2}\cos\varphi$. При разности фаз $\varphi = \pi$ и $J_1 = J_2$ общая интенсивность излучения становится равной нулю, а при $\varphi = 0$ она достигает максимального значения $J = (\sqrt{J_1} + \sqrt{J_2})^2$.

Полученный результат справедлив только в том случае, когда разность фаз между полями двух волн остается постоянной за время наблюдения. Такие волны называют *когерентными*. Однако практически никогда не имеют дела с колебаниями, длящимися бесконечно. Например, световые лучи образуются волнами, излучаемыми атомами короткими цугами, возникая обычно (в тепловых источниках света) с разными, нерегулярно меняющимися фазами. В этом случае и результирующая интенсивность J также меняется с течением времени и практически при наблюдениях за время τ приходится иметь дело со средней интенсивностью

$$I = \bar{J} = \frac{1}{\tau}\int_0^{\tau} J\,\mathrm{d}t = E_1^2 + E_2^2 + 2E_1E_2\frac{1}{\tau}\int_1^{\tau}\cos\varphi\,\mathrm{d}t.$$

Для когерентных колебаний фаза остается неизменной в течение времени наблюдения. Поэтому

$$\overline{\cos\varphi} = \frac{1}{\tau}\int_0^{\tau}\cos\varphi\,\mathrm{d}t = \cos\varphi.$$

Средняя интенсивность при $\varphi = 0$

$$I = (\sqrt{I_1} + \sqrt{I_2})^2.$$

Если разность фаз колебаний беспорядочно меняется, многократно пробегая за время наблюдения все значения от 0 до 360°, то колебания называют *некогерентными*. При этом $\overline{\cos\varphi} = 0$ и $I = I_1 + I_2$, т. е. при сложении волн наблюдается суммирование только интенсивностей и интерференция не имеет места.

При сложении большого числа n некогерентных волн суммарная интенсивность излучения $I = \sum_{i=1}^{n} I_i$ и при одинаковой интенсивности ($I_i = I_0$) $I = nI_0$. При сложении когерентных волн с одинаковыми фазами $I = \left(\sum_{i=1}^{n} \sqrt{I_i}\right)^2$, а при одинаковых интенсивностях $I = n^2 I_0$. Таким образом, максимальный выигрыш в интенсивности когерентного излучения по сравнению с некогерентным получается в n раз.

Две немонохроматические волны, исходящие из различных источников, также не интерферируют, так как поля этих волн независимы, разность фаз между каждыми двумя составляющими меняется случайным образом и $\overline{\cos\varphi} = 0$. Спектр излучения большинства источников света состоит из совокупности монохроматических волн с различными частотами, амплитудами и начальными фазами. Особенностями излучения квантовых генераторов являются относительно малый разброс частот и узкополосность спектра. В этом случае

$$E(t) = \sum_{i=1}^{I} E_i \sin(\omega_i t - \varphi_i) = E_0(t)\sin[\omega_0 t + \Phi(t)], \qquad (1.78)$$

т. е. поле излучения можно представить в виде волны, которая обладает средней частотой $\omega_0 = 2\pi\nu$, но модулирована по амплитуде и фазе. Такие волны называют *квазимонохроматическими*.

Форма спектра источников излучения связана с механизмом излучения и с состоянием излучающей среды. Квазимонохроматическая волна может возникнуть, в частности, в результате того, что источник излучает одну гармонику с постоянной фазой Ф, но конечной длительности τ:

$$\dot{E}(t) = \begin{cases} E_0 e^{j\omega_0 t} & \text{при } |t| \leq \tau/2; \\ 0 & \text{при } |t| > \tau/2. \end{cases}$$

Интервал времени τ в этом случае называют *временем когерентности*.

Нетрудно найти его связь с шириной спектра. Действительно, амплитудный спектр волнового пакета

$$F(\omega) = E_0 \int_{-\tau/2}^{\tau/2} e^{j(\omega_0-\omega)t}\, dt = E_0 \frac{\sin(\omega-\omega_0)\tau/2}{(\omega-\omega_0)\tau/2},$$

а энергетический спектр

$$I(\omega) = |F(\omega)|^2 = E_0^2 \left[\frac{\sin(\omega-\omega_0)\tau/2}{(\omega-\omega_0)\tau/2}\right]^2.$$

Ширина энергетического спектра на половинном уровне

$$\Delta\omega = 2\pi/\tau. \qquad (1.79)$$

Связь между шириной спектра (1.79) и временем когерентности сохраняется и в том случае, когда расширение спектра обусловлено и другими причинами, например затуханием излучения, доплеровским эффектом, столкновениями атомов и т. п.

Длина когерентности

$$l = c_0\tau = c_0/\Delta\nu = \lambda_0^2/\Delta\lambda. \qquad (1.80)$$

Для лучших тепловых квазимонохроматических источников излучения длина когерентности составляет около метра. С помощью лазеров удается достичь длины когерентности в сотни километров. Для лазерного излучения с шириной $\Delta\lambda$ = 0,005 Å линии λ_0 = 5480 Å — длина когерентности равна 60 см, а время когерентности 2 нс. При этом в одном цуге укладывается около 10 периодов электромагнитных колебаний.

§ 1.13. Восстановление волнового фронта по интерферограмме

Системы хранения и передачи изображений до возникновения голографии характеризовались тем, что объектом наблюдения являлась не сама световая волна, а изображение, получающееся непосредственно из волнового поля. Это волновое поле возникает обычно в результате отражения от предмета или прохождения через него излучения от внешнего источника. Изображение формирует волновое поле, имеющее в плоскости фотоприемника (xy), например фотопластинки, некоторое распределение амплитуд $E(x, y)$, фаз $\varphi(x, y)$ и комплексных амплитуд:

$$\dot{E} = E(x, y)e^{j\varphi(x,y)}. \qquad (1.81)$$

Запаздывание и фаза волн, посылаемых от удаленных или толстых частей предмета, будут больше, чем у волн от более близких или тонких

частей предмета. Следовательно, без фазовой характеристики волнового поля нельзя получить полное пространственное изображение предмета.

Реакция оптических приемников света (фотоэмульсии, фотоэлемента, глаза) зависит только от энергии или интенсивности светового потока

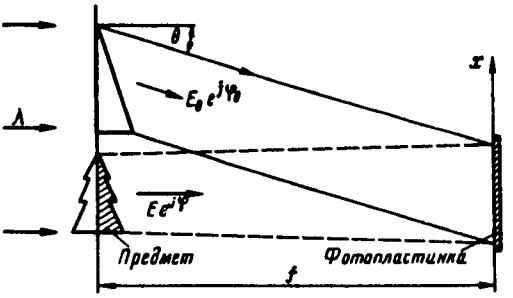


Рис. 1.7. Схема получения голограммы

$I(x, y) = \dot{E}\dot{E}^* = E^2$. Поэтому при непосредственной фотоиндикации лучей, пришедших только от предмета, информация о фазе φ волнового поля оказывается утраченной.

Голограммы относятся к новым средствам хранения и обработки электромагнитных сигналов. В отличие от фотографии, которая регистрирует двумерное распределение только *амплитуды* волны, распространяющейся от предмета, голограмма запоминает также и *фазу* волны. Именно поэтому первый корень («голос») слова «голограмма» в переводе с греческого означает «весь», «целый». Такие свойства фотоэмульсия приобретает благодаря тому, что она регистрирует одновременно с электромагнитным полем, образованным предметом, поле вспомогательной волны, называемой *опорной*, *подсветкой* или *фоном*.

Голограмма — это такая система полной записи пространственной структуры электромагнитной волны, которая образуется в результате ее интерференции с опорным лучом.

Простейший метод формирования голограммы связан с использованием наклонной когерентной подсветки, когда опорный луч образует однородная плоская волна, наклоненная под некоторым углом θ к первичной волне от предмета. Схема такой суперпозиции волн показана на рис. 1.7, где когерентность предметного и опорного лучей обеспечивается широкой однородной плоской волной, освещающей одновременно и предмет, и призму.

Призма, поворачивая фронт преломленной волны на угол θ, создает в плоскости фотопластинки (xy) поле с постоянной амплитудой E_0 и линейно меняющейся фазой $\varphi_0 = \alpha x$. Из схемы рис. 1.8, *а* видно, что изменение фазы опорного луча вызвано различной длиной пути $x' = |x| \sin\theta$ до точек пластинки с координатой x. Запаздывание фазы $\varphi_0 = -kx' = kx \sin\theta = \alpha x$, откуда для малых углов

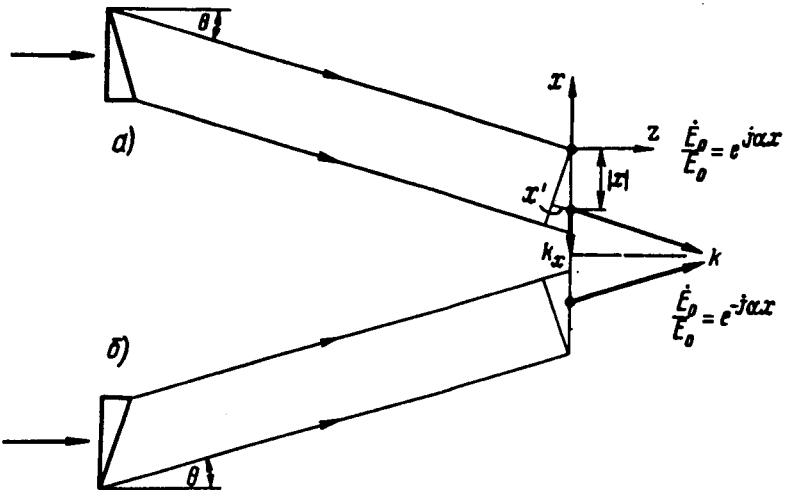


Рис. 1.8. Поворот волнового фронта призмой

$$\alpha = k\sin\theta = k_x = \frac{2\pi}{\lambda}\sin\theta \approx \frac{2\pi}{\lambda}\theta. \qquad (1.82)$$

Таким образом, коэффициент пропорциональности α определяется углом θ. Поскольку коэффициент характеризует скорость изменения фазы вдоль оси x, то естественно, что он совпадает с проекцией волнового вектора на эту ось.

Положительный знак фазы φ_0 опорного луча характерен для призмы, отклоняющей плоскую волну в отрицательную сторону ($\theta < 0$). При этом положительное направление оси x показывает направление роста опережающих фаз $\dot{E}_0$. Отрицательные фазы φ_0 соответствуют противоположному направлению отклонения ($\theta > 0$) опорного луча (рис. 1.8, *б*). В этом случае положительное направление оси x совпадает с направлением роста запаздывания фазы E_0. Рассмотренные выражения для комплексной амплитуды еще раз подчеркивают непосредственную связь между *фазой* волны и ее направлением в *пространстве*. В данном случае соответствие между ними выражает уравнение (1.82), а в общем случае — (1.2).

В результате интерференции предметного и опорного лучей на поверхности фотопластинки (см. рис. 1.7) амплитуда суммарного поля

$$\dot{E}_r = \dot{E}_0 + \dot{E} = E_0 e^{j\alpha x} + E(x)e^{j\varphi(x)}.$$

Реакция фотоэмульсии по-прежнему определяется интенсивностью

$$I_r(x) = \dot{E}_r\dot{E}_r^* = E_0^2 + E^2(x) + 2E_0E(x)\cos[\varphi(x) - \alpha x],$$

однако теперь она зависит не только от амплитуды предметного луча, но и от его фазы. Опорный луч позволяет сохранить на фотопластинке не только распределение амплитуд волны, но и распределение фаз, т. е. запоминает полную информацию о структуре регистрируемой волны (хотя и в закодированном виде).

Наблюдать полученную голограмму можно только после проявления и фиксирования фотопластинки. При этом почернение фотоэмульсии пропорционально интенсивности экспонировавшего светового потока, возведенной в некоторую степень, называемую *коэффициентом контрастности* ν. Энергетическое пропускание проявленной эмульсии обратно пропорционально почернению. Поэтому если работать на линейном участке характеристической кривой, то амплитудное пропускание $T(x) \sim I_r^{-\nu/2}$.

Обычно интенсивность опорного луча заметно выше, чем предметного ($E_0 >> E$). Это позволяет сохранить в разложении коэффициента пропускания в степенной ряд только члены первого порядка малости:

$$T(x) \doteq [E_0^2 + E^2 + 2E_0 E\cos(\varphi - \alpha x)]^{-\nu/2} \approx$$

$$\approx E_0^{-\nu/2}\left[E_0^2 - \frac{\nu}{2}E^2 - \nu E_0 E\cos(\varphi - \alpha x)\right] \doteq$$

$$\doteq 2E_0^2 - \nu E^2(x) - 2\nu E_0 E(x)\cos[\varphi(x) - \alpha(x)] =$$

$$= 2E_0^2 - \nu E^2(x) - \nu\dot{E}(x)E_0\, e^{-j\alpha x} - \nu\dot{E}^*(x)E_0\, e^{j\alpha x}. \qquad (1.83)$$

Полученную после проявления фотопластинку (пленку) называют голограммой, а уравнение (1.83) — *уравнением голограммы*.

Восстановление изображения исходного предмета осуществляется освещением голограммы однородной плоской волной (рис. 1.9). Плоская

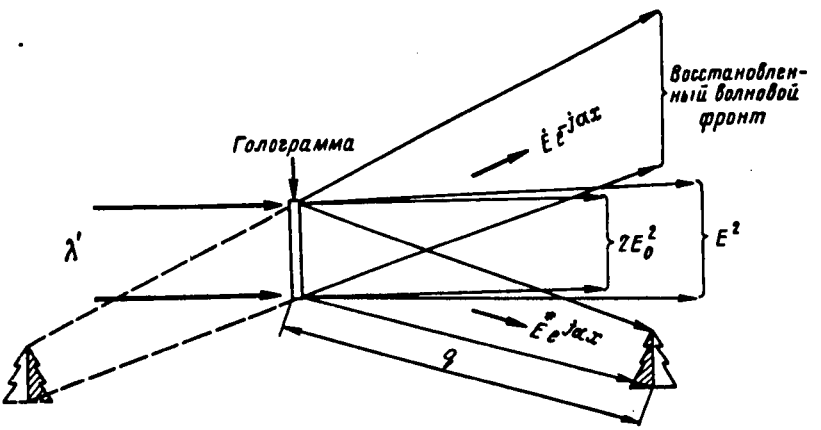


Рис. 1.9. Восстановление волнового фронта при освещении голограммы

волна с амплитудой $A(x) = \text{const}$ после прохождения через обработанную фотопластинку имеет амплитуду $AT(x)$, т. е. как бы умножается на значение пропускания $T(x)$. При этом на выходе пластинки возникают четыре компоненты, соответствующие четырем членам уравнения (1.83). Первый член является постоянной величиной, и поэтому он характеризует равномерное ослабление, не вызывающее вариации светового пучка в поперечном направлении (x). Второй член также соответствует ослаблению пучка, но неравномерному, что приводит к небольшой ($2E_0^2 \gg E^2$) дифракции прямого луча. Чтобы выяснить, какое преобразование однородной волны выражают третий и четвертый члены уравнения (1.83), необходимо вспомнить выражения для наклонных лучей.

Изменение направления распространения на угол $\pm\theta$ приводит, как показано на рис. 1.8, к появлению дополнительного сдвига фаз $\varphi_0 = \mp\alpha x$, а в выражении для комплексной амплитуды — дополнительного множителя $e^{\pm j\varphi_0}$. Поэтому третий член уравнения голограммы показывает, что соответствующий ему луч образуется в результате модуляции в поперечном направлении восстанавливающего луча A комплексной амплитудой предметной волны $\dot{E} = E(x)e^{j\varphi(x)}$ и отклонением его в положительном направлении на угол $\theta = \alpha\lambda/2\pi$ (за счет умножения на $e^{-j\alpha x}$). Аналогично, четвертая компонента уравнения голограммы означает умножение восстанавливающего луча на комплексно сопряженную амплитуду предметной волны и ее отклонение в отрицательном направлении на угол θ.

Поскольку распределение амплитуд и фаз по поперечному сечению третьего луча определяется амплитудой $\dot{E}$, то этот луч восстанавливает точную копию исходной волны (1.81). Волна, образуемая четвертым лучом, отличается от предметной только знаком фазы; она как бы распространяется «назад», т. е. как бы возвращается к исходному предмету.

Таким образом голография открыла принципиальную возможность с помощью двухмерной, плоской интерферограммы получить высококачественное трехмерное, *объемное* изображение, а также перспективы создания трехмерного цветного телевидения. Отображение и увеличение объектов могут быть получены без помощи линз. Можно смещать точку наблюдения, рассматривая предмет по голограмме под разными ракурсами. Голография позволяет осуществлять интерференцию между двумя световыми лучами, не перекрывающимися ни во времени, ни в пространстве. Наблюдатель может корректировать оптические свойства прибора, используемого в эксперименте, после того как эксперимент полностью закончен, т. е. апостериорно. Например, трехмерную схему можно рассматривать на произвольно выбранную глубину.

Одним из видов голограмм являются *трехмерные* или *объемные голограммы*. При записи статической трехмерной голограммы когерентное излучение, идущее от точечного источника, перед падением на объект проходит через светофоточувствительный материал. Засвечивает материал стоячая волна, возникающая в результате интерференции отраженной от объекта волны с излучением источника. Пучности стоячей волны образуют поверхности, которые после проявления можно рассматривать как систему своеобразных зеркал сложной формы.

Для восстановления изображения объекта на полученную рассмотренным способом голограмму направляется излучение точечного источника. Восстанавливающая волна, достигая какого-либо из упомянутых «зеркал», приобретает распределение фаз, совпадающее с распределением фаз объектной волны. После отражения от такого «изофазного зеркала» направление распространения восстанавливающей волны изменится на противоположное, и она по всем параметрам совпадет с объектной волной, т. е. преобразуется в нее.

Источник считывающей волны может быть и немонохроматическим и иметь сплошной спектр излучения. Структура следующих друг за другом «зеркал» сама выделяет из сплошного спектра и отражает излучение, длина волны которого совпадает с длиной волны излучения, экспонировавшего голограмму при записи.

Объемные голограммы имеют ряд преимуществ перед обычными плоскими голограммами. Наряду с высокой спектральной и угловой селективностью они обладают большой дифракционной эффективностью и низким уровнем шумов. Информацию о поляризации объектной волны позволяет сохранить запись голограммы в светочувствительной среде, которая под воздействием поляризованного излучения становится анизотропной (эффект Вейлерта).

Цифровая голография — это анализ и синтез волновых полей с помощью цифровых вычислительных устройств. Синтез голограмм на ЭВМ осуществляется в два этапа. Сначала строится математическая модель предмета и рассчитывается световая волна, создаваемая им в плоскости голограммы, затем изыскивается способ отображения результатов этих расчетов (в виде комплексных чисел), и полученная картина записывается на пленке или каком-либо носителе таким образом, чтобы с этой записи можно было восстановить исходную волну. При машинном восстановлении волнового поля производится сканирование плоскости голограммы чувствительным элементом и ввод данных в ЭВМ. Машина выполняет расчеты, необходимые для восстановления исходного предмета.

Использование ЭВМ для анализа и синтеза полей позволяет реализовать преимущества, присущие цифровой технике обработки сигналов: высокую точность и абсолютную воспроизводимость обработки, возможность реализации сложных нелинейных и логических преобразова-

ний, доступность результатов и простоту вмешательства на любой стадии обработки.

Машинный синтез голограмм дает возможность визуально наблюдать изображения математически заданных объектов и процессов, изготовлять голограммы эталонов и пространственных фильтров. Машинное восстановление голографических изображений позволяет избежать трудностей, связанных со съемом количественной информации с голографических изображений, так как в этом случае количественные характеристики изображения получаются на выходе машины.

Особенно полезно применение ЭВМ для СВЧ- и ультразвуковой голографии, для которых прямой ввод данных в машину устраняет масштабные искажения изображения, присущие оптическому восстановлению таких голограмм. Моделирование на ЭВМ всего голографического процесса позволяет довести до численного результата решение ряда сложных теоретических задач: определение разрешающей способности, влияние шумов, аббераций, дискретизации голограмм и пр. Это вызвано тем, что, во-первых, реализация голографического процесса затрудняется обычно влиянием трудно контролируемых факторов и не всегда дает точно воспроизводимые результаты, во-вторых, алгоритмическая гибкость цифровой техники значительно расширяет виды преобразований излучения по сравнению с теми, которые посильна выполнить имеющаяся оптическая система. Появляется также возможность реализации строгих аналитических зависимостей для формирования и корректировки волнового фронта.

§ 1.14. Голография в информационной технике

В области аналоговой вычислительной техники голография позволяет произвести многие математические операции над *комплексными числами*, заданными в виде пространственных распределений (матриц) амплитуд и фаз световых волн: сложения, вычитания, умножения, дифференцирования и некоторых видов интегральных операций.

Новые возможности открыла голография в организации запоминающих устройств цифровых ЭВМ, например осуществление голографического хранения информации не только в тонком слое, но и в трехмерных средах, например в щелочно-галоидных кристаллах или реоксане. Теоретическая емкость объемной записи интерференционных картин оптического диапазона достигает 10^{13} двоичных единиц в кубическом сантиметре. Таким образом, согласно теоретический расчетам, в каждом кубическом сантиметре кристалла можно разместить библиотеку из 5 млн. томов по 200 страниц каждый, если считать, что на каждой странице имеется 2500 знаков, состоящих из четырех элементов.

Логические операции в системах цифровой памяти можно осуществить при записи серии экспозиций голограмм в одном положении в таких непрерывно-изменяемых средах, как электрооптические кристаллы или непроявленная фотопленка. Селективное стирание голограмм в комплекте наложенных трехмерных голограмм, записанных в самопроявляющемся кристалле ниобата лития, было получено сдвигом фазы опорного луча на 180°. Этот процесс эквивалентен операции булевой алгебры ИСКЛЮЧАЮЩЕЕ ИЛИ. Не представляет труда в такой системе записи обращение контраста, эквивалентное логической операции ДОПОЛНЕНИЕ. На страницах цифровых данных была осуществлена также операция ИЛИ.

В булевой алгебре для образования других логических операций достаточно сочетания одной функции подобия (например, ИСКЛЮЧАЮЩЕЕ ИЛИ) и одной функции единственности (например, ИЛИ). Наличие таких операций позволяет производить сложение и вычитание n-битных слов и другие виды обобщенной обработки.

Добавление к перечисленным трем основным операциям булевой логической операции И существенно упрощает логическую обработку. Операция И становится возможной, если составители страниц последовательно накладываются друг на друга, образуя сандвич. Восстановленная страница данных будет светлой только в том адресуемом положении, для которого все отдельные составители страниц прозрачны. Приписывая светлым местам единицы, а темным нули, сложный составитель страниц осуществляет при голографической записи логическую операцию И. Операции И, ИСКЛЮЧАЮЩЕЕ ИЛИ, ДОПОЛНЕНИЕ и ИЛИ можно рассматривать как основные операции при голографической обработке, поскольку эти процессы можно произвести за один шаг.

Запись информации в интерференционной форме позволяет производить процесс поиска в блоке оптической памяти практически *одновременно* по всей памяти, как бы громоздка и многомерна она ни была. Операция поиска происходит по программе одновременного действия. Например, за одно включение (10^{-3} с) можно осуществить поиск нужной информации в памяти емкостью 10^8 бит, размещенной на плоской фотопластинке 70×70 мм (эквивалентно библиотеке из 300 книг по 200 страниц в каждой). Необходимо отметить, что высокая плотность записи и хранение информации являются достижением не столько голографии, сколько оптического диапазона, для которого характерны длины волн, соизмеримые с микрометром и определяющие размер одного бита информации (емкость блока памяти при поэлементной, дискретной форме записи и при интерференционной форме записи одинакова). Основное преимущество интерференционной памяти состоит в ее надежности, устойчивости по сравнению с поэлементной. Это объясняется тем, что волна от каждой точки диффузно отражающего или преломляющего объекта

распределяется по всей голограмме. Если такую голограмму разбить на куски, то по любому из них можно восстановить изображение всего предмета. Потеря части запоминающей среды, запыление, износ и повреждение ее не препятствуют проведению поиска по всей библиотеке и восстановлению нужного изображения (массива) по сохранившейся части интерферограммы. Столь высокая *помехоустойчивость* голографической памяти, чрезмерно избыточная в обычных условиях, может быть использована в *системах ассоциативного поиска.*

К числу самых необычных свойств интерференционной формы записи информации относится возможность получения изображения утерянной части искомого предмета. Такие изображения называют *фантомными.* Допустим, в интерференционном блоке памяти записана полная информация о каком-либо предмете и требуется путем сравнения с предметом ответить на вопрос: имеются ли в памяти сведения о предмете? Оказывается, что и в том случае, когда сравнение производится с частью предмета, можно получить не только ответ на интересующий вопрос, но и изображение целого предмета.

Допустим, что распределение напряженности поля в пучке от предъявляемого к поиску фрагмента предмета описывается функцией $\dot{E}_\phi(x) = E_\phi(x)e^{j\varphi(x)}$, в пучке остатка — $\dot{E}_0(x) = E_0(x)e^{j\rho}$ и в предметном пучке — $\dot{E} = \dot{E}_\phi + \dot{E}_0$.

Тогда в голограмме, полученной при контрастности $\nu = -2$ в результате интерференции между пучками лучей от фрагмента и остатка, амплитудное пропускание

$$T(x) \doteq I(x) \doteq \left|\dot{E}_\phi + \dot{E}_0\right|^2 = E_\phi^2 + E_0^2 + \dot{\tilde{E}}_\phi \dot{E}_0 + \dot{E}_\phi \dot{\tilde{E}}_0.$$

При освещении такой голограммы волной фрагмента $\dot{E}_\phi$ на выходе появится волна с амплитудой

$$\dot{E}_r = T\dot{E}_\phi = \dot{E}_\phi(E_\phi^2 + E_0^2) + E_\phi^2 \dot{E}_0 + \dot{E}_\phi^2 \dot{\tilde{E}}_0. \qquad (1.84)$$

При диффузном освещении фаза фрагмента $\varphi(x)$ является быстро и случайно меняющейся функцией координат и ее удвоенное значение приводит к полному искажению волнового фронта. В результате компонента волны, описываемая третьим слагаемым $\dot{E}_\phi^2 \dot{\tilde{E}}_0 = E_\phi^2 E_0 \times e^{j(2\varphi_0 - \rho)}$, полностью рассеивается во всех направлениях и создает практически равномерный фон.

Интенсивности волн E_ϕ^2 и E_0^2 можно считать постоянными при диффузном освещении. Следовательно, согласно (1.84), при диффузном освещении на выходе голограммы кроме компоненты от фрагмента (пер-

вое слагаемое) будет присутствовать и компонента остатка (второе слагаемое) — фантомное изображение.

Таким образом, с помощью интерференционной памяти удается вызвать фантом потерянной части предмета и беспрепятственно произвести ассоциативный поиск в блоке памяти. С помощью такого блока памяти можно, например, ответить на вопрос, есть ли интересующая нас страница в библиотеке, которая записана в голограмме размером 70×70 мм (300 книг). Для получения определенного ответа согласно расчетам достаточно иметь лишь 1/30 страницы, например одну строку. Страница будет распознана и локализована, а ее фантомное изображение будет хорошо различимо (отношение сигнала к шуму составит примерно 20).

Кибернетическая машина с ассоциативной памятью может решать задачу «мгновенной» выдачи инструкции по заданной ситуации без перебора всех инструкций, вложенных в машину. Для этого в голографической памяти в процессе ее изготовления следует производить запись каждой ситуации одновременно с соответствующей ей инструкцией. Если в процессе поиска нужной инструкции осветить блок памяти когерентным светом через транспарант (диапозитив) с изображением заданной ситуации, то на экране возникает фантомное изображение инструкции. Ее остается только прочесть. В машине для переводов роль ситуации могут играть фразы на иностранном языке, а инструкции — переводы этих фраз на родной язык.

Исследование трехмерных голограмм стимулируется сходством их некоторых свойств со свойствами мозга человека и гипотезой физиологов о том, что мозг хранит каждый бит информации не в одиночной пространственно-локализованной ячейке, а в виде одиночной пространственной гармоники возбуждения, заполняющей весь его объем. При этом повреждение одного или нескольких участков мозга не вызывает полного исчезновения какой-либо части записанной в ней информации. В отличие от блока памяти на пространственных ячейках, когда каждая расположенная внутри объема ячейка должна быть соединена с входным устройством специальным нервным волокном, в данном случае каждый элемент соединен только с ближайшими соседями. Плотность записи и объем такой памяти грандиозны.

Аналогия трехмерной голограммы с мозгом относится также к ассоциативной памяти и к операции, напоминающей «концентрацию внимания». Этот эффект возникает в среде, реагирующей на свет непосредственно в момент его воздействия. Восстановленная волна, проходя через объем голограммы в такой динамической среде, начинает снова записываться, что приводит к усилению записи и постепенному «разгоранию» восстановленного изображения.

Возможность моделирования некоторых функций головного мозга с помощью объемной голограммы пока не нашла практического примене-

ния из-за отсутствия светочувствительного материала, обладающего необходимыми качествами. Голограммы в щелочно-галогенидных кристаллах быстро разрушаются под действием считывающего излучения; реоксан, хотя и позволяет довести дифракционную эффективность почти до 100%, однако имеет низкую светочувствительность, равную нескольким джоулям на квадратный сантиметр.

Изложенное далеко не исчерпывает область применения голографии в информационных системах. В последующих разделах к этому вопросу мы будем неоднократно возвращаться.

§ 1.15. Голографические методы в измерительной технике

Голография находит многочисленные практические применения в измерительной технике. К ним, в частности, относятся объемные измерения *стереоскопическими методами*. Они позволяют определить форму объекта по его голографическому изображению и координатам характерных точек объекта. Этим методом были определены формы осколков при взрыве, факела центробежной форсунки, пузырьков газа в дизельном топливе, следы треков в пузырьковых камерах, а также треки осколков взрывающейся проволочки. В отличие от стереоскопических методов в обычной фотометрии голографическое изображение можно рассматривать с различных ракурсов, вплоть до получения изображения с углом обзора 360°.

Голография преодолевает ограниченность регистрации объемных изображений, связанную с небольшой глубиной поля зрения микроскопов с большим увеличением, когда можно наблюдать только небольшие участки сцены, находящейся в непосредственной близости от фокальной плоскости микроскопа. При наблюдении движущихся микрообъектов, например живых организмов, перемещение за время перефокусировки может создать искаженное представление об изучаемых объектах и их количестве. В то же время скоростное стробоскопическое голографирование изучаемой области через объектив с большой глубиной резкости позволяет рассматривать объемные изображения «застывших» микроорганизмов через окуляр микроскопа, который можно фокусировать на различные плоскости без потери информации о промежуточных стадиях процесса.

Одним из важных свойств восстановленных голографических изображений является возможность их взаимодействия с реальным объектом или другим голографическим изображением. На этом основаны голографическая интерференция, пространственная фильтрация и распознавание образов.

Объемные измерения *интерферометрическим методом* основаны на сравнении волновых фронтов реальных объектов с восстановленными

голограммами образцовых объектов, либо на сравнении волновых фронтов от одного и того же объекта, полученных в различные моменты времени. В отличие от классических интерферометров, в которых используются два точно выравненных оптических тракта (интерференция возникает только при одновременном существовании обоих сравниваемых изображений), в голографическом интерферометре используется один оптический тракт (благодаря регистрации волновых фронтов в различные моменты времени). Если в классической интерферометрии применяются, как правило, только плоские объекты, то голографическая интерферометрия позволяет изучать деформации отражающих трехмерных объектов сложной формы и объемное распределение различных физических параметров внутри преломляющих (фазовых) объектов.

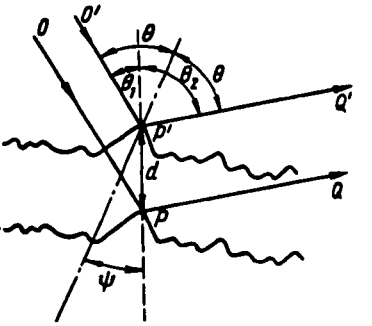


Рис. 1.10. Геометрия поверхностной деформации

Изображение, получаемое при восстановлении голограммы, должно накладываться точно на объект. При этом происходит интерференция восстановленного волнового фронта с волновым фронтом, возникающим непосредственно на объекте. При точном их совмещении объект будет покрывать одна темная интерференционная полоса, так как фаза голографически образованного волнового фронта является обратной по отношению к волне реального объекта вследствие отрицательных характеристик (негативных свойств) фотографического процесса. Слабое смещение поверхности объекта образует серию интерференционных полос, число n которых тем больше, чем больше смещение d. Направление движения интерференционных полос определяет, произошло смещение в сторону наблюдателя или от него.

На рис 1.10 показаны положения шероховатой поверхности до и после смещения в направлении PP' на расстояние d. Направление освещения перед отклонением представлено отрезком OP, после отклонения — $O'P'$, а соответствующие направления наблюдения — отрезками PQ и $P'Q'$. Разность оптического пути

$$\Delta = d(\cos\theta_1 + \cos\theta_2) = 2d\cos\frac{\theta_1+\theta_2}{2}\cos\frac{\theta_1-\theta_2}{2} = 2d\cos\theta\cos\psi = n\lambda.$$

Смещение

$$d = n\lambda(2\cos\theta\cos\psi)^{-1} \qquad (1.85)$$

подсчитывают по числу n интерференционных полос. Число полос должно быть достаточно малым, чтобы обеспечить их различимость.

В голографической интерферометрии используются два основных метода: метод двойной экспозиции и метод наблюдения в реальном масштабе времени. В первом из них на одной фотопластинке при неизменном опорном луче регистрируются две голограммы объекта: одна — до приложения возмущающих сил или до начала процесса, а вторая — через некоторый интервал времени. Необходимую информацию получают из интерпретации интерференционных полос, образующихся при восстановлении двух волновых фронтов одновременно. Волновые фронты от участков объектов, в которых не произошли изменения за время между экспозициями (например, в неподвижных стенках сосуда), не создадут интерференционных полос в восстановленном изображении объекта. Поэтому результаты голографической интерферометрии в отличие от классической не зависят от качества прозрачных стенок, за которыми происходит изучаемый процесс. Основные преимущества метода заключаются в том, что: 1) пластинка остается в фиксированном положении в течение всего испытания; 2) исключается влияние усадки эмульсии; 3) при восстановлении неподвижные интерференционные полосы (их иногда называют «замороженными») легче поддаются измерению. Метод двойной экспозиции обычно применяют при изучении быстро протекающих процессов.

Метод наблюдения в реальном масштабе времени используют в большинстве случаев при исследовании сравнительно медленно протекающих процессов либо при контроле с помощью одной голограммы серии стационарных объектов. При использовании этого метода, называемого иногда «живым», производят непосредственное, обычно визуальное наблюдение за интерференционными полосами и их движением на объекте, совмещенным с его восстановленным голографическим изображением. Этот метод дает возможность управления процессом деформации.

Проведение неразрушающего контроля и испытаний во многих случаях оказывается весьма эффективным, даже при качественном применении голографической интерферометрии, например, при рассмотрении деформаций по всей поверхности объекта в результате приложения к нему небольших усилий. Подобные испытания были проведены на тонких мембранах, используемых в датчиках давления. При нагружении мембраны в центре незначительным давлением интерференционные полосы образуют симметричную кольцевую картину. В дефектной мембране локальная «слабая» область проявляется в нарушении симметрии интерференционных колец. Поскольку каждое кольцо соответствует смещению соседнего на половину длины волны (около 0,3 мкм), проведенный контроль оказался чрезвычайно чувствительным.

Возможности голографии были показаны применительно к оптическому диапазону длин волн. Однако они реализуются и в других диапазонах электромагнитных и акустических волн.

При переходе к колебаниям СВЧ необходимо учитывать, что увеличение длины волны вызывает значительный рост всех размеров излучателей и объектов и расстояний между ними. Например, для получения приемлемого разрешения на длине волны 3 см размер квадратной голограммы должен быть равен примерно 2 м, а расстояние от источника излучения до объекта — 10 м. Отсутствие индикаторов СВЧ, близких по свойствам к фотопленкам, приводит к использованию сканирующих СВЧ-приемников, сигнал с которых поступает либо в запоминающее устройство ЭВМ, либо на модулятор электронно-лучевой трубки, либо на газосветную лампочку. Сформированную с их помощью голограмму после обработки и уменьшения можно использовать для восстановления изображения обычными методами лазерной голографии.

От оптической голографии СВЧ-голографию отличает ряд существенных преимуществ, связанных с тем, что: а) длина когерентности генераторов СВЧ составляет сотни километров, поэтому возможны неразрушающий контроль и обнаружение сильно удаленных крупномасштабных предметов; б) при достаточно малом числе отсчетов интерференционной картины для восстановления приемлемого по качеству изображения перспективно применение хорошо отработанных методов радиотехнической обработки полезной информации, позволяющих быстро и с высокой достоверностью проводить сравнение аналогичных изделий или их сортировку (например, на базе теории распознавания образов); в) передача опорного сигнала непосредственно на детектор без излучения его в свободное пространство избавляет от сферичности опорного сигнала, чем существенно облегчается получение трехмерного изображения; г) высокая проникающая способность в такие диэлектрические материалы, как бетон и пластмассы, позволяет в системах со сканированием или многоэлементным антенным полем в «ближней зоне» строить безлинзовые системы формирования изображения, при котором сам материал является прозрачным, а дефекты — непрозрачными, обеспечивая большое отношение сигнала к фону (таким образом, становится возможным измерение профиля, глубины залегания и физических параметров твердого грунта Луны под слоем сухого песка, а также обнаружение скрытого оружия); д) измерение объектов возможно в темноте и за непрозрачными перегородками.

Для непрозрачных объектов решать задачи объемной интроскопии применительно к контролю различных металлических, керамических и других изделий можно также методами *акустической голографии*. В сейсмических и океанографических исследованиях и при изучении фарватеров рек акустическая голография открывает новые возможности об-

наружения, определения местонахождения и опознания различных объектов, слоев грунта и т. п. В биологии и медицине акустическая голография позволяет вести исследования различных биологических объектов и, в частности, решать задачи диагностики различных опухолевых заболеваний. При исследовании внутренних органов человека акустическая голография обеспечивает четкое восстановление изображения изучаемого объекта в частности, мягких тканей, кровеносных сосудов, внутренних органов, что недоступно рентгеновским методам. В большинстве случаев конечной формой голограмм является оптический диапозитив, который способен образовывать оптическую копию исходного акустического фронта. В качестве среды для последующей оптической регистрации ультразвукового поля часто используется пульсация поверхности воды. Если к свободной поверхности жидкости прикладывается некоторое распределенное давление, то поверхность приобретает рельеф волнообразной формы с гребнями и впадинами, причем высота гребней прямо пропорциональна местному избыточному давлению. Если распределенное давление определяется ультразвуковой волной, то абсолютное значение давления, а следовательно, и высота отдельного гребня пропорциональны квадрату амплитуды волны. Рельеф поверхности образует обычную голограмму, но в другом масштабе. Ее можно фотографировать или, освещая когерентным светом, наблюдать непосредственно восстановленное изображение.

§ 1.16. Спекл-интерферометрия

Поверхность, освещенная когерентным светом, например, лазером для наблюдателя, кажется имеющей зернистую структуру. Это явление называется *спекл-эффектом*. Спекл-эффект наблюдается только в том случае, когда изменения рельефа поверхности имеют порядок длины волны падающего света (оптически грубая поверхность).

Каждая точка отражающей поверхности в соответствии с принципом Гюйгенса-Френеля действует как источник сферических волн. Поскольку высота рельефа поверхности меняется случайным образом на λ или больше, то амплитуда напряженности поля в наблюдаемой точке образуется суммированием векторов со случайными фазами. Она меняется от нуля до максимального значения в зависимости от значения и фазы отдельных компонент. При изменении положения наблюдаемой точки интенсивности рассеянного ею света принимают различные случайные значения. Эти случайные изменения интенсивности и обуславливают спекл-эффект.

Интенсивности рассеянного света в двух точках мало отличаются, если точки расположены близко друг от друга. Различие между ними растет по

мере увеличения расстояния между точками до тех пор, пока они не становятся полностью независимыми от этого расстояния. Это расстояние может быть названо *размером спекла.* Оценить его значение можно, рассматривая площадку $L \times L$, освещаемую однородным пучком света. Допустим, точки P_1 и P_2 ограничивают освещаемую поверхность вдоль оси x (рис. 1.11).

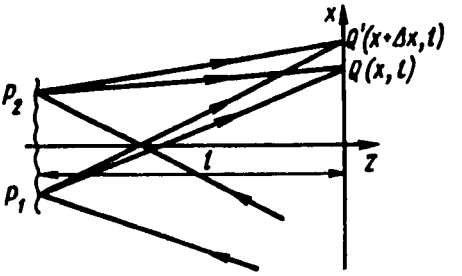


Рис. 1.11. Рассеяние светового потока оптически грубой поверхностью

Интенсивность света в точке $Q(x, l)$ в плоскости наблюдения рассматривают как результат сложения интенсивности световых пучков, рассеянных всеми точками поверхности предмета. Разность хода волн, проходящих пути P_1Q и P_2Q, $s = P_1Q - P_2Q \approx xL/z + L^2/2z$, где z — расстояние между плоскостью наблюдения и предметной плоскостью.

Разность хода к другой точке наблюдения $Q'(x + \Delta x, l)$ $s' \approx P_1Q' - P_2Q' \approx xL/z + L^2/2z + \Delta xL/z$, т. е. разность изменилась на $\Delta s = s' - s = \Delta xL/z$. Если $\Delta s \ll \lambda$, то разность фаз всех компонент будет одной и той же. Если $\Delta s \approx \lambda$, то фазы будут настолько различными, что интенсивности света в точках Q и Q' не коррелируют. Следовательно, средним размером спекла можно считать

$$\Delta x = \lambda z / L. \qquad (1.86)$$

Из формулы (1.86) следует, что размер спекла зависит от положения плоскости наблюдения и не зависит от системы, которая формирует изображение и используется для наблюдения. Такие спеклы являются *спеклами в пространстве предмета.* Они не совпадают со *спеклами в пространстве изображения.*

Вероятностное распределение спеклов в пространстве изображения то же самое, что и для спеклов в пространстве предмета. Однако пространственное распределение в этом случае ограничено дифракционным пределом системы, формирующей изображение. Если, например, изображение формируется сферической линзой диаметром D, находящейся на расстоянии f от плоскости изображения, то первый минимум дифракционного распределения интенсивности излучения с максимумом в точке Q будет находиться на расстоянии радиуса Эйри $r_0 = 1{,}22\lambda f / D$.

Если точка P_2 расположена таким образом, что положение первого минимума ее дифракционной картины совпадает с точкой Q, то она уже

не влияет на амплитуду напряженности поля светового луча в точке Q. Составляющими этой амплитуды, обусловленные рассеянием света от точек, отстоящих от P_1 дальше P_2, можно пренебречь, так как второй и последующие максимумы дифракционного распределения много меньше центрального. Следовательно, интенсивность света в точке Q определяется световыми лучами, приходящими от площадки предметной поверхности с центром в точке P_1 и диаметром, в два раза превышающем расстояние между точками P_1 и P_2. За *размер* спекла в пространстве изображения принимается диаметр именно этой поверхности:

$$d = 2{,}44\lambda f/D. \qquad (1.87)$$

Таким образом, помещая между предметом диафрагму или линзу с апертурой меньше размера зрачка глаза, можно увеличить кажущийся размер спекла по сравнению с непосредственным визуальным наблюдением.

Спекл-интерферометрия позволяет измерять смещения и исследовать форму в случае оптически грубой поверхности с точностью порядка длины световой волны. Для этого используются два основных метода: корреляционная спекл-интерферометрия и спекл-фотография. В обоих методах поверхность наблюдается в начальном и смещенном положениях и изучается картина интерференционных полос. Чувствительность этих полос к значению и направлению смещения поверхности предмета может меняться в существенно более широких пределах, нежели в голографической интерферометрии. Кроме того, при этом требуется меньшая разрешающая способность регистрирующей среды. Эти качества делают метод спекл-интерферометрии более гибким, чем голографическая интерферометрия.

В методе *спекл-фотографии* с помощью линзы осуществляется запись на одной фотопластинке двух рассеянных предметом световых волн. В плоскости изображения возникают две смещенные спекл-структуры: одна из них образована волной, рассеянной предметом в начальном положении, другая — волной, рассеянной предметом после его смещения. Сдвиг этих спекл-структур зависит от смещения предмета. Измерив относительный сдвиг двух записанных спекл-структур, можно определить смещение предмета. Для получения интерференционных полос здесь не требуется опорный пучок. Для определения смещения спекл-структуры используется фурье-преобразование функции пропускания дважды экспонированной фотопластинки. Для этого достаточно, например, осветить фотопластинку плоской волной и наблюдать дифрагированный ею световой поток в фокальной плоскости линзы.

Динамические смещения можно определять спекл-фотографией с усреднением по времени, стробоскопической и двухимпульсной спекл-фотографией подобно аналогичным методам в голографической интерферометрии.

Корреляционная спекл-интерферометрия основана на анализе интерференционной картины, возникающей при одновременном наложении волн, рассеянных от поверхности предмета в начальном и смещенных положениях. Используемый для этого интерферометр расщепляет входящий луч на два одинаковой интенсивности. Один из них освещает неподвижную поверхность, а другой — такую же поверхность, но смещенную. Волны, отраженные интерферометром, накладываются друг на друга и через линзу попадают на плоскость изображения. Схема этого устройства отличается от интерферометра Майкельсона только тем, что зеркала заменены на две шероховатые поверхности.

Интерференционную картину корреляции спеклов получают с помощью фото- или телекамеры. Требования к разрешающей способности регистрирующей среды в данном методе более низкие, чем в голографической интерферометрии (3000...5000 линий/мм). Это связано с тем, что минимальные размеры спеклов лежат в пределах 5...100 мкм. Здесь требуется обеспечить лишь разрешение спекл-структуры (50 линий/мм), а не тонкой структуры полос, возникающих на голограмме в результате интерференции предметной и опорной волн. Именно поэтому для получения картины спекл-корреляционных полос стало возможным использование стандартной телекамеры. Этот метод называют *электронной спекл-интерферометрией.*

Видеозапись позволяет выводить динамическую картину корреляционных полос прямо на экран дисплея, минуя какую-либо фотозапись, точную установку фотопластинки и т. д. Относительная простота этого метода дает возможность при определении формы поверхности и анализе деформаций решать более сложные задачи. Применение широкополосного фильтра и запоминающего устройства позволяет использовать сложение и вычитание видеосигналов с последующим выводом результатов корреляции интенсивностей на экран телемонитора.

Глава 2
Стоячие электромагнитные волны и объемные резонаторы

При нормальном падении плоской волны на металлическую поверхность отраженная волна обладает амплитудой, равной амплитуде падающей волны, и противоположной фазой ($\rho_{11} = -1$). В результате интерференции падающей и отраженных волн в диэлектрике около металлической поверхности возникает стоячая волна. Если волновое число диэлектрика $k = 2\pi/\lambda$ и падающая волна распространяется вдоль оси z, то в стоячей волне напряженности

$$\dot{E} = \dot{E}^+ + \dot{E}^- = E_0(e^{-jkz} - e^{jkz}) = -j2E_0 \sin kz = \dot{E}_m \sin\frac{2\pi z}{\lambda};$$

$$\dot{H} = \dot{H}^+ - \dot{H}^- = \frac{E_0}{z_c}(e^{-jkz} + e^{jkz}) = 2\frac{E_0}{z_c}\cos kz = \dot{H}_m \cos\frac{2\pi z}{\lambda}.$$

Если в бегущей волне отсутствовал сдвиг фаз между электрическим и магнитным векторами, то в стоячей волне эти составляющие поля электромагнитной волны оказываются сдвинутыми по фазе на 90°. Таким образом, в момент прохождения электрического поля через максимальное значение магнитное поле отсутствует, и наоборот.

В бегущей волне направление вектора Пойнтинга в любой момент времени совпадает с направлением ее распространения (кроме момента прохождения фазы через нуль, когда мгновенное значение вектора Пойнтинга равно нулю). В бегущей волне энергия электрического поля в любой момент времени равна энергии магнитного поля: $\varepsilon E^2/2 = \mu H^2/2$, а среднее значение вектора Пойнтинга $EH/2 = E^2/(2z_c) = v(\varepsilon E^2/2) = v(\mu H^2/2)$.

В стоячей волне среднее значение вектора Пойнтинга в любой точке равно нулю (ввиду сдвига фаз между $\dot{E}$ и $\dot{H}$ на 90°). Энергия поля стоячей волны изменяется во времени и пространстве. Происходит процесс преобразования энергии электрического поля, сосредоточенной вблизи пучности электрического поля, в энергию магнитного поля, сосредоточенную вблизи своей пучности. Этот процесс носит колебательный характер как во времени, так и в пространстве. Однако движение энергии ограничено расстоянием $\lambda/4$ между пучностями электрического и магнитного полей. В стоячей волне максимальные плотности энергии электрического и магнитного полей также одинаковы:

$$\frac{\varepsilon E_m^2}{2} = 2\varepsilon E_0^2 = 2\mu H_0^2 = \frac{\mu H_m^2}{2}.$$

В отличие от бегущей волны они локализованы в различных точках пространства ($\Delta z = \lambda/4$) и наступают в различные моменты времени ($\Delta t = T/4$).

Явления в стоячей электромагнитной волне подобны в энергетическом отношении явлениям в простейшем колебательном LC-контуре при отсутствии потерь. При резонансной частоте, когда период $T = 2\pi\sqrt{LC}$, в контуре происходят колебания энергии, в процессе которых энергия сосредоточивается то в магнитном поле катушки, то в электрическом поле конденсатора. В момент максимума энергия электрического поля конденсатора равна максимуму энергии магнитного поля катушки: $CU_m^2/2 = LI_m^2/2$, хотя, как и в стоячей волне, в момент максимума напряженности поля в конденсаторе поле в катушке отсутствует. Подобно тому как в стоячей волне в среде с волновым сопротивлением $z_c = \sqrt{\mu/\varepsilon}$ $\dot{E}_m = -jz_c\dot{H}_m$, в колебательном контуре напряжение на конденсаторе $\dot{U}_m = -j\dot{I}_m/(\omega C) = -jz\dot{I}_m$, где $z_c = \sqrt{L/C}$ — волновое сопротивление контура.

Следовательно, колебательный контур можно рассматривать как своеобразную схему замещения той части поля стоячей волны, которая сосредоточена между пучностями электрического и магнитного полей.

§ 2.1. Структура поля в прямоугольном резонаторе

Общность свойств колебательного контура и поля стоячей волны лежит в основе использования *объемных резонаторов*, т. е. полостей, ограниченных со всех сторон металлическими стенками. Если в ограниченном металлическими стенками объеме возбудить электромагнитное поле, то образуются стоячие волны, расположение и частота которых зависят от формы и размеров полости.

Простейшая структура поля образуется в прямоугольном резонаторе (рис. 2.1) без потерь в результате многократного отражения плоской волны (1.1), причем компоненты волнового вектора (1.2) связаны соотношением

$$k^2 = k_x^2 + k_y^2 + k_z^2 = (2\pi/\lambda)^2. \qquad (2.1)$$

Последнее уравнение получено ранее простым разложением волнового вектора плоской волны на его ортогональные компоненты. Однако можно убедиться, что при любой форме волны компоненты волнового

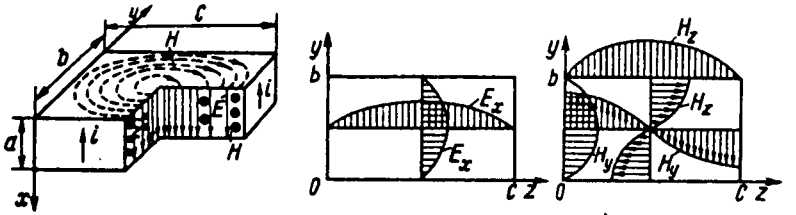


Рис. 2.1. Схема распределения поля основного типа в прямоугольном резонаторе

вектора связаны таким же соотношением, вытекающим из возможности решения любого однородного волнового уравнения в прямоугольной системе координат методом разделения переменных. Знаки «±» в уравнении (1.1) указывают на возможность образования после многократных отражений волн с разнообразными направлениями. Различные сочетания знаков перед одинаковыми компонентами позволяют получить восемь различных волновых векторов, одинаковых по значению и одинаково удовлетворяющих уравнению (2.1). Для положительных знаков, например, соответствующее выражение проекции комплексного вектора напряженности электрического поля на ось x будет иметь вид

$$\dot{E}_{xl} = \dot{E}_{0xl}(\cos k_x x + j\sin k_x x)(\cos k_y y + j\sin k_y y)(\cos k_z z + j\sin k_z z).$$

Аналогичный вид имеют проекции на оси y и z.

Выбор из всех возможных решений одного, дающего решение конкретной задачи, проводится с помощью граничных условий и теоремы Гаусса, согласно которой в однородной изотропной среде без объемных зарядов ($\rho = 0$)

$$\operatorname{div}\mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0. \qquad (2.2)$$

В прямоугольном резонаторе (см. рис. 2.1) с идеально проводящими стенками ($\sigma = \infty$) граничные условия определяются условиями отсутствия тангенциальных составляющих вектора напряженности электрического поля на стенках:

$$\begin{aligned} &E_y = E_z = 0 \text{ при } x = 0 \text{ и } x = a; \\ &E_x = E_z = 0 \text{ при } y = 0 \text{ и } y = b; \\ &E_x = E_y = 0 \text{ при } z = 0 \text{ и } z = c. \end{aligned} \qquad (2.3)$$

Решение волнового уравнения с учетом (2.2) и (2.3) приводит к следующим значениям составляющих волнового вектора:

$$k_x = \pi m/a, \quad k_y = \pi n/b, \quad k_z = \pi p/c \qquad (2.4)$$

и составляющих напряженности электрического поля

$$\begin{aligned} \dot{E}_x &= \dot{E}_1 \cos\frac{m\pi x}{a}\sin\frac{n\pi y}{b}\sin\frac{p\pi z}{c}; \\ \dot{E}_y &= \dot{E}_2 \sin\frac{m\pi x}{a}\cos\frac{n\pi y}{b}\sin\frac{p\pi z}{c}; \\ \dot{E}_z &= \dot{E}_3 \sin\frac{m\pi x}{a}\sin\frac{n\pi y}{b}\cos\frac{p\pi z}{c}, \end{aligned} \qquad (2.5)$$

где m, n, p — произвольные целые числа, называемые *характеристическими числами*. Как следует из (2.2), решение (2.5) учитывает только вих-

ревое поле. Если бы потребовалось решение общей краевой задачи, то в общее решение необходимо было бы включить кроме компонент (2.5) вихревого поля компоненты градиентного поля, для которого $\operatorname{div}\mathbf{E} \neq 0$ и $\mathbf{E} = -\operatorname{grad}\varphi$.

Полученное решение представляет собой уравнение стоячих волн, причем характеристические числа оказываются равными числу полуволн, укладывающихся внутри резонатора соответственно вдоль осей x, y, z.

Длину волны собственных колебаний резонатора найдем из уравнений (2.1) и (2.4):

$$\lambda_0 = \frac{v}{f} = \frac{2\pi}{k} = \frac{2}{\sqrt{(m/a)^2 + (n/b)^2 + (p/c)^2}}. \qquad (2.6)$$

Таким образом, характеристические числа определяют структуру поля в резонаторе, его резонансную частоту, направления и фазовые скорости волн, образующих стоячую волну внутри полости. Направляющие косинусы $\cos\alpha = k_x/k = m\lambda_0/(2a)$, $\cos\beta = n\lambda_0/(2b)$, $\cos\gamma = p\lambda_0/(2c)$.

Фазовые скорости в направлении соответствующих координатных осей

$$\begin{aligned} v_{\phi x} &= \pm\omega_0/k_x = \pm 2av/(m\lambda_0) = \pm v/\cos\alpha; \\ v_{\phi y} &= \pm\omega_0/k_y = \pm 2bv/(n\lambda_0) = \pm v/\cos\beta; \\ v_{\phi z} &= \pm\omega_0/k_z = \pm 2cv/(p\lambda_0) = \pm v/\cos\gamma, \end{aligned} \qquad (2.8)$$

причем

$$v = 1/\sqrt{\varepsilon\mu}\,. \qquad (2.9)$$

Разным характеристическим числам в резонаторе соответствуют колебания с разными частотами. Однако два характеристических числа не могут одновременно обратиться в нуль, так как при этом в уравнениях системы (2.5) все три составляющие обращаются в нуль в любой момент времени, что свидетельствует об отсутствии колебаний и поля. Простейший тип колебаний (H_{011}) образуется, когда одно из характеристических чисел, например m, равно нулю, а другие $n = p = 1$. Тогда

$$\dot{E}_x = \dot{E}_1 \sin(\pi y/b)\sin(\pi z/c), \qquad E_y = E_z = 0. \qquad (2.10)$$

Из второго уравнения Максвелла $\operatorname{rot}\dot{\mathbf{E}} = -j\omega\mu\dot{\mathbf{H}}$ могут быть получены компоненты напряженности магнитного поля:

$$\begin{aligned} \dot{H}_x &= 0; \\ \dot{H}_y &= \frac{j}{\omega_0\mu}\frac{\pi}{c}\dot{E}_1 \sin\frac{\pi y}{b}\cos\frac{\pi z}{c}; \\ \dot{H}_z &= \frac{1}{j\omega_0\mu}\frac{\pi}{b}\dot{E}_1 \cos\frac{\pi y}{b}\sin\frac{\pi z}{c}. \end{aligned} \qquad (2.11)$$

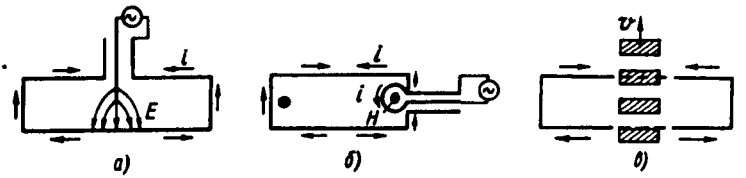


Рис. 2.2. Возбуждение резонатора штырем (*а*), петлей (*б*) и модулированным электронным потоком (*в*)

Распределение поля для данного случая изображено на рис. 2.1.

Основной частотой резонатора называют наименьшую из возможных частот. Она может быть получена, когда напряженность электрического поля направлена вдоль самой короткой стороны резонатора: $c \geqslant b > a$ при $m = 0$.

Другие сочетания характеристических чисел определяют так называемые *высшие типы волн*. Их возникновение в резонаторе с определенными размерами зависит от частоты генератора и способа возбуждения поля. Для этого часто применяют либо электрический вибратор в виде стержня или штыря (рис. 2.2, *а*), либо магнитный вибратор в виде витка с током, называемого *петлей связи* (рис. 2.2, *б*), либо модулированный электронный поток (рис. 2.2, *в*). Размещение возбуждающего штыря в направлении электрического поля в точке его максимума для данного типа стоячей волны способствует возбуждению волны именно этого типа. Для возбуждения этого же типа колебаний витком связи ось витка должна располагаться в направлении силовых линий магнитного поля в его пучности.

§ 2.2. Добротность объемного резонатора

Максимальная энергия поля прямоугольного резонатора при основном типе колебаний

$$W_{max} = \int_V \frac{\varepsilon E^2}{2} dV = \frac{1}{2} \int_0^a \int_0^b \int_0^c \varepsilon E_1^2 \sin^2 \frac{\pi y}{b} \sin^2 \frac{\pi z}{c} dx\, dy\, dz = \frac{\varepsilon E_1^2}{8} abc. \quad (2.12)$$

Реактивную мощность резонатора определяют по формуле

$$Q = \omega_0 W_{max} \quad (2.13)$$

подобно тому, как находят реактивную мощность резонансного LC-контура

$$Q = U^2 / x_c = \omega_0 C U_m^2 / 2 = \omega_0 W_{max}.$$

В предыдущих расчетах предполагалось, что потери отсутствуют не только в диэлектрике, заполняющем резонатор, но и в стенках. Это допущение позволяет с достаточной точностью рассчитать поле внутри полости. Однако в других случаях, например при расчете эквивалентных схем реального резонатора, необходимо учитывать все потери. Потери мощности, приходящиеся на единицу поверхность стенки резонатора, нетрудно подсчитать по формуле (1.60): $P_0 = H_t^2/(2\sigma\Delta) = \omega_0\mu\Delta H_t^2/4$.

В прямоугольном резонаторе для волн основного типа потери мощности:

в стенке при $z = 0$

$$P_{xy} = \int_0^a\int_0^b \frac{H_y^2}{2\sigma\Delta}\,dx\,dy = \frac{\omega_0\mu\Delta}{4}\left(\frac{\pi E_1/c}{\omega\mu}\right)^2\int_0^a\int_0^b \sin^2\frac{\pi x}{a}\,dx\,dy = \frac{\pi^2\Delta E_1^2}{8\omega_0\mu}\frac{ab}{c^2},$$

в стенке при $y = 0$

$$P_{xz} = \int_0^a\int_0^c \frac{H_z^2}{2\sigma\Delta}\,dx\,dz = \frac{\pi^2\Delta E_1^2}{8\omega_0\mu}\frac{ac}{b^2},$$

в стенке при $x = 0$

$$P_{yz} = \int_0^b\int_0^c \frac{H_y^2 + H_z^2}{2\sigma\Delta}\,dy\,dz = \frac{\pi^2\Delta E_1^2}{16\omega_0\mu}\left(\frac{b}{c} + \frac{c}{b}\right).$$

Потери мощности во всех стенках

$$P = 2(P_{xy} + P_{xz} + P_{yz}) = \frac{\pi^2\Delta E_1^2}{8\omega_0\mu}\left(\frac{b}{c} + \frac{c}{b} + 2\frac{ab}{c^2} + 2\frac{ac}{b^2}\right).$$

Исходя из определения добротности колебательного контура в виде отношения его реактивной мощности к активной при резонансе для прямоугольного резонатора добротность ненагруженного резонатора

$$Q_0 = \frac{\omega_0 W_{max}}{P} = \frac{\omega_0^2\varepsilon\mu}{\pi^2\Delta}\frac{abc}{b/c + c/b + 2ab/c^2 + 2ac/b^2}. \qquad (2.14)$$

Если учесть, что глубина проникновения $\Delta = \sqrt{2/(\omega_0\mu\sigma)}$, проводимость стенки $\sigma = 2/(\omega_0\mu)\Delta^2)$, относительная магнитная проницаемость $\mu_r = 1$ и размеры резонатора составляют обычно порядка длины волны, т. е. в первом приближении знаменатель (2.14) имеет величину порядка 6, то нетрудно с помощью уравнения (2.14) убедиться в том общем правиле, что собственная добротность резонаторов в первом приближении про-

порциональна отношению длины волны к глубине проникновения: $Q_0 \approx \lambda/\Delta$. В дециметровом диапазоне собственная добротность полых резонаторов может доходить до 10^5, что недостижимо для контуров с сосредоточенными параметрами при обычных температурах.

§ 2.3. Эквивалентная схема резонатора

Основные свойства резонатора как колебательной системы могут быть описаны двумя эквивалентными схемами, отличающимися способом включения эквивалентного активного сопротивления R или проводимости G (рис.2.3, *а*, *б*), которые учитывают потери энергии в резонаторе. При рассмотрении электронных приборов с резонаторами предпочитают вариант, представленный на рис. 2.3, *б*, позволяющий более просто суммировать эквивалентные проводимости резонатора с электропроводностью электронного потока. Принципиальное различие между резонатором и его эквивалентной схемой состоит в том, что резонатор представляет собой объемную систему с распределенными параметрами и размерами, соизмеримыми с длиной волны, а схема содержит элементы с сосредоточенными параметрами. Кроме тех немногих типов резонаторов, у которых имеется явно выраженная сосредоточенная емкость или индуктивность, нет смысла отдельно рассматривать в схеме эквивалентную емкость C и индуктивность L. Их заменяют обычно реактивной проводимостью jB (рис.2.3, *в*).

Реактивная проводимость при частотах, незначительно отличающихся от резонансной,

$$B = \omega C - \frac{1}{\omega L} = \omega C\left(1 - \frac{1}{\omega^2 LC}\right) = \frac{C(\omega - \omega_0)(\omega + \omega_0)}{\omega} \approx 2C(\omega - \omega_0). \quad (2.15)$$

Условность понятия эквивалентной активной проводимости полых резонаторов связана с неоднозначностью понятия «напряжение»:

$$U = \int_a^b \mathbf{E}\, d\mathbf{l}.$$

Поэтому активная проводимость в отличие от добротности в полом резонаторе не является инвариантом, а зависит от того, к каким точкам a и b в резонаторе она отнесена, т. е. зависит от точек отсчета.

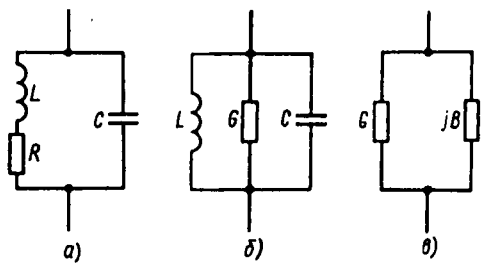


Рис. 2.3. Эквивалентные схемы резонатора

Иногда в качестве таких точек выбирают крайние точки ввода или вывода энергии. В электронных приборах такими точками считают те, в которых резонатор пронизывается электронным потоком. В этом случае активная проводимость имеет порядок $10^{-4}..:10^{-5}$ См.

В общем случае

$$G = \frac{P}{U^2} = \frac{\int\limits_S \frac{H_t^2}{2\sigma\Delta}\,\mathrm{d}S}{\left(\int\limits_a^b \mathbf{E}\,\mathrm{d}\mathbf{l}\right)^2} = \sqrt{\frac{\omega\mu}{2\sigma}}\,\frac{\int\limits_S H_t^2\,\mathrm{d}S}{\left(\int\limits_a^b \mathbf{E}\,\mathrm{d}\mathbf{l}\right)^2} \sim f^{-3/2}. \qquad (2.16)$$

В узкой полосе частот (порядка 1% от резонансной) активная проводимость практический не зависит от частоты.

Уравнения (2.14) и (2.16) редко используются даже в тех немногих случаях, когда возможно несложное аналитическое решение. Значительная ошибка получается за счет неучета потерь из-за шероховатости стенок, в контактах на стыках частей резонатора и в элементах механической настройки. Эти дополнительные потери в стенках не поддаются точному расчету. Поэтому добротность и параметры эквивалентных схем обычно находят экспериментальным путем. Из рассмотрения параметров эквивалентного колебательного контура (см. рис. 2.3, *б*, *в*) следует, что вблизи резонансной частоты: а) реактивная проводимость (2.15) линейно зависит от частоты; б) реактивная проводимость (2.15), проходя через нуль, меняет знак; в) активная проводимость практически постоянна.

Очевидно, что эквивалентные схемы, изображенные на рис.2.3, *б*, *в* можно применять только в том случае, когда полученные экспериментальным путем параметры схемы замещения данного резонатора удовлетворяют перечисленным трем условиям. Эти условия можно считать критериями аппроксимации полого резонатора параллельным контуром (см. рис. 2.3, *в*) с сосредоточенными постоянными.

§ 2.4. КПД резонатора

В том случае, когда часть энергии резонатора передается во внешнюю цепь или линию передачи, добротность резонатора уменьшается, поскольку рассеиваемая мощность $P = P_{\text{рез}} + P_{\text{н}} > P_{\text{рез}}$, где $P_{\text{рез}}$ — мощность потерь в резонаторе; $P_{\text{н}}$ — мощность, потребляемая нагрузкой. В этом случае добротность резонатора называют *нагруженной*: $Q_{\text{н}} = \omega_0 W_{\max}/P = \omega_0 W_{\max}/(P_{\text{рез}} + P_{\text{н}})$, откуда

$$1/Q_{\text{н}} = 1/Q_0 + 1/Q_{\text{вн}}, \qquad (2.17)$$

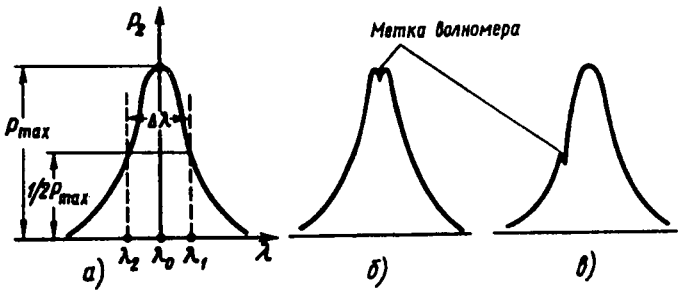


Рис. 2.4. Резонансная кривая резонатора и осциллограммы, полученные при измерении частоты и добротности

где $Q_0 = \omega_0 W_{max} / P_{рез} = \omega_0 W_{max} / GU^2$ — собственная, или ненагруженная, добротность; $Q_{вн} = \omega_0 W_{max} / P_н$ — внешняя добротность полого резонатора.

Резонатор, присоединенный к нагрузке и возбуждаемый изнутри, например с помощью электронного потока, можно рассматривать как активный двухполюсник. С учетом уравнения (2.17) коэффициент полезного действия резонатора в таком режиме

$$\eta = \frac{P_н}{P_{рез} + P_н} = \frac{Q_н}{Q_{вн}} = 1 - \frac{Q_н}{Q_0}.$$

У резонаторов, применяемых в генераторах сантиметровых волн, нагруженная добротность имеет порядок 200, а ненагруженная — 1000...2000. В этих случаях КПД резонатора составляет 80...90%.

Широкое использование понятия добротности связано со сравнительно простым ее измерением. Один из методов измерения добротности основан на анализе кривой зависимости проходящей мощности P_2 от частоты или длины волны генератора λ (рис. 2.4, *а*). Из кривой, которая может быть получена на экране осциллографа, находят резонансную длину волны λ_0 (или частоту f_0) и длины волн $\lambda_1(f_1)$ и $\lambda_2(f_2)$ (или частоты), соответствующие прохождению через резонатор половины мощности по сравнению с ее максимальным значением при резонансе. После этого подсчитывают $Q = \lambda_0 / \Delta\lambda \approx f_0 / \Delta f$, где $\Delta\lambda = \lambda_1 - \lambda_2$; $\Delta f = f_2 - f_1$.

Длины волн и частоты определяют с помощью метки, образующейся на осциллограмме резонансной кривой от параллельно включенного в тракт эталонного волномера. Перестройка волномера с помощью регулировочного винта с делениями перемещает метку по резонансной кривой. Совмещая метку с максимумом кривой (рис. 2.4, *б*) можно найти резонансную длину волны или частоту. Располагая метку в точках половин-

§ 2.6. Перестройка частоты деформацией стенок резонатора

Если возмущение вызвано только незначительным локальным изменением объема δV за счет деформации стенок резонатора, то связанное с этим изменение частоты $\delta\omega$ может быть найдено, если учесть, что в этом случае $\delta\varepsilon = \delta\mu = 0$. Тогда в отсутствие потерь в резонаторе и стенках из (2.27)

$$\frac{\delta\omega}{\omega} = \frac{\int\limits_{\delta V}(\varepsilon E^2 - \mu H^2)\,dV}{\int\limits_{V}(\varepsilon E^2 + \mu H^2)\,dV} = \frac{w_э - w_м}{2W_{max}}\delta V, \qquad (2.28)$$

где $w_э$, $w_м$ — плотности энергий электрического и магнитного полей вблизи деформируемого участка поверхности стенок; $W_{max} = \int\limits_V \frac{\varepsilon E^2}{2}\,dV$ — энергия электромагнитного поля до деформации.

Качественно результаты деформации можно предсказать, учитывая преобладание вблизи различных участков стенок либо нормального электрического поля, либо тангенциального магнитного поля. Небольшое вдавливание ($\delta V < 0$) стенки в области, где магнитное поле больше электрического ($w_м > w_э$) увеличивает частоту собственных колебаний резонатора ($\delta\omega > 0$). Вдавливание стенки там, где больше электрическое поле, понижает согласно (2.28) частоту собственных колебаний ($\delta\omega < 0$).

§ 2.7. Резонаторный метод измерения диэлектрической проницаемости

Если предварительно откачанный резонатор наполняется *газом* с незначительной диэлектрической проницаемостью, мало отличающейся от единицы, то единственным источником возмущения поля является изменение диэлектрической проницаемости $\delta\varepsilon_r = \varepsilon_r - 1$, а другие вариации отсутствуют ($\delta\mu = \delta V = 0$). Тогда с учетом (2.26) согласно (2.27)

$$\frac{\delta\omega}{\omega} = -\frac{\int\limits_V \delta\varepsilon E^2\,dV}{\int\limits_V(\varepsilon_0 E^2 + \mu_0 H^2)\,dV} = -\frac{\delta\varepsilon\int\limits_V E^2\,dV}{2\varepsilon_0\int\limits_V E^2\,dV} = \frac{1-\varepsilon_r}{2} = \frac{\delta f_0}{f_0}. \qquad (2.29)$$

При этом предполагается, что откачанный резонатор был возбужден на резонансную частоту f_0 и в порядке первого приближения $\mu = 1$. Следовательно, искомая диэлектрическая проницаемость

$$\varepsilon_r = 1 - 2\delta f_0 / f_0 . \qquad (2.30)$$

Для всех резонаторных методов исследования газов характерно одновременное использование двух резонаторов, имеющих до заполнения газом одинаковые резонансные частоты. Заполняется газом один из резонаторов (измерительный), а затем определяется не его новая частота, а *разность частот* измерительного и второго (опорного) резонатора. Это обеспечивает высокую точность, недостижимую при измерениях прямыми методами. Например, при введении воздуха в резонатор с f_0 = 10 ГГц уход частоты составит по уравнению (2.30)

$$\delta f_0 = 0{,}5 f_0 (\varepsilon_r - 1) = 0{,}5 \cdot 10^4 (1{,}000575 - 1) \approx 2{,}28 \text{ МГц}.$$

Такие изменения при прямых методах измерения можно обнаружить только в том случае, когда ошибка измерений будет не больше 10^{-2}%. В то же время при использовании метода сравнения двух частот ошибка измерения ±0,3% обеспечивает правильное значение проницаемости с точностью до пятого знака после запятой. Разность резонансных частот измерительного и опорного резонаторов измеряется методом биений либо осциллографическими методами.

Диэлектрическую проницаемость *твердых* образцов определяют резонаторным методом по изменению резонансной частоты и добротности. Соотношениями, которые дает метод малых возмущений, можно пользоваться только в том случае, когда вносимые в резонатор образцы имеют объем V_1, значительно меньший объема самого резонатора V, и существенно не искажают первоначальное поле.

Если образец размещен в пучности электрического поля с амплитудой E_1, то аналогично (2.29)

$$\frac{\delta\omega}{\omega} = \frac{-\int\limits_V \delta\varepsilon E^2 \, dV}{\int\limits_V (\varepsilon_0 E^2 + \mu_0 H^2) \, dV} = \frac{\int\limits_{V_1} (1 - \varepsilon_r)\varepsilon_0 E_1^2 \, dV}{2\int\limits_V \varepsilon_0 E^2 \, dV} \approx \frac{(1 - \varepsilon_r) E_1^2 V_1 \varepsilon_0}{4 W_{max}} . \qquad (2.31)$$

В отличие от измерений диэлектрической проницаемости газов уход резонансной частоты при внесении твердых диэлектриков зависит не только от размеров и размещения образца, но и от формы резонатора. Если образец имеет форму тонкого длинного цилиндра и расположен вдоль силовых линий электрического поля в центре прямоугольного резонатора с основным типом колебаний, то согласно (2.12) $W_{max} = \varepsilon_0 E_1^2 V / 8$ и по (2.31)

$$\varepsilon_r = 1 - \frac{\delta f_0}{f_0} \frac{V}{2V_1} . \qquad (2.32)$$

ной мощности (рис.2.4, *в*), можно определить полуширину $\Delta\lambda$ кривой проходящей мощности.

Нижний предел измеряемых этим способом добротностей ограничен диапазоном электронной перестройки генератора (например, клистрона). Достаточная точность измерений частоты может быть получена только в том случае, когда нагруженная добротность волномера значительно больше нагруженной добротности исследуемого резонатора.

§ 2.5. Метод малых возмущений

Объемные резонаторы получили широкое применение в электронных приборах СВЧ и измерительной технике. Большинство методов измерения связано с определением изменения частоты резонаторов. В процессе эксплуатации и наладки резонансных электронных приборов требуется подстройка или изменение собственной частоты колебаний резонатора. Количественную оценку изменений частоты и в том и в другом случае производят обычно с помощью теории малых возмущений.

Метод малых возмущений основан на сравнении двух близких стационарных состояний поля. Одно из них характеризуется значениями ω, $\mathbf{E}$, $\mathbf{H}$, V, S другое отличается достаточно малым изменением первоначальных величин и соответствует значениям $\omega + \delta\omega$, $\mathbf{E} + \delta\mathbf{E}$, $\mathbf{H} + \delta\mathbf{H}$, $V + \delta V$, $S + \delta S$, причем предполагается одинаковый порядок малости для всех рассматриваемых изменений.

В первом случае исходное поле в отсутствие токов может быть описано уравнениями Максвелла

$$\operatorname{rot}\dot{\mathbf{H}} = j\omega\varepsilon\dot{\mathbf{E}}, \qquad \operatorname{rot}\dot{\mathbf{E}} = -j\omega\mu\dot{\mathbf{H}}. \tag{2.18}$$

Плотность истоков вектора Пойнтинга для исходного поля

$$\operatorname{div}[\dot{\mathbf{E}}\dot{\mathbf{H}}] = \dot{\mathbf{H}}\operatorname{rot}\dot{\mathbf{E}} - \dot{\mathbf{E}}\operatorname{rot}\dot{\mathbf{H}} = j\omega(\varepsilon\mathbf{E}^2 - \mu\mathbf{H}^2), \tag{2.19}$$

а поток вектора Пойнтинга

$$N_S = \frac{1}{2}\int_S [\dot{\mathbf{E}}\dot{\mathbf{H}}]\,\mathrm{d}\mathbf{S} = \frac{1}{2}j\omega\left(\int_V \varepsilon E^2\,\mathrm{d}V - \int_V \mu H^2\,\mathrm{d}V\right). \tag{2.20}$$

Во втором случае после небольшой вариации ($\delta x \ll x$) всех параметров

$$\operatorname{rot}(\dot{\mathbf{H}} + \delta\dot{\mathbf{H}}) = j(\omega + \delta\omega)(\varepsilon + \delta\varepsilon)(\dot{\mathbf{E}} + \varepsilon\dot{\mathbf{E}}). \tag{2.21}$$

Пренебрегая величинами второго порядка малости и учитывая уравнения (2.18), (2.19), (2.20) и (2.21), получаем следующие соотношения:

$$\operatorname{rot}\delta\dot{\mathbf{H}} = j\delta\omega\varepsilon\dot{\mathbf{E}} + j\omega\delta\varepsilon\dot{\mathbf{E}} + j\omega\varepsilon\delta\dot{\mathbf{E}}; \tag{2.22}$$

$$2\delta N_{S+\delta S} = j\delta\omega\left(\int_V \mu H^2\,dV - \int_V \varepsilon E^2\,dV\right) +$$

$$+ j\omega\int_{\delta V}(\mu H^2 - \varepsilon E^2)\,dV + j\omega\left(\int_V \delta\mu H^2\,dV - \int_V \delta\varepsilon E^2\,dV\right) +$$

$$+ j\omega\int_V (\mu\dot{\mathbf{H}}\delta\dot{\mathbf{H}} + \mu\dot{\mathbf{H}}\delta\dot{\mathbf{H}} - \varepsilon\dot{\mathbf{E}}\delta\dot{\mathbf{E}} - \varepsilon\dot{\mathbf{E}}\delta\dot{\mathbf{E}})\,dV. \qquad (2.23)$$

Для упрощения уравнения (2.23) необходимо учесть несколько тождественных преобразований. Так, аналогично (2.19), с учетом (2.18) и (2.22)

$$\operatorname{div}[\dot{\mathbf{E}}\delta\dot{\mathbf{H}}] = \delta\dot{\mathbf{H}}\operatorname{rot}\dot{\mathbf{E}} - \dot{\mathbf{E}}\operatorname{rot}\delta\dot{\mathbf{H}} =$$

$$= j\omega\mu\dot{\mathbf{H}}\delta\dot{\mathbf{H}} - j\delta\omega\varepsilon\dot{\mathbf{E}}\dot{\mathbf{E}} - j\omega\delta\varepsilon\dot{\mathbf{E}}\dot{\mathbf{E}} - j\omega\varepsilon\dot{\mathbf{E}}\delta\dot{\mathbf{E}},$$

$$\operatorname{div}[\dot{\mathbf{E}}\delta\dot{\mathbf{H}}] = \delta\dot{\mathbf{H}}\operatorname{rot}\dot{\mathbf{E}} - \dot{\mathbf{E}}\operatorname{rot}\delta\dot{\mathbf{H}} =$$

$$= -j\omega\mu\dot{\mathbf{H}}\delta\dot{\mathbf{H}} + j\delta\omega\varepsilon\dot{\mathbf{E}}\dot{\mathbf{E}} + j\omega\varepsilon\dot{\mathbf{E}}\dot{\mathbf{E}} + j\omega\varepsilon\dot{\mathbf{E}}\delta\dot{\mathbf{E}}.$$

Вычитая из первого равенства второе и производя перегруппировку, можно убедиться, что

$$j\omega\mu(\dot{\mathbf{H}}\delta\dot{\mathbf{H}} + \dot{\mathbf{H}}\delta\dot{\mathbf{H}}) - j\omega\varepsilon(\dot{\mathbf{E}}\delta\dot{\mathbf{E}} + \dot{\mathbf{E}}\delta\dot{\mathbf{E}}) =$$

$$= \operatorname{div}[\dot{\mathbf{E}}\delta\dot{\mathbf{H}}] - \operatorname{div}[\dot{\mathbf{E}}\delta\dot{\mathbf{H}}] + 2j\delta\omega\varepsilon\dot{\mathbf{E}}\dot{\mathbf{E}} + 2j\omega\delta\varepsilon\dot{\mathbf{E}}\dot{\mathbf{E}}. \qquad (2.24)$$

При этом

$$\int_V \{\operatorname{div}[\dot{\mathbf{E}}\delta\dot{\mathbf{H}}] - \operatorname{div}[\dot{\mathbf{E}}\delta\dot{\mathbf{H}}]\}\,dV = \oint_S \{[\dot{\mathbf{E}}\delta\dot{\mathbf{H}}] - [\dot{\mathbf{E}}\delta\dot{\mathbf{E}}]\}\,dS = 0, \qquad (2.25)$$

так как на поверхности S металлических стенок, ограничивающих поле в резонаторе, электрическое поле имеет только нормальные составляющие, а магнитное — только касательные.

Поскольку полость ограничена проводящей поверхностью, то потоки энергии, проходящие через эту поверхность, $N_S = 0$, $N_S + \delta N_S = 0$ и вариация энергии $\delta N_S = 0$. Поэтому согласно (2.20)

$$\int_V \mu H^2\,dV = \int_V \varepsilon E^2\,dV, \qquad (2.26)$$

а в соответствии с (2.23) с учетом (2.24) и (2.25)

$$\delta\omega\int_V (\varepsilon E^2 + \mu H^2)\,dV + \omega\int_{\delta V}(\mu H^2 - \varepsilon E^2)\,dV +$$

$$+ \omega\int_V (\delta\mu H^2 + \delta\varepsilon E^2)\,dV = 0. \qquad (2.27)$$

Рассмотрим некоторые следствия из теории малых возмущений.

Приведенный вывод не является достаточно строгим, так как в твердых образцах не выполняются два исходных допущения: $\delta\varepsilon \ll \varepsilon$ и $\varepsilon'' = 0$. Более строгий вывод показывает, что формулы (2.31) и (2.32) не дают заметной ошибки в том случае, когда $V_1 \ll V$, $\varepsilon'' \ll \varepsilon'$ и длина образца много больше его диаметра. Обычно эти условия выполняются. Однако надо помнить, что выражение (2.32) соответствует в этом случае вещественной составляющей ε_r' комплексной диэлектрической проницаемости $\underline{\varepsilon}_r = \varepsilon_r' - j\varepsilon_r''$.

Мнимая компонента $\varepsilon'' = \sigma/\omega$ может быть найдена, если использовать метод малых возмущений при расчете добротности резонатора. Пусть добротность резонатора без образца $Q_0 = \omega_0 W_{\max}/P$, а добротность резонатора с образцом

$$Q_1 = \frac{(\omega_0 + \delta\omega_0)(W_{\max} + \delta W_{\max})}{P + \delta P}.$$

Если учесть, что $\delta\omega_0/\omega_0 \ll 1$, $\delta W_{\max}/W_{\max} \ll 1$ и $V_1 \ll V$, то

$$\left(\frac{1}{Q_1} - \frac{1}{Q_0}\right)\omega_0 W_{\max} = \frac{P + \delta P}{(1 + \delta\omega_0/\omega_0)(1 + \delta W_{\max}/W_{\max})} - P = \delta P =$$
$$= \sigma E_1^2 V_1/2 = \omega_0 \varepsilon'' E_1^2 V_1/2,$$

откуда

$$\varepsilon'' = \left(\frac{1}{Q_1} - \frac{1}{Q_0}\right)\frac{2W_{\max}}{E_1^2 V_1}.$$

В прямоугольном резонаторе с основным типом колебаний с учетом (2.12)

$$\varepsilon_r'' = \left(\frac{1}{Q_1} - \frac{1}{Q_0}\right)\frac{V}{4V_1}. \qquad (2.33)$$

Диэлектрическая проницаемость *жидкостей* может быть измерена резонаторным методом аналогично тому, как это проводится для твердых образцов. Жидкости при этом заливают в тонкую изоляционную трубку из кварца или другого изоляционного материала с малыми потерями. Присутствие трубки оказывает существенное влияние на результаты и должно учитываться в расчетах.

§ 2.8. Открытые резонаторы

Резонаторы, объем которых на большом протяжении не ограничен металлическими поверхностями, называются *открытыми*. К ним могут

быть отнесены *диэлектрические резонаторы*, которые изготовляют из диэлектрика с высокой проницаемостью, выбранной так, что при заданных его форме и размерах выполняются условия резонанса электромагнитных колебаний. В качестве резонатора могут быть применены образцы различной формы, например призмы, шайбы и сферы. В диэлектрических резонаторах используется явление полного внутреннего отражения для волн, угол падения которых больше критического. Таким образом, физические принципы действия полых и диэлектрических резонаторов сходны, поскольку и в том и в другом случае внутри резонатора образуются стоячие волны. В отличие от полого металлического резонатора в диэлектрическом резонаторе отражающей поверхностью является граница раздела воздуха и диэлектрика. Другой его особенностью является наличие внешнего поля.

Диэлектрические резонаторы также отличаются от полых значительно меньшими размерами и в некоторых случаях отсутствием специальных элементов связи с линией передачи (могут быть связаны непосредственно с полем передающей линии), что позволяет создавать на основе таких резонаторов простые и малогабаритные устройства.

Спектр собственных частот диэлектрического резонатора определяется его формой, геометрическими размерами и вещественной компонентой диэлектрической проницаемости, а собственная добротность — мнимой компонентой диэлектрической проницаемости материала.

Остановимся подробнее на другом типе открытых резонаторов, которые образованы двумя *отражателями*, направленными друг на друга. Отражатели располагают в свободном пространстве, либо между ними помещают диэлектрик. Они обычно применяются на миллиметровых и субмиллиметровых волнах, а также в оптическом диапазоне. В этих диапазонах отражатели позволяют получить такие добротности, мощности колебаний и малые плотности спектра собственных колебаний, которые недоступны объемным резонаторам внутри замкнутых металлических полостей. На основных типах колебаний объемные резонаторы в этой области частот имеют слишком малые размеры, а на высших типах колебаний теряют частотно-избирательные свойства из-за чрезмерной густоты спектра их собственных колебаний.

В оптическом диапазоне в качестве отражателей чаще всего используют многослойные диэлектрические зеркала, поскольку коэффициент отражения металла в этом диапазоне недостаточен. В диапазоне миллиметровых волн вместо сплошных зеркал применяют металлические решетки или зеркала с отверстиями (перфорацией), обеспечивающими удобную связь резонатора с другими устройствами.

Все размеры таких резонаторов значительно больше длины волны, поэтому их основные рабочие параметры в первом приближении могут быть получены с помощью геометрической оптики в предположении на-

личия плоских однородных волн конечных размеров. На резонансной (рабочей) частоте резонаторы образуют стоячие волны. Для их возникновения необходимо, чтобы расстояние между зеркалами L равнялось целому числу q полуволн, т. е. $q\lambda_q/2 = L/\cos\theta\cos\varphi$, где θ — угол между направлением распространения луча и оптической осью резонатора; φ — меридианальный угол цилиндрической системы координат. Обычно *оптической осью резонатора* называют общую нормаль к обеим отражающим поверхностям.

Таким образом, резонанс существует одновременно для нескольких частот

$$\nu_{qrn} = \frac{qv}{2L}\cos\theta_m \cos\varphi_n, \tag{2.34}$$

эквидистантно расположенных по шкале частот, и для волн, распространяющихся под углами θ_m, φ_n к оси. Для параксиальных лучей ($\theta_m \approx 0$) расстояние между резонансными частотами, согласно (2.34),

$$\Delta\nu = v/(2L) \tag{2.35}$$

называют *частотой межмодовых биений*. Для метрового воздушного ($v = c$) резонатора $\Delta\nu = 3\cdot10^{10}/(2\cdot10^2) = 150$ МГц. Ширина отдельной спектральной линии определяется добротностью резонатора ($\Delta\nu_0 = \nu/Q_0$).

Добротность открытого резонатора, заполненного диэлектриком с запасенной энергией W_{max},

$$Q_0 = \frac{\omega_0 W}{P} = \frac{2\pi\nu W_{max}}{\alpha W_{max}\, v/L} = \frac{2\pi}{\alpha}\frac{L}{\lambda}, \tag{2.36}$$

где α — коэффициент суммарных потерь за один полный проход $2L$ резонатора лучом.

Потери в резонаторе складываются из рассеяния или поглощения в диэлектрике, потерь в зеркалах (в основном за счет прохождения волны), дифракционных потерь (часть энергии выходит за края зеркал благодаря волновому рассеянию) и иногда потерь на торцах диэлектрика. Излучением через боковые стенки, потерями из-за непараллельности зеркал и поглощения в них в большинстве случаев можно пренебречь. Если потери определяются только частичным выходом излучения из зеркал с коэффициентами отражения ρ_1 и ρ_2, то коэффициент потерь

$$\alpha = (1-\rho_1^2)(1-\rho_2^2). \tag{2.37}$$

При малых дифракционных потерях коэффициент потерь может быть доведен до 0,01, поэтому в оптических резонаторах с L = 1м, λ = 0,63 мкм добротность достигает 10^9. В миллиметровом диапазоне добротность от-

крытого резонатора может доходить до 10^5. В обычных полых резонаторах такие добротности в этом диапазоне недостижимы вследствие роста потерь в металлических стенках.

Предположение о существовании колебаний с частотами (2.34) основано на формировании поля резонатора плоскими волнами. В действительности распределение поля оказывается гораздо более сложным. Строгое исследование электромагнитного поля в открытом резонаторе основано на рассмотрении неоднородной системы временны́х уравнений Максвелла с нулевыми начальными и граничными условиями на зеркалах. После некоторого числа прохождений волн через резонатор поле на поверхностях зеркал становится стационарным. Существует несколько таких распределений, отличных друг от друга и соответствующих различным *нормальным типам колебаний резонатора*. Каждое из них обладает своими потерями и сдвигом фаз φ_{mn} за один проход. Поэтому условие резонанса q-го колебания любого типа приобретает вид

$$2\pi L/\lambda_q + \varphi_{mn} = \pi q,$$

откуда

$$\nu_q = v_0 q/(2L) - v_0\varphi_{mn}/(2\pi L), \qquad (2.38)$$

т. е. резонансная частота будет различной для разных типов колебаний даже при одних и тех же q и L. Правда, разность частот соседних колебаний (q и $q \pm 1$) одного и того же типа совпадает с (2.35) и одинакова для всех типов.

Нормальные типы колебаний называют также *модами*. Запаздывание фазы у краев зеркала по сравнению с его центром составляет несколько десятков градусов, что соответствует долям длины волны. По сравнению с числом длин волн, укладывающихся по диаметру зеркала, это очень небольшая величина, поэтому кривизна фронта волны очень мала, и электромагнитное поле имеет очень малую продольную составляющую. Следовательно, можно считать, что в резонаторе существуют только поперечные волны. Это послужило причиной того, что моды этих волн обозначают так же, как это принято для поперечных электромагнитных волн, т. е. T_{mnq} (T — transverse), где m, n — целые числа (0, 1, 2, ...), обозначающие число изменений знака (направления) напряженности поля на поверхностях зеркал, а q равно числу полуволн, укладывающихся на длине резонатора. Индексы m и n называют поперечными, определяющими *угловые моды*, а q — продольными или аксиальными индексами, определяющими *аксиальные моды*. Число q очень велико по сравнению с m и n, а разность частот между модами (2.35) очень мала по сравнению с самими частотами. Поэтому в обозначениях мод индекс q часто опускают (T_{mn}). Для круглых зеркал индекс m означает число изменений знака поля

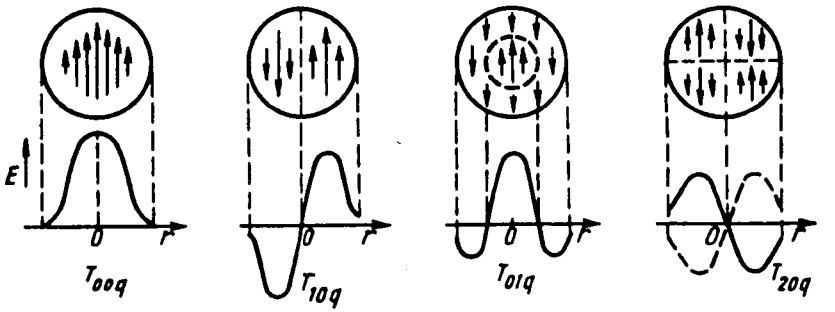


Рис 2.5. Нормальные типы колебаний и распределение электрического поля для круглых зеркал

по углу 180° (для одного и того же радиуса), а n — число изменений знака поля вдоль радиуса (рис. 2.5).

Однородная плоская волна с одинаковой амплитудой по всей поверхности зеркала имеет сравнительно большие потери на отражение, поскольку существует дифракционное расширение отраженного пучка и часть его энергии оказывается за краем зеркала. Малые потери имеют только такие моды, у которых амплитуда спадает к краю пучка лучей. С увеличением m и n повышается напряженность поля на краю зеркала и возрастают дифракционные потери. Наименьшими потерями обладают колебания типа T_{00}, для которых эпюра поля в поперечном сечении имеет один максимум и наиболее плавно спадает к краям зеркала. Эти колебания называются *основными*. Их частоты находят по (2.38).

Режим считают *одномодовым*, если в открытом резонаторе возбуждены только колебания T_{00}, хотя при этом допускаются колебания на нескольких резонансных частотах (2.38). В многомодовом режиме одновременно возникают несколько угловых T_{mn} с отличающимися m и n.

Применение вогнутых зеркал значительно уменьшает дифракционное расширение пучка. В *конфокальном резонаторе* используются два сферических зеркала с радиусом кривизны, равным расстоянию между ними или удвоенному фокусному расстоянию. По законам геометрической оптики параллельный пучок лучей должен был бы собираться каждым из зеркал в фокусе. Однако вследствие дифракции волн, отраженных от зеркал конечных размеров, сечение пучка в фокальной плоскости конечно, хотя по-прежнему минимально; это сечение называют *перетяжкой*.

Волновые поверхности пучка собственных типов колебаний большинства оптических резонаторов (поперечные размеры много меньше продольных) близки к сферическим. Пучок колебаний расширяется при удалении от наименьшего сечения, определяя его продольную структуру в виде гиперболоидной каустической поверхности. Каждое поперечное

сечение такого пучка характеризуется своим радиусом кривизны сферического волнового фронта R и расстоянием w от оси z, на котором амплитуда основной моды уменьшается в e раз. Такого рода пучки называют *гауссовыми*. Основные закономерности распространения гауссова пучка следуют непосредственно из анализа скалярного волнового уравнения для амплитуды напряженности электромагнитного поля

$$\nabla^2 E + k^2 E = 0. \tag{2.39}$$

Его решение можно представить в виде

$$E = E_0 f(x, y, z) e^{-jkz}. \tag{2.40}$$

Функция $f(x, y, z)$, определяющая пространственное распределение амплитуды и фазы гауссова пучка, характеризует его отличия от плоской волны: неоднородность поперечного распределения, расширение по мере удаления от перетяжки и кривизну волнового фронта.

Эта функция является решением уравнения, которое получается после подстановки выражения (2.40) в (2.39):

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} - 2jk\frac{\partial f}{\partial z} = 0, \tag{2.41}$$

если пренебречь $\partial^2 f / \partial z^2$. Последнее допустимо в силу параксиального характера гауссова пучка, при котором изменение рассматриваемой функции вдоль пучка (по координате z) значительно слабее ее изменения в поперечном направлении (по координатам x и y).

Параболическое уравнение (2.41) напоминает по форме стационарное уравнение Шредингера, решение которого хорошо известно. Для основного типа колебаний («фундаментальной моды») его решение имеет вид

$$f(x, y, z) = e^{-j[\underline{p} - kr^2/(2\underline{q})]}, \tag{2.42}$$

где $r^2 = x^2 + y^2$; $\underline{p}$, $\underline{q}$ — комплексные функции координаты z.

Вещественная часть $\mathrm{Re}\,\underline{p}$ определяет смещение фазы пучка на его оси относительно плоской волны, описываемой экспонентой в выражении (2.40). Мнимая часть $\mathrm{Im}\,\underline{p}$ учитывает уменьшение амплитуды пучка на оси, связанное с его расширением. Аналогичным образом комплексный параметр $\underline{q}$ характеризует изменение амплитуды и фазы поля в поперечном направлении.

Учитывая, что волновой фронт представляет собой сферу радиусом R, а амплитуда волны уменьшается в e раз на расстоянии w от оси пучка, получаем

$$\mathrm{Re}(kr^2/2\underline{q}) = kr^2/2R; \quad \mathrm{Im}(kr^2/2\underline{q}) = -r^2/w^2,$$

т. е.

$$\frac{1}{\underline{q}} = \frac{1}{R} - j\frac{2}{kw^2}. \tag{2.43}$$

Если координату z отсчитать от перетяжки и учесть, что в минимальном сечении пучка ($z = 0$) волновой фронт плоский, то параметр (2.43) должен быть чисто мнимым: $\underline{q}(0) = jkw_0^2/2 = jR_0/2$, где w_0 — размер пятна в перетяжке; $R_0 = kw_0$ — конфокальный параметр.

После подстановки выражения (2.42) в (2.41) получим уравнение

$$2j\frac{k}{\underline{q}} + 2k\frac{\partial \underline{p}}{\partial z} + \left(\frac{kr}{\underline{q}}\right)^2\left(1 - \frac{\partial \underline{q}}{\partial z}\right) = 0,$$

которое превращается в тождество только в том случае, когда

$$\partial \underline{p}/\partial z = -j/\underline{q}; \tag{2.44}$$

$$\partial \underline{q}/\partial z = 1. \tag{2.45}$$

Следовательно, выражения (2.44) и (2.45) являются теми условиями, при которых соотношение (2.42) является решением уравнения (2.41). Из формулы (2.45), в частности, для произвольного сечения

$$\underline{q}(z) = \underline{q}(0) + z = z + jR_0/2. \tag{2.46}$$

Вещественная часть (2.46) равняется расстоянию до перетяжки, а мнимая позволяет определить размер пучка в перетяжке $w_0 = R_0/k$.

Подставляя выражение (2.46) в (2.44) и интегрируя его в пределах от 0 до z, получаем

$$\underline{p}(z) = -j\ln(1 - j2z/kw_0^2) =$$

$$= \operatorname{arctg}(2z/kw_0^2) - j\ln\sqrt{1 + (2z/kw_0^2)^2}.^{*} \tag{2.47}$$

Вещественная часть (2.47) представляет собой сдвиг фаз

$$\varphi = \operatorname{arctg}(2z/kw_0^2) \tag{2.48}$$

на оси гауссова пучка относительно фазы однородной плоской волны, излучаемой из сечения перетяжки. Он равен нулю в перетяжке и возрастает при удалении от нее, стремясь в пределе ($z \to \pm\infty$) к $\pm\pi/2$. Мнимая часть (2.47) определяет уменьшение амплитуды поля на оси по мере удаления от перетяжки

$$\operatorname{Im}\underline{q}(z) = -\ln\sqrt{1 + (2z/kw_0^2)^2} = \ln(w_0/w)$$

---

* Здесь учтено, что $\operatorname{Re}\ln(a + jb) = \ln\sqrt{a^2 + b^2}$; $\operatorname{Im}\ln(a + jb) = \operatorname{arctg}(b/a)$.

обратно пропорционально увеличению сечения гауссова пучка.

Таким образом, уравнение основной моды гауссовой волны можно записать в следующем окончательном виде

$$E(r,z)=\frac{w_0}{w}E_0\exp\left[-\frac{r^2}{w^2}-j\left(kz+\frac{kr^2}{2R}-\operatorname{arctg}\frac{2z}{kw_0^2}\right)\right]. \qquad (2.49)$$

Наличие поперечных мод приводит к резкому увеличению дифракционной расходимости излучения и его потерям. Если требуется выделить основную моду и подавить остальные, то используют различие распределения мод с неодинаковыми поперечными индексами. Простейшим способом селекции является диафрагмирование пучка внутри резонатора. В первом приближении поперечный размер диафрагмы должен быть равен поперечному размеру распределения поля моды, следующей за основной. Располагать диафрагму следует там, где размеры мод отличаются наиболее сильно. Однако при этом вносятся потери излучения и в основную моду.

Дискриминацию продольных мод можно выполнить тонким частично прозрачным поглотителем (слой металла толщиной 1...10 нм на прозрачной подложке), расположенном в узле выделяемой моды, где напряженность ее электрического поля равна нулю. Напряженность электрического поля остальных мод в этой точке на оси резонатора имеет ненулевое значение, и поэтому эти моды будут испытывать потери.

§ 2.9. Представление поля в резонаторе в виде набора осцилляторов

До сих пор предполагалось наличие в резонаторе гармонических колебаний только одной частоты. Вместе с тем различным типам колебаний могут соответствовать не только различные пространственные распределения стоячих волн, но, как следует из (2.6), и различные собственные частоты. Поскольку каждый тип колебаний удовлетворяет линейному волновому уравнению, то ему будет удовлетворять любая их линейная комбинация. Поэтому в общем случае электромагнитное поле в резонаторе, изменяющееся во времени по произвольному закону, может быть представлено как соответствующая суперпозиция набора нормальных типов колебаний $\mathbf{E}_l(\mathbf{R})$ и $\mathbf{H}_l(\mathbf{R})$:

$$\mathbf{E}=-\frac{1}{\sqrt{\varepsilon}}\sum_l p_l(t)\mathbf{E}_l(\mathbf{R}); \qquad \mathbf{H}=\frac{1}{\sqrt{\mu}}\sum_l \omega_l q_l(t)\mathbf{H}_l(\mathbf{R}), \qquad (2.50)$$

где временна́я зависимость выражается множителями $p_l(t)$ и $q_l(t)$, а суммирование по l учитывает все характеристические числа и две поля-

ризации возможных типов колебаний (по m, n, p, α), т. е. l является сокращенным обозначением полного набора четырех индексов.

Каждая из компонент (2.50)

$$\mathbf{E}_l = -\frac{1}{\sqrt{\varepsilon}} p_l(t)\mathbf{E}_l(\mathbf{R}), \qquad \mathbf{H}_l = \frac{1}{\sqrt{\mu}} \omega_l q_l(t)\mathbf{H}_l(\mathbf{R}) \tag{2.51}$$

удовлетворяет своим волновым уравнениям

$$\nabla^2\mathbf{E} = \frac{1}{v^2}\frac{\partial^2\mathbf{E}}{\partial t^2}, \qquad \nabla^2\mathbf{H} = \frac{1}{v^2}\frac{\partial^2\mathbf{H}}{\partial t^2} \tag{2.52}$$

и исходит из возможности их решения методом разделения переменных, поскольку при подстановке (2.50) в (2.52) они распадаются на две пары уравнений:

$$\left.\begin{aligned} &\nabla^2\mathbf{E}_l(\mathbf{R}) + \frac{\omega_l^2}{v^2}\mathbf{E}_l(\mathbf{R}) = 0, \ \nabla^2\mathbf{H}_l(\mathbf{R}) + \frac{\omega_l^2}{v^2}\mathbf{H}_l(\mathbf{R}) = 0, \\ &\frac{\mathrm{d}^2 p_l}{\mathrm{d}t^2} + \omega_l^2 p_l = 0, \ \frac{\mathrm{d}^2 q_l}{\mathrm{d}t^2} + \omega_l^2 q_l = 0, \end{aligned}\right\} \tag{2.53}$$

где ω_l^2 — постоянные.

Необходимо обратить внимание на то, что между временны́ми функциями должны иметь место определенные соотношения

$$\frac{\partial q_l}{\partial t} = p_l, \qquad \frac{\partial p_l}{\partial t} = -\omega_l^2 q_l, \tag{2.54}$$

поскольку только при этом уравнения (2.50) будут удовлетворять уравнениям Максвелла. Действительно,

$$\operatorname{rot}\mathbf{E}_l(\mathbf{R}) = \frac{\omega_l}{v}\mathbf{H}_l(\mathbf{R}), \qquad \operatorname{rot}\mathbf{H}_l(\mathbf{R}) = \frac{\omega_l}{v}\mathbf{E}_l(\mathbf{R}), \tag{2.55}$$

и

$$\operatorname{rot}\mathbf{E} = -\frac{1}{\sqrt{\varepsilon}}\sum_l p_l \operatorname{rot}\mathbf{E}_l(\mathbf{R}) = -\frac{1}{\sqrt{\varepsilon}}\sum_l p_l \frac{\omega_l}{v}\mathbf{H}_l(\mathbf{R}) =$$

$$= -\sqrt{\mu}\sum_l \omega_l \frac{\partial q_l}{\partial t}\mathbf{H}_l(\mathbf{R}) = -\mu\frac{\partial\mathbf{H}}{\partial t},$$

$$\operatorname{rot}\mathbf{H} = \frac{1}{\sqrt{\mu}}\sum_l \omega_l q_l \operatorname{rot}\mathbf{H}_l(\mathbf{R}) = \frac{1}{\sqrt{\mu}}\sum_l \frac{\omega_l^2}{v} q_l\mathbf{E}_l(\mathbf{R}) =$$

$$= -\sqrt{\varepsilon}\sum_l \frac{\partial p_l}{\partial t}\mathbf{E}_l(\mathbf{R}) = \varepsilon\frac{\partial\mathbf{E}}{\partial t}.$$

Таким образом, функции $\mathbf{E}_l(\mathbf{R})$ и $\mathbf{H}_l(\mathbf{R})$ являются *собственными функциями* оператора $\nabla\times\nabla\times$, соответствующими *собственным значениям* ω_l^2/v^2. Эти функции в резонаторе с идеально проводящими стенками ортогональны и могут быть нормированы по единице:

$$\int \mathbf{E}_\alpha(\mathbf{R})\mathbf{E}_\beta(\mathbf{R})\,dV = \delta_{\alpha\beta}, \qquad \int \mathbf{H}_\alpha(\mathbf{R})\mathbf{H}_\beta(\mathbf{R})\,dV = \delta_{\alpha\beta}, \tag{2.56}$$

где $\delta_{\alpha\beta}$ — оператор Кронекера.

Энергию, запасенную в резонаторе, можно представить в виде

$$W = \frac{1}{2}\int(\varepsilon\mathbf{E}\mathbf{E} + \mu\mathbf{H}\mathbf{H})\,dV = \frac{1}{2}\sum_l (p_l^2 + \omega_l^2 q_l) = \mathcal{H}, \tag{2.57}$$

если векторы напряженностей заменить их разложениями (2.50) и учесть условие (2.56).

Из уравнений (2.54) и (2.57) следует, что входящие в них параметры удовлетворяют классическим уравнениям Гамильтона

$$\frac{\partial\mathcal{H}}{\partial p_l} = \frac{\partial q_l}{\partial t} = p_l, \qquad \frac{\partial\mathcal{H}}{\partial q_l} = -\frac{\partial p_l}{\partial t} = \omega_l^2 q_l. \tag{2.58}$$

Таким образом, выражение для энергии электромагнитного поля (2.57), запасенной в резонаторе, является *функцией Гамильтона*, а q_l и p_l представляют собой «обобщенные координаты» и «канонически сопряженные импульсы».

Глава 3
Направляемые волны и волноводы

Переход от обычного колебательного контура с сосредоточенными параметрами L, C и R к объемному резонатору был определен особенностями электромагнитных колебаний сверхвысокой частоты (СВЧ), охватывающих участок электромагнитного спектра от 30 МГц до 1500 ГГц (с длинами волн 10 м ... 200 мкм и до длин инфракрасных волн). При этом линейные размеры колебательных цепей оказываются сравнимыми с длиной волны и само представление о контуре с сосредоточенными параметрами теряет смысл. В этих условиях резонатор с замкнутой полостью, предотвращающий излучение, представляет собой наиболее простое конструктивное решение колебательной электромагнитной системы с минимальными потерями энергии.

Для канализации энергии СВЧ используют направляющие системы также принципиально иного типа, называемые *волноводами*. Наиболее характерные из них представляют собой металлическую трубу, по внутренней полости которой осуществляется передача энергии. Такая конструкция позволяет, как и в случае полых резонаторов, полностью исключить излучение во внешнее пространство, резко возрастающее по мере того, как поперечные размеры линий передач приближаются к длине волны. Однако этим не исчерпываются их преимущества перед двухпроводными линиями, которые обладают низкой помехозащищенностью, значительными потерями в металле и диэлектрике на больших частотах, низкой электрической прочностью (не в состоянии обеспечить передачу больших мощностей, получаемых от современных генераторов и усилителей СВЧ) и требуют использования устройств для поддержания недостаточно жестких проводников. Уменьшение расстояния между проводниками, необходимое при повышении частоты, затрудняет изготовление линии с требуемыми жесткими допусками на взаимное расположение проводников .

Коаксиальная линия обеспечивает практически полную экранировку и отсутствие излучения. Однако им присущи остальные недостатки двухпроводных линий. По сравнению с ними полые волноводы обладают наибольшей простотой и жесткостью конструкции, а также большей электрической прочностью, отсутствием опор для внутреннего проводника и потерь на его поверхности, в диэлектрике.

Широкое распространение волноводов и резонаторов связано со все увеличивающейся областью использования СВЧ и их спецификой. СВЧ применяют в радиолокации, так как геометрические размеры антенн СВЧ превышают длину волны и позволяют, подобно оптическим устройствам, получать остро направленные излучения для нахождения координат удаленных объектов. Наземная радиоастрономия и космическая связь возможны только в диапазоне СВЧ, поскольку только такие волны способны проникать через ионизированные слои, окружающие Землю, без заметных потерь. Распространение СВЧ в телефонной и телевизионной связи обу-

словлено широкими возможностями для перехода на многоканальные системы. Создание квантовомеханических генераторов и усилителей СВЧ, как и применение СВЧ в радиоспектроскопии для анализа строения вещества, обусловлено тем, что кванты излучения на СВЧ имеют энергию, соизмеримую с разностью близко расположенных энергетических уровней многих атомов и молекул. Успехи, достигнутые во всех отраслях техники СВЧ, связаны с развитием той отрасли электроники, которая использует инерцию электронов (клистроны, магнетроны, лампы бегущей волны). Соизмеримость периода СВЧ-колебаний с временем пролета электронов в межэлектродном промежутке электровакуумных приборов позволила создать генераторы СВЧ значительной мощности.

§ 3.1. Направляемые электромагнитные волны

В волноводах или направляющих системах электромагнитные волны используются для передачи энергии и информации от источника (генератора) к потребителю по определенному пути. Электромагнитную волну, распространяющуюся вдоль такой системы, называют *направляемой*. Направляемые волны в отличие от свободно распространяющихся в пространстве могут существовать только при наличии каких-либо направляющих элементов, например металлических, диэлектрических или полупроводящих поверхностей. Полые металлические трубы и двухпроводные линии являются частными случаями направляющих систем. Одна из главных особенностей направляемой волны — отсутствие потока энергии и излучения в поперечном направлении. Поле этой волны сосредоточено в поперечном сечении ограниченных размеров, становясь равным нулю или очень малым на некотором расстоянии от оси системы. Следовательно, направляемая волна всегда неоднородна.

В однородной направляющей системе, в которой передача электромагнитной энергии происходит только вдоль одного направления, например z,

$$\dot{\mathbf{E}}(x,y,z,t) = \dot{\mathbf{E}}(x,y)\mathrm{e}^{j(\omega t - k_z z)};$$
$$\dot{\mathbf{H}}(x,y,z,t) = \dot{\mathbf{H}}(x,y)\mathrm{e}^{j(\omega t - k_z z)}. \qquad (3.1)$$

В этом случае частная производная $\partial/\partial z = -jk_z$, оператор набла $\nabla = \nabla_\perp - jk_z\mathbf{e}_z$ и лапласиан $\Delta = \nabla^2 = \nabla_\perp^2 - k_z^2$, где $\nabla_\perp$ — оператор набла по поперечным координатам (x, y). Поэтому трехмерные волновые уравнения

$$\nabla^2\dot{\mathbf{E}}^2 + k^2\dot{\mathbf{E}} = 0, \qquad \nabla^2\dot{\mathbf{H}}^2 + k^2\dot{\mathbf{H}} = 0 \qquad (3.2)$$

преобразуются здесь в двумерные (для поперечной плоскости направляющей системы):

$$\nabla_\perp^2 \dot{\mathbf{E}}^2 + \chi^2 \dot{\mathbf{E}} = 0, \qquad \nabla_\perp^2 \dot{\mathbf{H}}^2 + \chi^2 \dot{\mathbf{H}} = 0, \tag{3.3}$$

где *поперечный волновой коэффициент* (волновое число стоячей волны)

$$\chi = \pm\sqrt{k^2 - k_z^2}. \tag{3.4}$$

Дальнейшее упрощение расчета поля в волноводе (решение двумерной «мембранной» задачи) может быть получено, если учесть, что поперечные составляющие $\mathbf{E}_\perp$ и $\mathbf{H}_\perp$ в направляющих системах являются однозначными функциями продольных. В этом можно убедиться, если перейти от системы из двух векторных уравнений Максвелла $\operatorname{rot}\dot{\mathbf{H}} = j\omega\varepsilon\dot{\mathbf{E}}$, $\operatorname{rot}\dot{\mathbf{E}} = -j\omega\mu\dot{\mathbf{H}}$ к системе из шести скалярных уравнений для их проекций и затем решить эту систему относительно поперечных составляющих. В прямоугольной системе координат это приводит к выражениям

$$\begin{aligned}
j\chi^2 \dot{E}_x &= k_z \frac{\partial \dot{E}_z}{\partial x} + \omega\mu \frac{\partial \dot{H}_z}{\partial y};\\
j\chi^2 \dot{E}_y &= k_z \frac{\partial \dot{E}_z}{\partial y} - \omega\mu \frac{\partial \dot{H}_z}{\partial x};\\
j\chi^2 \dot{H}_x &= -j\omega\varepsilon \frac{\partial \dot{E}_z}{\partial y} + k_z \frac{\partial \dot{H}_z}{\partial x};\\
j\chi^2 \dot{H}_y &= j\omega\varepsilon \frac{\partial \dot{E}_z}{\partial x} + k_z \frac{\partial \dot{H}_z}{\partial y}.
\end{aligned} \tag{3.5}$$

Таким образом, если известно распределение продольных составляющих поля по поперечному сечению $\dot{E}_z(x,y)$ и $\dot{H}_z(x,y)$, то можно вычислить поперечные составляющие:

$$\begin{aligned}
\dot{\mathbf{E}}_\perp &= \dot{E}_x\mathbf{e}_x + \dot{E}_y\mathbf{e}_y = \\
&= -j\frac{k_z}{\chi^2}\left(\frac{\partial \dot{E}_z}{\partial x}\mathbf{e}_x + \frac{\partial \dot{E}_z}{\partial y}\mathbf{e}_y\right) + j\frac{\omega\mu}{\chi^2}\left(\frac{\partial \dot{H}_z}{\partial x}\mathbf{e}_x - \frac{\partial \dot{H}_z}{\partial y}\mathbf{e}_y\right) = \\
&= -j\frac{k_z}{\chi^2}\operatorname{grad}_\perp \dot{E}_z + j\frac{\omega\mu}{\chi^2}[\mathbf{e}_z \operatorname{grad}_\perp \dot{H}_z];\\
\dot{\mathbf{H}}_\perp &= \dot{H}_x\mathbf{e}_x + \dot{H}_y\mathbf{e}_y = \\
&= j\frac{\omega\varepsilon}{\chi^2}\left(\frac{\partial \dot{E}_z}{\partial y}\mathbf{e}_x - \frac{\partial \dot{E}_z}{\partial x}\mathbf{e}_y\right) - j\frac{k_z}{\chi^2}\left(\frac{\partial \dot{H}_z}{\partial x}\mathbf{e}_x + \frac{\partial \dot{H}_z}{\partial y}\mathbf{e}_y\right) = \\
&= -j\frac{\omega\varepsilon}{\chi^2}[\mathbf{e}_z \operatorname{grad}_\perp \dot{E}_z] - j\frac{k_z}{\chi^2}\operatorname{grad}_\perp \dot{H}_z,
\end{aligned} \tag{3.6}$$

где индекс $\perp$ при градиенте означает, что производные берутся только по поперечным координатам.

Следовательно, для нахождения структуры полного поля в общем случае достаточно решить с учетом граничных условий только два дифференциальных уравнения

$$\nabla_\perp^2 \dot{E}_z + \chi^2 \dot{E}_z = 0; \qquad \nabla_\perp^2 \dot{H}_z + \chi^2 \dot{H}_z = 0, \tag{3.7}$$

являющихся проекциями векторных уравнений (3.3), а затем найти поперечные составляющие из (3.5) или (3.6).

Поскольку поперечные составляющие поля определяются скалярными функциями E_z и H_z и пропорциональны их градиентам, то исходя из аналогии с электростатическим полем можно утверждать, что E_z и H_z являются для составляющих $E_\perp$ и $H_\perp$ *потенциальными функциями.*

В однородной линии без потерь волновой процесс передачи энергии характеризуется неизменностью амплитуды распространяющейся волны (отсутствует затухание) и постоянством ее скорости (фаза изменяется по линейному закону). Волна (3.1) обладает такими свойствами только в том случае, когда коэффициент фазы волны k_z будет вещественным числом. Из уравнения (3.4) следует, что это возможно для определенных значений поперечного волнового коэффициента

$$\chi \leqslant k = \omega\sqrt{\varepsilon\mu} = 2\pi/\lambda = 2\pi f\sqrt{\varepsilon\mu}$$

или частоты

$$f = v/\lambda \geqslant \chi/(2\pi\sqrt{\varepsilon\mu}).$$

Минимальную частоту, при которой по направляющей системе возможно свободное распространение волны, находят из условия

$$f_{кр} = \chi/(2\pi\sqrt{\varepsilon\mu}). \tag{3.8а}$$

Соответствующее значение максимальной длины волны, при которой еще возможен волновой процесс,

$$\lambda_{кр} = v/f_{кр} = 2\pi/\chi. \tag{3.8б}$$

Длина волны Λ в направляющей системе и, в частности, в волноводе равна минимальному расстоянию между поперечными сечениями z, в которых колебания сдвинуты по фазе на 2π. Поэтому в соответствии с (3.1), (3.4), (3.7), (3.8а) и (3.8б) длина волны

$$\Lambda = \frac{2\pi}{k_z} = \frac{\lambda}{\sqrt{1-(\lambda/\lambda_{кр})^2}} = \frac{\lambda}{\sqrt{1-(f_{кр}/f)^2}} \geqslant \lambda, \tag{3.9}$$

а фазовая скорость волны в волноводе

$$v_\phi = \frac{\omega}{k_z} = \Lambda f = \frac{v}{\sqrt{1-\lambda^2/\lambda_{кр}^2}} = \frac{v}{\sqrt{1-f_{кр}^2/f^2}} \geqslant v = \frac{1}{\sqrt{\varepsilon\mu}}. \qquad (3.10)$$

Если $\chi > k$ или $f < f_{кр}$, то продольное волновое число становится согласно (3.14) мнимой величиной, а фаза электромагнитного поля (3.1) постоянной (не меняющейся вдоль z); амплитуда убывает вдоль волновода по экспоненциальному закону, что является признаком затухания поля, отсутствия бегущей волны и переноса энергии вдоль направляющей системы. Диапазон частот $f < f_{кр}$ и длин волн $\lambda > \lambda_{кр}$ называют *областью отсечки* или запредельной областью. В этом режиме линия и волновод не могут быть использованы для передачи энергии. Условиями распространения волн по всякой направляющей системе являются неравенства

$$\lambda < \lambda_{кр}, \quad f > f_{кр}. \qquad (3.11)$$

§ 3.2. Классификация волн

За основу классификации обычно принимают наличие или отсутствие продольных составляющих поля E_z и H_z, параллельных направлению распространения. Что касается поперечных составляющих $E_\perp$ и $H_\perp$, то ими обязательно обладает поле любой волны. Их существование является необходимым условием наличия продольной компоненты вектора Пойнтинга Π_z, обусловливающей передачу энергии вдоль направляющей системы. Рассмотрим основные типы направляемых волн.

Поперечные электромагнитные волны (типа Т) отличаются отсутствием продольных составляющих ($E_z = 0$, $H_z = 0$).

Поперечные составляющие напряженностей при этом могут отличаться от нуля только в том случае, когда в (3.6) поперечный волновой коэффициент $\chi = 0$. В этом случае, как видно из (3.14), коэффициент распространения волны типа Т $k_z = k = \omega\sqrt{\varepsilon\mu}$ всегда равен волновому числу среды, которой заполнена данная направляющая система.

Критическая частота (3.8а) при этом равна нулю, критическая длина волны (3.8б) стала бесконечно большой, а длина волны в направляющей системе (3.9) оказывается той же, что и у волны в свободном пространстве (λ). Таким образом, если в направляющей системе могут существовать волны типа Т, то они не имеют отсечки; волны могут распространяться при любой частоте, включая нулевую, т. е. при постоянном токе. Это согласуется с тем обстоятельством, что структура поля волны типа Т в поперечной плоскости идентична структуре электростатического и стацио-

нарного магнитного полей в направляющей системе. Действительно, при $\chi = 0$ уравнения (3.3) приобретают тот же вид ($\nabla_\perp^2 \mathbf{E}_\perp = 0$, $\nabla_\perp^2 \mathbf{H}_\perp = 0$), что и уравнения Лапласа. Следовательно, волна типа Т может распространяться лишь в таких направляющих системах, где возможно существование электростатического поля, т. е. в системах из двух или нескольких изолированных проводников (в поперечном сечении они образуют многосвязную область). Волны типа Т не могут возбуждаться в полых металлических трубах, диэлектрических волноводах, линиях поверхностной волны, системах, состоящих из нескольких разнородных диэлектрических слоев.

Характеристическим (волновым) сопротивлением передающей линии принято называть отношение напряженностей поперечных электрического и магнитных полей. Для волн типа Т характеристическое сопротивление

$$Z_T = \frac{\dot{E}_x}{\dot{H}_y} = -\frac{\dot{E}_y}{\dot{H}_x} = \frac{k_z}{\omega\varepsilon} = \sqrt{\frac{\mu}{\varepsilon}}, \qquad (3.12)$$

т. е. не зависит от частоты и совпадает с волновым сопротивлением Z_c однородной плоской волны.

Остальные типы волн отличаются наличием продольной компоненты. Продольную составляющую имеет только магнитное поле магнитных волн (типа Н). Для волн этого типа согласно (3.5) характеристическое сопротивление

$$Z_H = \frac{\dot{E}_x}{\dot{H}_y} = -\frac{\dot{E}_y}{\dot{H}_x} = \frac{\omega\mu}{k_z} = \frac{Z_T}{\sqrt{1-(\lambda/\lambda_{кр})^2}} = \frac{Z_T}{\sqrt{1-(f_{кр}/f)^2}}, \qquad (3.13)$$

а поперечные компоненты $\dot{\mathbf{E}}_\perp = Z_H \dot{\mathbf{H}}_\perp \times \mathbf{e}_z$ оказываются синфазными, взаимно перпендикулярными и пропорциональными друг другу в любой точке волновода без потерь.

Волны, имеющие продольную компоненту только у электрического поля, называют *электрическими волнами* (типа Е). Для них по (3.5)

$$Z_E = \frac{k_z}{\omega\varepsilon} = Z_T\sqrt{1-\left(\frac{\lambda}{\lambda_{кр}}\right)^2} = Z_T\sqrt{1-\left(\frac{f_{кр}}{f}\right)^2} \qquad (3.14)$$

и $\dot{\mathbf{E}}_\perp = Z_E \dot{\mathbf{H}}_\perp \times \mathbf{e}_z$ т. е. сохраняются синфазность, пропорциональность и ортогональность поперечных компонент.

Общим свойством волн магнитного и электрического типов является зависимость их параметров от частоты. Это относится и к характеристическим сопротивлениям (3.13) и (3.14), и к длинам волн (3.9), и к фазо-

вым скоростям (3.10). Это свойство называют *дисперсией*, а волны соответствующего типа — диспергирующими волнами. Частотную зависимость характеристических сопротивлений иллюстрирует рис. 3.1. Аналогичный вид имеют частотные характеристики фазовой скорости (3.10) и скорости распространения энергии, групповой скорости:

$$v_э = v_{гр} = \frac{\partial\omega}{\partial k_z} = v\sqrt{1-\left(\frac{\lambda}{\lambda_{кр}}\right)^2} = v\sqrt{1-\left(\frac{f_{кр}}{f}\right)^2}. \qquad (3.15)$$

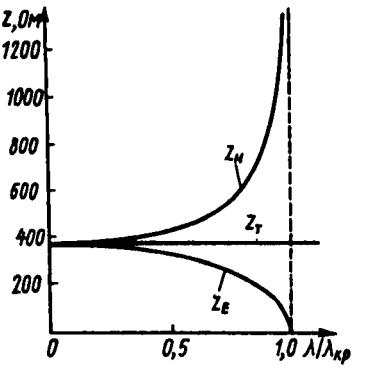


Рис. 3.1. Частотная зависимость характеристических сопротивлений прямоугольных волноводов

Волны типа Т принципиально отличаются от волн магнитного и электрического типов тем, что не имеют, как отмечалось, критической частоты ($f_{кр} = 0$); их характеристическое сопротивление (3.12) и скорости распространения ($v_ф = v_э = v$) не зависят от частоты. Эти волны не обладают дисперсией, относятся к недиспергирующим волнам.

Поле в направляющих системах в общем случае можно рассматривать как линейную комбинацию (наложение) рассмотренных ранее волн. Волны сложного типа, для которых одновременно не равны нулю все шесть составляющих поля, называют *гибридными*.

§ 3.3. Прямоугольные волноводы

Изучение свойств разнообразных типов волноводов начнем с волноводов, имеющих прямоугольную форму поперечного сечения. Для нахождения уравнений поля в волноводе нет необходимости решать волновое уравнение (3.7) с самого начала, поскольку прямоугольный волновод можно рассматривать как полученный из прямоугольного резонатора (см. рис. 2.1) большой длины ($c \to \infty$) после удаления металлических стенок, нормальных к оси z, и замены их генератором ($Z = 0$) и нагрузкой без отражения ($Z = Z_c$). Как и в резонаторе, установившееся поле можно представить как результат распространения и многократных отражений плоской волны (1.1), компоненты волнового вектора которой по-прежнему связаны уравнением (2.1). Как и в резонаторе, отражения волн стенками волновода образуют в поперечных направлениях (по осям x и y) стоячие волны.

В отличие от резонатора в согласованном волноводе отсутствуют отражения при движении волны в продольном направлении. Поэтому в направлении оси z образуется не стоячая, а бегущая волна, с помощью которой и осуществляется распространение энергии в волноводе от генератора к нагрузке. Это различие учитывается при переходе от системы уравнений резонатора (2.5) к системе уравнений поля в волноводе так: последний сомножитель $\begin{pmatrix}\sin\\ \cos\end{pmatrix} k_z z$, характеризующий стоячую волну вдоль оси z, должен быть заменен множителем $e^{-jk_z z}$, определяющим изменение амплитуды и фазы бегущей волны:

$$\begin{aligned}\dot{E}_x &= \dot{E}_1 \cos k_x x \sin k_y y\, e^{-jk_z z};\\ \dot{E}_y &= \dot{E}_2 \sin k_x x \cos k_y y\, e^{-jk_z z};\\ \dot{E}_z &= \dot{E}_3 \sin k_x x \sin k_y y\, e^{-jk_z z}.\end{aligned} \tag{3.16}$$

Структура стоячей волны в поперечном сечении определяется, как и в резонаторе, целыми значениями характеристических чисел (m и n) и компонентами волнового вектора для осей x и y:

$$k_x = \pi m/a, \quad k_y = \pi n/b. \tag{3.17}$$

Следовательно, в прямоугольном волноводе поперечный волновой коэффициент (3.4)

$$\chi = \sqrt{k^2 - k_z^2} = \sqrt{k_x^2 + k_y^2} = \pi\sqrt{(m/a)^2 + (n/b)^2} \tag{3.18}$$

и критическая длина волны (3.8б)

$$\lambda_{кр} = \frac{2}{\sqrt{(m/a)^2 + (n/b)^2}}. \tag{3.19}$$

При заполнении волновода воздухом ($v \approx c_0$) фазовая скорость волны вдоль волновода оказывается больше скорости света. Это явление объясняется тем, что передача энергии вдоль волновода происходит в результате многократных отражений плоской волны, движущейся под некоторым углом γ к оси волновода, причем

$$\cos\gamma = k_z/k = \lambda/\Lambda = \sqrt{1-(\lambda/\lambda_{кр})^2} = \sqrt{1-(f_{кр}/f)^2}.$$

Скорость волн вдоль волновода и угол наклона волнового вектора к оси волновода зависят от поперечных размеров a и b волновода, типа стоячей волны (m и n) и частоты f электромагнитных колебаний. При критической частоте $f_{кр}$ (угол $\gamma = 90°$) плоская волна движется перпендикулярно оси волновода и распространение энергии вдоль волновода пре-

кращается ($v_{эн} = 0$). По мере увеличения частоты направление движения плоской волны все больше приближается к продольной оси волновода.

Структура электромагнитного поля в волноводе зависит от частоты, размеров волновода и типа возбужденных волн.

Для волн электрического типа уравнения, описывающие структуру поля, можно найти, исходя из полученной ранее зависимости от координат электрических компонент (3.16). Так, если принять $E_з = \chi^2 D$, то из (3.5)

$$\dot{E}_x = -j\frac{k_z}{\chi^2}\frac{\partial \dot{E}_z}{\partial x} = -jk_z D\frac{\pi m}{a}\cos\frac{\pi mx}{a}\sin\frac{\pi ny}{b}e^{-jk_z z};$$

$$\dot{E}_y = -j\frac{k_z}{\chi^2}\frac{\partial \dot{E}_z}{\partial y} = -jk_z D\frac{\pi n}{b}\sin\frac{\pi mx}{a}\cos\frac{\pi ny}{b}e^{-jk_z z}; \qquad (3.20а)$$

$$\dot{E}_z = \chi^2 D\sin\frac{\pi mx}{a}\sin\frac{\pi ny}{b}e^{-jk_z z};$$

$$\dot{H}_x = j\frac{\omega\varepsilon}{\chi^2}\frac{\partial \dot{E}_z}{\partial y} = j\omega\varepsilon D\frac{\pi n}{b}\sin\frac{\pi mx}{a}\cos\frac{\pi ny}{b}e^{-jk_z z};$$

$$\dot{H}_y = -j\frac{\omega\varepsilon}{\chi^2}\frac{\partial \dot{E}_z}{\partial x} = j\omega\varepsilon D\frac{\pi m}{a}\cos\frac{\pi mx}{a}\sin\frac{\pi ny}{b}e^{-jk_z z}; \qquad (3.20б)$$

$$\dot{H}_z = 0.$$

Система (3.20) показывает, что в общем случае в волноводе может существовать бесконечное число волн электрического типа, отличающихся значениями характеристических чисел m и n. Поэтому в обозначении каждого типа волн указываются в виде индексов значения m и n, например E_{mn}.

Уравнения (3.20) также показывают, что поле волн типа Е может существовать в волноводе, если ни одно из характеристических чисел m и n не равно нулю, т. е. в прямоугольных волноводах не могут быть возбуждены волны типа E_{00}, E_{0n}, E_{m0}.

Структура поля некоторых волн типа Е изображена на рис. 3.2. Эти волны могут быть образованы различными сочетаниями ячеек простейшей волны типа E_{11}. Структура силовых линий волн типа Е особенно отчетливо показывает, что в теории волноводов теряет свой смысл понятие напряжения, широко используемое в обычной теории цепей.

Для волн магнитного типа структура поля может быть найдена, если в (3.16) обозначить $j(\pi/\omega\mu)(m\dot{E}_2/a - n\dot{E}_1/b) = \chi^2 C$. Тогда из второго уравнения Максвелла и из (3.5)

$$\dot{H}_z = j\frac{1}{\omega\mu}\left(\frac{\partial \dot{E}_y}{\partial x} - \frac{\partial \dot{E}_x}{\partial y}\right) = \chi^2 C \cos\frac{\pi m x}{a}\cos\frac{\pi n y}{b} e^{-jk_z z};$$

$$\dot{H}_y = -j\frac{k_z}{\chi^2}\frac{\partial \dot{H}_z}{\partial y} = jk_z C \frac{\pi n}{b}\cos\frac{\pi m x}{a}\sin\frac{\pi n y}{b} e^{-jk_z z};$$

$$\dot{H}_x = -j\frac{k_z}{\chi^2}\frac{\partial \dot{H}_z}{\partial x} = jk_z C \frac{\pi m}{a}\sin\frac{\pi m x}{a}\cos\frac{\pi n y}{b} e^{-jk_z z}; \qquad (3.21)$$

$$\dot{E}_x = -j\frac{\omega\mu}{\chi^2}\frac{\partial \dot{H}_z}{\partial y} = j\omega\mu C \frac{\pi n}{b}\cos\frac{\pi m x}{a}\sin\frac{\pi n y}{b} e^{-jk_z z};$$

$$\dot{E}_y = j\frac{\omega\mu}{\chi^2}\frac{\partial \dot{H}_z}{\partial x} = -j\omega\mu C \frac{\pi m}{a}\sin\frac{\pi m x}{a}\cos\frac{\pi n y}{b} e^{-jk_z z};$$

$$\dot{E}_z = 0.$$

Основной волной или волной основного типа называют волну с наибольшей критической длиной. Волны других типов называют *высшими*. Индексы m и n не могут одновременно равняться нулю, так как при этом поле, определяемое уравнениями (3.21), отсутствует. Следовательно, волна основного типа характеризуется тем, что вдоль широкой стенки прямоугольного волновода укладывается только одна полуволна, а вдоль узкой стенки поле не меняется.

Размеры прямоугольного волновода принято обозначать так, чтобы ось x была направлена вдоль широкой стенки волновода, т. е. всегда $a \geqslant b$. В этом случае основная волна типа H_{10} соответствует значениям $m = 1$ и $n = 0$. В соответствии с (3.19) для волны типа H_{10} $\lambda_{кр} = 2a$; для волны типа H_{01} $\lambda_{кр} = 2b$; для всех волн типа H_{m0} $\lambda_{кр} = 2a/m$, для волн типа H_{0n} $\lambda_{кр} = 2b/n$.

При заданной частоте передатчика наименьшие размеры, потери и массу волновода получают при использовании волн основного типа. Чтобы избежать возникновения волн высших типов (H_{20}, H_{01} и т. д.), длина волны в свободном пространстве должна находиться в пределах

$$2a > \lambda > a > 2b. \qquad (3.22)$$

Структура поля волн типа H_{mn} в прямоугольном волноводе определяется уравнениями (3.21) и зависит от числа полуволн m и n, укладывающихся вдоль широкой и узкой сторон, волновода. На рис.3.3 изображено расположение силовых линий поля некоторых типов магнитных волн. Из рисунка видно, что структура всех волн типа H_{mn} может быть образована из двух простейших ячеек, соответствующих волнам типов H_{10} и H_{11}.

Элементы и способы возбуждения волн в волноводах в принципе те же, что и в резонаторах: штырь, виток с током или связь с другим СВЧ-

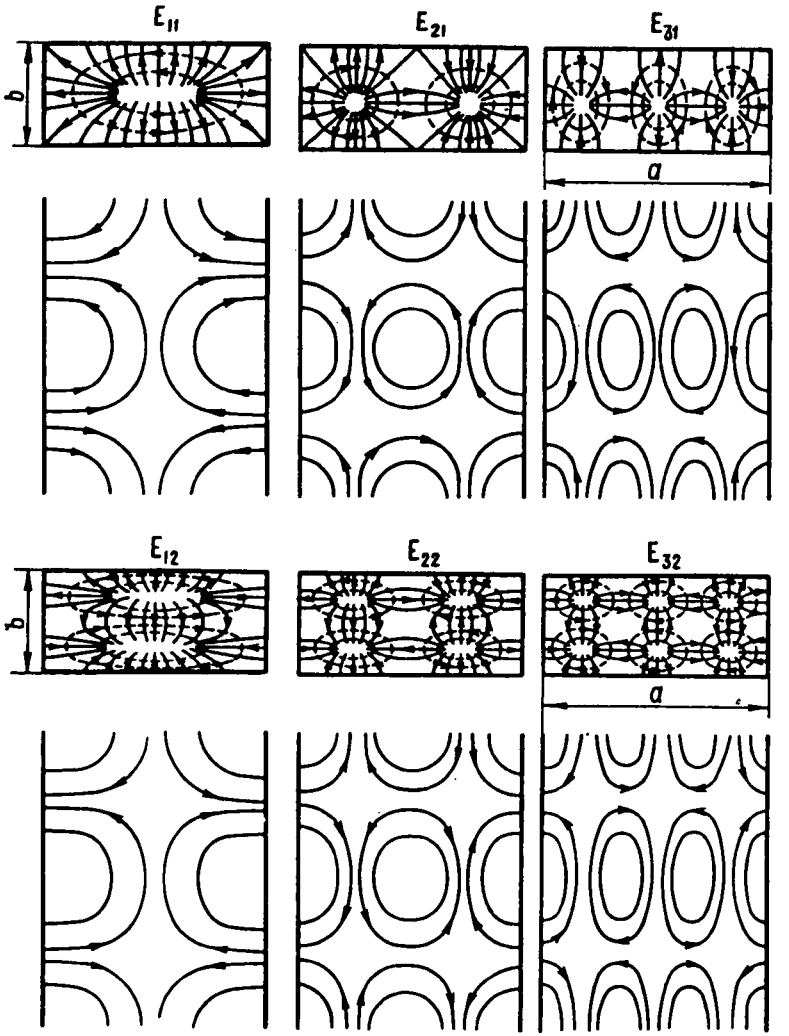


Рис. 3.2. Структура электрических волн в прямоугольном волноводе

устройством (волноводом или резонатором) через щель или диафрагму. Устройство должно возбуждать волну какого-либо одного типа и делать невозможным распространение волн других типов. Штырь должен быть расположен так, чтобы создавать в волноводе электрическое поле, совпадающее по направлению с электрическими силовыми линиями поля вол-

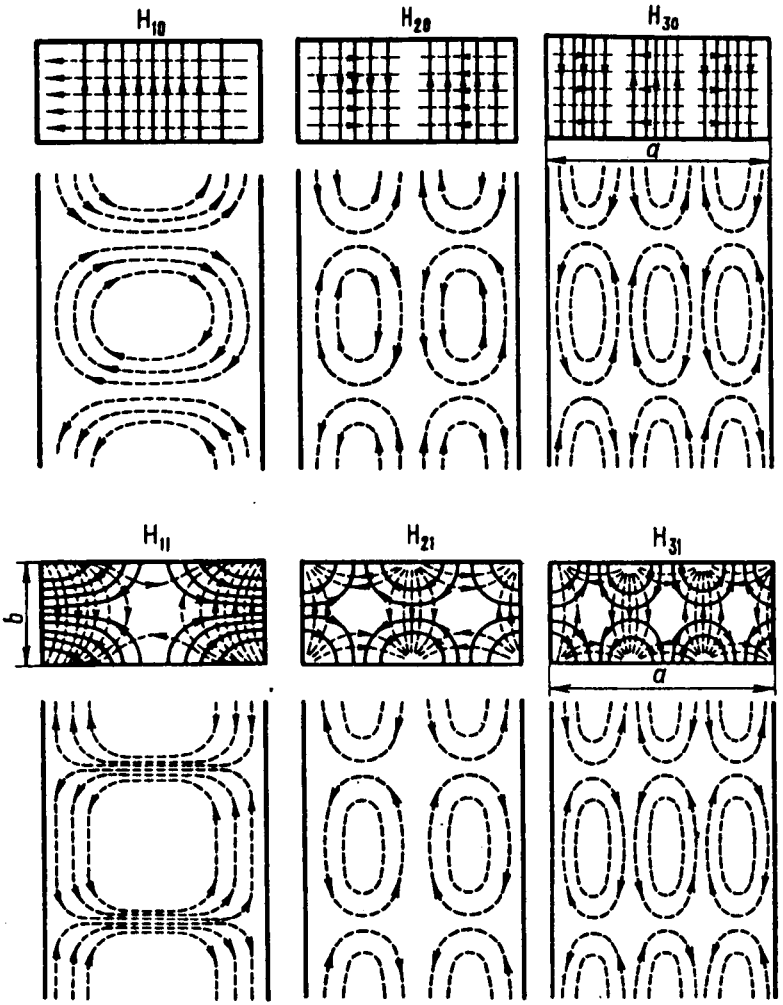


Рис. 3.3. Структура магнитных волн в прямоугольном волноводе

ны желаемого типа. Петля с током должна возбуждать магнитное поле такого направления, которое имеет магнитное поле желаемого типа. Отверстия в стенках волновода должны формировать поверхностные токи с тем же распределением, что и у волн желаемого типа.

Волна типа E_{11} может быть возбуждена в прямоугольном волноводе штырем (рис.3.4, *а*). Если ввести штырь через боковую стенку волновода нормально к оси, то кроме волн типа E_{11} будут возбуждаться и волны

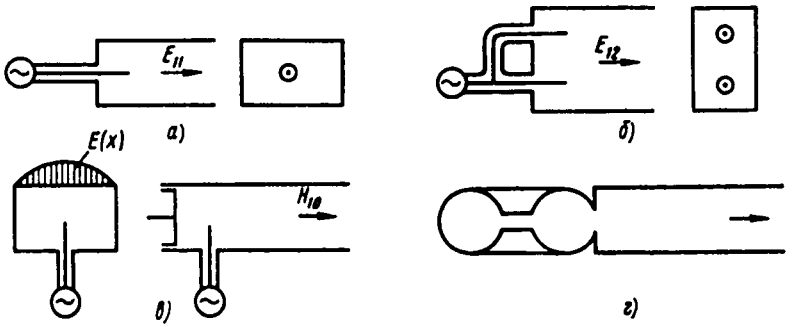


Рис. 3.4. Возбуждение волновода с помощью штыря (*а* ... *в*) и через отверстие в общей с резонатором стенке (*г*)

типа Н. Два штыря (рис.3.4, *б*) образуют волну типа E_{21}, если разветвление коаксиальной линии, связывающей штыри с генератором, обеспечит сдвиг фаз на π между полями этих штырей. Практическое значение имеют в большинстве случаев устройства для возбуждения основного типа волн. В устройстве с электрическим диполем (рис.3.4, *в*) длину штыря и положение поршня, отражающего левую волну в нужную сторону, подбирают так, чтобы обеспечить согласование питающей коаксиальной линии с волноводом. При этом в коаксиальной линии не образуется отраженных волн. Следует иметь в виду, что устройство, изображенное на рис.3.4, *в*, возбуждает кроме волны типа H_{10} множество волн других типов (H_{30}, H_{50}, ..., E_{11}, E_{31}, H_{11}, H_{31}, H_{32} и т. д.). Распространяться по волноводу могут волны только тех типов, для которых критическая длина волны больше длины свободной волны λ. В частности, подбором размеров поперечного сечения можно создать условия для распространения волны только одного типа. Однако оказывается принципиально невозможным в ближней зоне, около возбуждающего устройства исключить образование волн высших типов и обеспечить структуру поля «чистой» волны какого-либо одного типа.

Устройства для отбора энергии от волновода принципиально ничем не отличаются от возбуждающих устройств. Это следует непосредственно из принципа взаимности. Например, на сантиметровых и миллиметровых волнах часто для возбуждения волноводов применяют резонаторы с отверстием в общей стенке (рис. 3.4, *г*). Не менее часто такой вид связи используют и для решения обратной задачи — возбуждения резонатора электромагнитной энергией, поступающей из волновода.

Мощность, передаваемая волноводом,

$$P = \frac{1}{2}\operatorname{Re}\int[\dot{\mathbf{E}}\dot{\mathbf{H}}]\,d\mathbf{S} = \frac{1}{2}\operatorname{Re}\int\underline{\Pi}_z\,dS_{xy} =$$

$$= \frac{1}{2}\operatorname{Re}\int_0^a\int_0^b(\dot{E}_x\dot{H}_y - \dot{E}_y\dot{H}_x)\,dx\,dy = \frac{1}{2Z_н}\int_0^a\int_0^b(E_x^2 + E_y^2)\,dx\,dy. \qquad (3.23)$$

Например, для волны основного типа H_{10} из уравнений (3.21) $E_x = 0$ и

$$E_y = E_m \sin \pi x/a, \qquad (3.24)$$

где $E_m = \omega\mu\pi C/a$.

Тогда согласно уравнениям (3.9), (3.13), (3.23) и (3.24) передаваемая мощность

$$P = \frac{1}{4}ab\frac{E_m^2}{Z_H} = \frac{ab}{4Z_T}\frac{\lambda}{\Lambda}E_m^2. \qquad (3.25)$$

Можно показать, что в общем случае для любого однородного волновода и волны любого типа

$$P = \xi\frac{S}{Z_T}\frac{\lambda}{\Lambda}E_m^2, \qquad (3.26)$$

где S — площадь поперечного сечения волновода; $\xi = \left(\int_V E^2\,dV\right):(E_m^2V)$ — коэффициент, характеризующий степень равномерности заполнения энергией электрического поля волновода в объеме $V = S\Lambda/2$ (для H_{10} $\xi = 0{,}25$).

Уравнение (3.26) часто используют в волноводной технике для: 1) нахождения амплитуды электрического поля, поскольку проходящая мощность поддается непосредственному измерению несравнимо легче, чем градиенты поля; 2) определения предельной мощности, которую может пропустить волновод данного сечения при определенной частоте.

Прямоугольные волноводы получили широкое распространение при передаче больших мощностей для волн короче 0,5 м. Для меньших частот их применение нецелесообразно из-за чрезмерной громоздкости и большой массы при значительном недоиспользовании образующегося запаса электрической прочности.

§ 3.4. Круглые волноводы

Волноводы круглого сечения проще в изготовлении, чем прямоугольные. Однако неустойчивость основного вида колебаний приводит к тому, что в настоящее время короткие отрезки круглых волноводов ис-

пользуют только в качестве вспомогательных и переходных элементов в волноводных передачах (вращающиеся сочленения, преобразователи типа волн и др.). Ведется исследование возможностей использования круглых волноводов для передачи на большие расстояния больших мощностей и для многоканальных линий связи с малым затуханием.

Волновые явления в круглых волноводах по своему характеру ничем не отличается от явлений в прямоугольных волноводах. Однако электромагнитное поле в них образуется в результате распространения и наложения не плоских, а цилиндрических волн. Радиальное и вращательное движения цилиндрических волн создают стоячие волны в радиальном и азимутальных направлениях. Продольное движение цилиндрической волны обеспечивает передачу энергии вдоль оси волновода.

Исходным для расчета поля в круглом волноводе служит волновое уравнение (3.7) в цилиндрической системе координат:

$$\frac{\partial^2 \dot{E}_z}{\partial r^2} + \frac{1}{r}\frac{\partial \dot{E}_z}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \dot{E}_z}{\partial \varphi^2} + \chi^2 \dot{E}_z = 0. \qquad (3.27)$$

Аналогичный вид имеет в этом случае волновое уравнение и для H_z. Их решение с учетом граничных условий позволяет получить структуру поля волн магнитного и электрического типов.

Волны магнитного типа (типа H_{mn}) в цилиндрической системе координат описываются следующей системой уравнений:

$$\dot{E}_r = D\omega \frac{m}{r} J_m\left(\frac{\nu_{mn}}{a} r\right) \sin m\varphi \, e^{-jk_z z};$$

$$\dot{E}_\varphi = D\omega \frac{\nu_{mn}}{a} J'_m\left(\frac{\nu_{mn}}{a} r\right) \cos m\varphi \, e^{-jk_z z};$$

$$\dot{E}_z = 0;$$

$$\dot{H}_r = -D\frac{k_z}{\mu}\frac{\nu_{mn}}{a} J'_m\left(\frac{\nu_{mn}}{a} r\right) \cos m\varphi \, e^{-jk_z z};$$

$$\dot{H}_\varphi = jD\frac{k_z m}{\mu r} J_m\left(\frac{\nu_{mn}}{a} r\right) \sin m\varphi \, e^{-jk_z z};$$

$$\dot{H}_z = -jD\frac{1}{\mu}\left(\frac{\nu_{mn}}{a}\right)^2 J_m\left(\frac{\nu_{mn}}{a} r\right) \cos m\varphi \, e^{-jk_z z}.$$

Экспоненциальный множитель описывает зависимость векторов поля от координаты z, характерную для бегущей волны с коэффициентом фазы k_z и фазовой скоростью $v_\phi = \omega/k_z$. Зависимость от азимутальной координаты выражают тригонометрические функции $\sin m\varphi$ или $\cos m\varphi$,

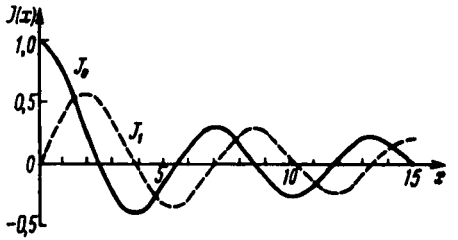


Рис. 3.5. Функции Бесселя первого рода нулевого и первого порядка

причем m равно числу (целое) стоячих волн, угадывающихся вдоль окружности волновода радиусом a.

Зависимость от радиальной координаты r выражается в несколько отличном от прямоугольного волновода виде — в форме функций Бесселя первого рода m-го порядка $J_m(\nu_{mn} r/a)$ и ее производной:

$$J'_m\left(\nu_{mn}\frac{r}{a}\right)=\frac{ma}{\nu_{mn}r}J_m\left(\nu_{mn}\frac{r}{a}\right)-J_{m+1}\left(\nu_{mn}\frac{r}{a}\right).$$

Обе функции — периодические с убывающей амплитудой и несколько изменяющимся периодом вблизи оси волновода (рис.3.5). Убывание функции учитывает уменьшение плотности электромагнитной энергии волны по мере удаления от оси волновода. Аналогом функции $J_0(x)$ является функция $\cos x$, а функции $J'_0(x)$ — функция $\sin(x)$.

Функции Бесселя, как и тригонометрические функции, табулированы и входят в уравнения стоячих волн. Функции Бесселя в данном случае представляют собой математическое описание стоячих радиальных волн в круглом волноводе; параметры $\nu_{mn} = \chi_{mn} a$ — корни первых производных бесселевых функций m-го порядка. Индекс n соответствует номеру корня и определяет число полуволн, укладывающихся вдоль радиуса волновода, поскольку движущаяся к центру волна испытывает на оси как бы полное отражение.

В круглом волноводе критическая длина волны типа Н

$$\lambda_{кр} = 2\pi a/\nu_{mn}.$$

Структура поля некоторых волн типа Н изображена на рис. 3.6. Наибольшую критическую длину $\lambda_{кр} = 3{,}41a$ в круглом волноводе имеет волна типа H_{11}, обладающая структурой, сходной со структурой поля волны H_{10} в прямоугольном волноводе. Волна типа H_{01} не имеет аналога в прямоугольном волноводе. Ее особенностями являются замкнутость силовых линий электрического поля (отсутствие их окончаний на стенках волновода) и малые потери в стенках. Затрудняет использование волны типа H_{01} для передачи энергии то, что эта волна не является низшей.

Волны электрического типа (типа Е) в круглом волноводе имеют следующие компоненты поля:

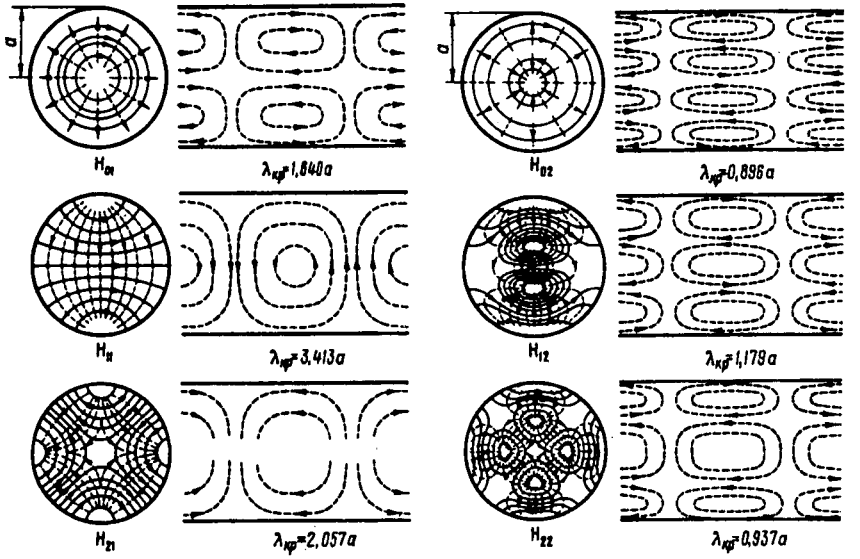


Рис. 3.6. Структура магнитных волн в круглом волноводе

$$\dot{E}_r = D\frac{k_z}{\varepsilon}\frac{\varkappa_{mn}}{a}J'_m\left(\varkappa_{mn}\frac{r}{a}\right)\cos m\varphi\, e^{-jk_z z};$$

$$\dot{E}_\varphi = -D\frac{k_z}{\varepsilon}\frac{m}{r}J_m\left(\varkappa_{mn}\frac{r}{a}\right)\sin m\varphi\, e^{-jk_z z};$$

$$\dot{E}_z = jD\frac{1}{\varepsilon}\left(\frac{\varkappa_{mn}}{a}\right)^2 J_m\left(\varkappa_{mn}\frac{r}{a}\right)\cos m\varphi\, e^{-jk_z z}; \qquad (3.28)$$

$$\dot{H}_r = D\omega\frac{m}{r}J_m\left(\varkappa_{mn}\frac{r}{a}\right)\sin m\varphi\, e^{-jk_z z};$$

$$\dot{H}_\varphi = D\omega\frac{\varkappa_{mn}}{a}J'_m\left(\varkappa_{mn}\frac{r}{a}\right)\cos m\varphi\, e^{-jk_z z};$$

$$\dot{H}_z = 0,$$

где $\varkappa_{mn} = \chi_{mn}a$ — n-й корень функции Бесселя m-го порядка.

Критическая длина волны типа E_{mn}: $\lambda_{кр} = 2\pi a/\varkappa_{mn}$. В круглом волноводе волны типа E_{m0} существовать не могут.

Структура некоторых электрических волн в круглом волноводе изображена на рис. 3.7. Отсутствие вариации поля волны типа E_{01} по азимуту и его полная круговая симметрия находят практическое использование

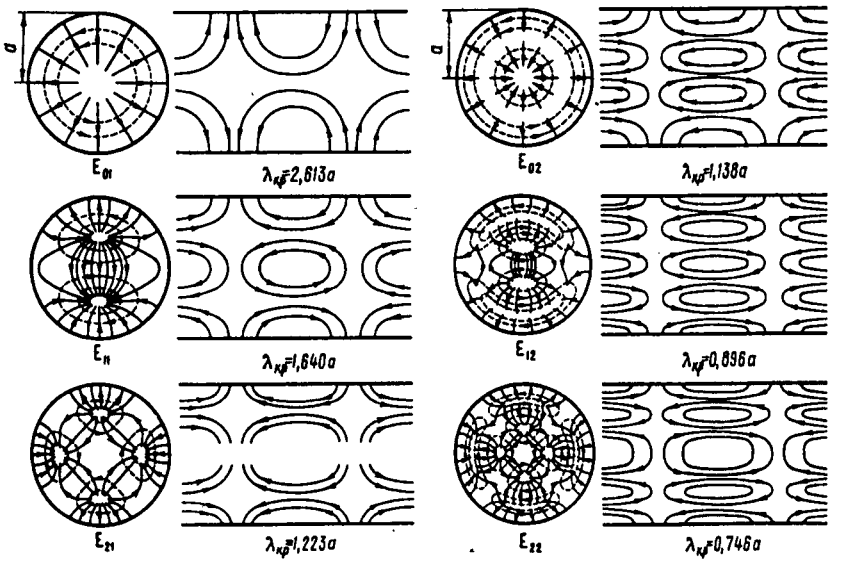


Рис. 3.7. Структура электрических волн в круглом волноводе

во вращающихся соединениях антенн. Этот тип волны также применяют в линейных ускорителях, где основную роль играет продольная составляющая электрического поля бегущей волны.

§ 3.5. Эквивалентное сопротивление волновода

Однородные направляющие системы характеризуются постоянством формы сечения и структуры полей вдоль линии. Условия однородности в реальных волноводных трактах выполняются только на небольших участках, поскольку тракт включает в себя многочисленные устройства и элементы, нарушающие структуру поля в однородном волноводе. К таким элементам относятся, в частности, изгибы, скрутки, изломы (рис.3.8, *а...в*), т. е. места перехода от одной формы и размеров сечения волновода к другим.

Сложность граничных условий в большинстве случаев не дает возможности точно рассчитать поля неоднородностей с учетом волн высших типов и в этом часто нет необходимости, поскольку для определения энергии, передаваемой по линии, достаточно знать только волны основного типа в *дальней зоне* неоднородности, где высшие типы волн отсутствуют. Рассмотрение волновода только как передающей линии позволя-

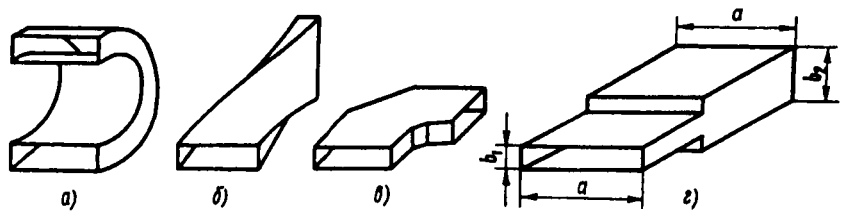


Рис. 3.8. Элементы волнового тракта:
а — изгиб; *б* — скрутка; *в* — излом; *г* — вертикальная ступенька

ет заменить неоднородность эквивалентным сопротивлением или четырехполюсником, создающими такие же отраженную и преломленную волны в дальней зоне, как и действительная неоднородность. Чаще всего параметры эквивалентного четырехполюсника находят более доступным, экспериментальным путем, а не решением уравнений поля.

Определение отраженных волн около неоднородностей требует правильного нахождения эквивалентного сопротивления волновода при согласованной нагрузке. К сожалению, характеристические сопротивления волновода (3.12)—(3.14) не всегда позволяют учесть все особенности волновых явлений. Например, при переходе волны типа H_{10} из прямоугольного волновода с высотой b_1 в прямоугольный волновод высотой b_2 той же ширины a (рис. 3.8, *г*) характеристическое сопротивление остается неизменным, так как критическая длина волны зависит только от ширины. Однако при эксперименте обнаруживаются значительные отражения от неоднородности такого вида, а также то, что стык волноводов не дает отражения при равенстве эквивалентных сопротивлений, подсчитываемых по интегральным параметрам волноводов. Например, в прямоугольном металлическом волноводе эквивалентное сопротивление для волны типа H_{10} может быть подсчитано одним из следующих интегральных способов:

$$z'_{\text{эк}} = \frac{U}{I} = \frac{\pi}{2}\frac{b}{a}\frac{\sqrt{\mu/\varepsilon}}{1-\lambda^2/\lambda_{\text{кр}}^2} = \frac{\pi}{2}\frac{b}{a}z_H;$$

$$z''_{\text{эк}} = \frac{U_m^2}{2P} = 2\frac{b}{a}z_H; \qquad z'''_{\text{эк}} = \frac{2P}{I_m^2} = \frac{\pi^2}{8}\frac{b}{a}z_H,$$

$$U_m = \int_0^b (E_y)_{x=a/2}\,\mathrm{d}y = bE_m; \qquad I_m = \int_0^a H_x\,\mathrm{d}x = \frac{2a}{\pi}H_m.$$

Расхождение результатов, полученных различными интегральными способами, указывает на приближенный характер понятия эквивалентно-

го сопротивления, однако не мешает его практическому использованию, поскольку согласование определяется отношением эквивалентных сопротивлений, а не их абсолютными значениями. Ввиду того, что абсолютное значение числового коэффициента в приведенных выражениях не имеет практического значения, его обычно принимают равным единице, и эквивалентное сопротивление прямоугольного волновода записывают в виде

$$z_{эк} = \frac{b}{a} z_H = \frac{b}{a} \frac{\sqrt{\mu/\varepsilon}}{\sqrt{1-\lambda^2/\lambda_{кр}^2}}. \qquad (3.29)$$

Эквивалентное сопротивление прямоугольного волновода (3.29) необходимо вводить только при сопряжениях волноводов с разными размерами поперечного сечения. В большинстве других случаев можно пользоваться более общим понятием характеристических сопротивлений z_H и z_E.

§ 3.6. Полосковые волноводы

Полосковые волноводы относятся к направляющим системам открытого типа, состоящим из нескольких изолированных друг от друга металлических лент. Широкое применение находят несколько классов полосковых волноводов. К первому из них относятся полосковые линии.

Поперечное сечение полосковых линий различной конструкции показано на рис. 3.9, *а...д*. Несимметричная полосковая линия (см. рис. 3.9, *б*) выполняется путем нанесения слоя металла достаточной ширины с одной стороны подложки и проводника конечной ширины — с другой стороны. Несимметричные полосковые линии конструктивно наиболее просты и технологичны, а при использовании диэлектрика с большим значением диэлектрической проницаемости обладают также и меньшими по сравнению с симметричными полосковыми линиями потерями в диэлектрике. Однако в отличие от полосковых линий, у которых диэлектриком служит воздух (см. рис. 3.9, *а*), они имеют бóльшие потери в диэлектрике и сильный разброс характеристик при изготовлении. К недостаткам несимметричных линий можно отнести жесткость допусков на поперечные размеры металлической ленты, толщину и однородность диэлектрика, а также большие нежелательные связи с другими элементами схемы через поле поверхностной волны.

Несимметричную полосковую линию, использующую диэлектрическую подложку с $\varepsilon_r > 10$, называют *микрополосковой*. Концентрация электрического поля в подложке позволяет в этом случае уменьшить потери на излучение, ослабить паразитные связи и обеспечить миниатюризацию всего устройства.

Симметричные полосковые линии с твердым диэлектриком (см. рис. 3.9, *в*) обладают по сравнению с несимметричными меньшими потерями на излучение при использовании диэлектриков с малой диэлектрической проницаемостью. При ширине внешних пластин $a \geqslant w + (4...5)b$ поле на их краях практически отсутствует. Симметричные полосковые линии с воздушным диэлектриком и токонесущей полоской, нанесенной с двух сторон на опорную диэлектрическую плату (см. рис. 3.9, *г*), позволяют совместить технологические преимущества печатных схем и малые потери волноводов с воздушным заполнением. Симметричные полосковые линии с воздушным заполнением (см. рис. 3.9, *д*), хотя и обладают наименьшими потерями, менее технологичны по сравнению с рассмотренными типами волноводов. Недостатком симметричной линии является затрудненный доступ к внутреннему проводнику.

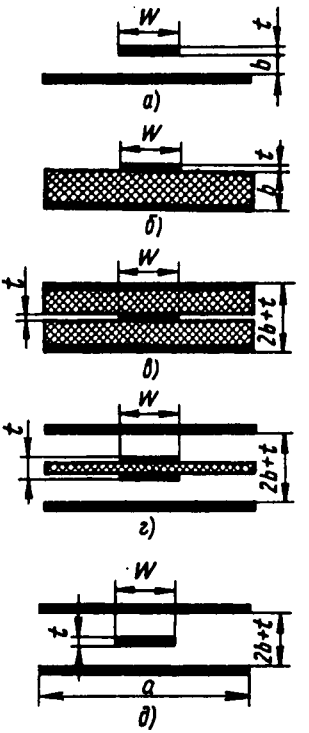


Рис. 3.9. Полосковые линии:

а — несимметричная с воздушным заполнением; *б* — несимметричная с твердым диэлектриком; *в* — симметричная с твердым диэлектриком; *г* — симметричная с воздушным заполнением и токонесущей полоской, нанесенной с двух сторон опорной диэлектрической пластины; *д* — симметричная с воздушным заполнением

При однородном диэлектрическом заполнении полосковой линии основным типом волны является волна типа Т, для которой электрическое поле совпадает со статическим, а магнитное — со стационарным полем в той же линии (рис. 3.10).

В полосковых линиях с неоднородным диэлектрическим заполнением, в частности, в микрополосковых линиях волна типа Т существовать не может. Здесь основной волной является волна гибридного типа, структура поля которой показана на рис. 3.11. Продольные составляющие напряженностей электрического и магнитного полей в этой волне обычно значительно меньше поперечных, поэтому волна получила название волны *квази-Т*.

Волна квази-Т типа обладает дисперсией. Однако со стороны низких частот отсечка отсутствует.

Спектр продольного волнового числа k_z состоит из двух множеств собственных значений: а) конечного множества дискретных собственных значений, соответствующих конечному числу волн различных типов,

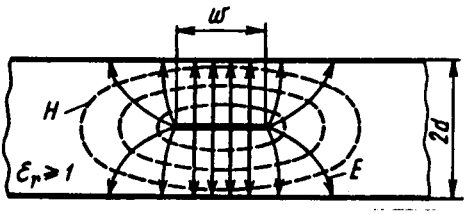


Рис. 3.10. Структура поля Т-волны в симметричной полосковой линии

включающих основной тип волны («дискретный спектр»); б) последовательности собственных значений, соответствующих полю излучения («непрерывный спектр»).

На больших частотах могут возбуждаться паразитные поверхностные волны, что приводит к нежелательным связям между элементами схемы или к паразитным резонансным явлениям при отражениях от корпуса модуля, краев подложки и пр.

Для предотвращения распространения высших типов волн, имеющих вариации поля по ширине полоскового проводника, его ширина должна быть меньше половины длины волны в материале: $w < \lambda_0 K / 2\sqrt{\varepsilon_r}$, где типичная величина коэффициента $K \approx 0{,}9$. Чтобы не допустить возбуждения поверхностных волн, достаточно выбрать толщину подложки микрополосковой линии из условия $b < \lambda_0 / 4\sqrt{\varepsilon_r - 1}$.

Предельная мощность, передаваемая по полосковым линиям, определяется в импульсном режиме электрическим пробоем диэлектрика, а в режиме непрерывных колебаний — нагревом диэлектрика («тепловой пробой»). Строгие расчеты пробивной мощности для полосковых линий отсутствуют. Расчетные и опытные оценки показывают возможность передачи по микрополосковым линиям средних мощностей около 1 кВт на частоте 10 ГГц. На практике передаваемую мощность ограничивают не линии, а локальные концентрации поля на неоднородностях.

К полосковым волноводам относятся также *щелевые линии*. Симметричная щелевая линия представляет собой узкую щель, вырезанную в металлической плоскости, расположенной на одной из сторон плоскопараллельного магнитодиэлектрического слоя, образующего подложку. Основным типом колебаний является гибридный. Однако структура поля волны (рис. 3.12) позволяет рассматривать ее как волну квази-Н.

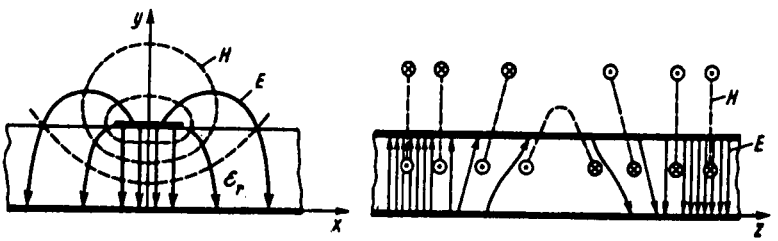


Рис. 3.11. Структура поля квази-Т-волны в микрополосковой линии

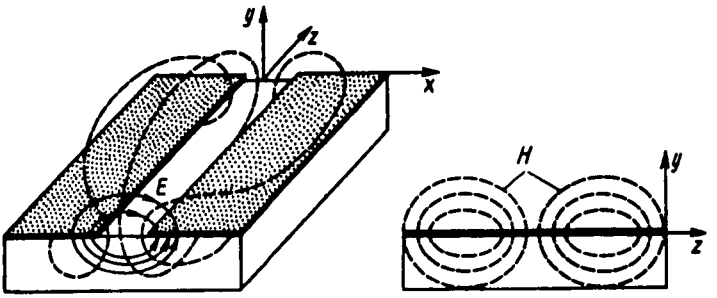


Рис. 3.12. Структура электромагнитного поля симметричной щелевой линии

Компланарная линия образована тремя проводниками, расположенными на одной стороне диэлектрической подложки (рис. 3.13). Распределение поля в зазоре между узким проводником и наружными широкими полосками напоминает распределение поля в щелевой линии. В компланарной линии распространяются волны квази-Т и Н типа.

Кроме рассмотренных конструкций полосковых волноводов имеются различные их модификации. Все виды полосковых волноводов составляют основу элементной базы интегральных сверхвысокочастотных устройств.

§ 3.7. Элементы и узлы микроволноводных устройств

Большая насыщенность электронной аппаратурой производства и научных исследований приводит к необходимости, с одной стороны, повышения скорости передачи информации (емкости канала связи), а с другой — к принятию мер по снижению габаритов и массы аппаратуры. В связи с этим все большее количество электронных устройств производят на основе широкого использования *интегральных схем* (ИС). Развитие электроники идет по пути непрерывного продвижения в область все более коротких волн не только в традиционных направлениях радиотехники, но в микросхемах цифровых и аналоговых ЭВМ. Скорости переключения со-

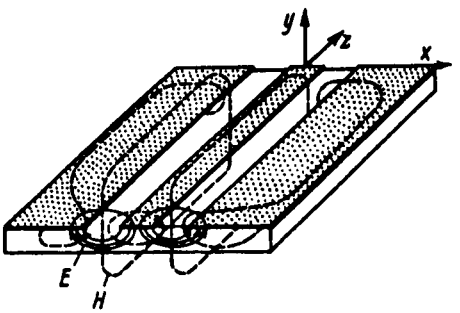


Рис. 3.13. Структура электромагнитного поля четной волны в компланарной линии

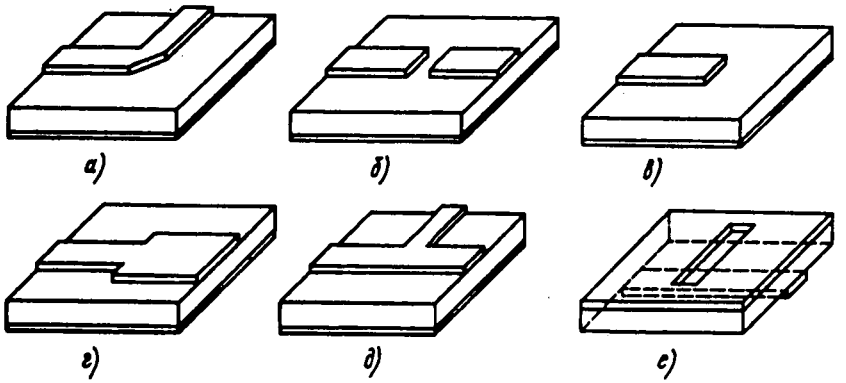


Рис. 3.14. Неоднородности в микрополосковых линиях:

а, б — излом и разрыв полоски; *в, г* — обрыв и скачкообразное изменение ширины полоски; *д* — разветвление; *е* — поперечная щель в заземляющем основании

временных элементов микросхем соответствуют частотам, достигающим 1 ГГц. В близком будущем планируются тактовые частоты 10 ГГц. Возможны и существенно бóльшие частоты: 100 и даже 500 ГГц. Решение проблем интеграции в технике СВЧ приводит не только к значительному улучшению массогабаритных показателей, важнейших свойств и эксплуатационных параметров электронной аппаратуры, ее надежности и воспроизводимости, но и к существенному повышению эффективности производства электронной промышленности и промышленности средств связи, экономии трудовых и материальных ресурсов.

Планарные интегральные схемы состоят из набора неоднородностей, выполняемых на каком-то одном типе полосковых линий. Неоднородности можно разбить на три класса: 1) неоднородности в Н-плоскости; 2) неоднородности в Е-плоскости; 3) неоднородности в компланарных линиях. На рис. 3.14, *а...е* приведены некоторые неоднородности первого класса в микрополосковых линиях, к которым относятся полосковые структуры, при этом токонесущие полоски расположены в плоскости, перпендикулярной вектору **E**.

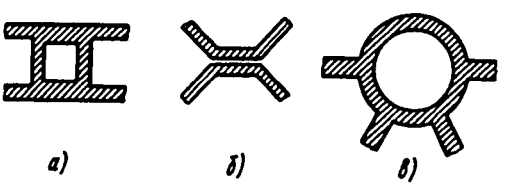


Рис. 3.15. Микрополосковые направленные ответвители:

а — двухшлейфовый; *б* — с распределенной связью; *в* — кольцевой мост

Неоднородности второго класса располагаются в плоскости, перпендикулярной вектору **H**. К ним относятся неоднородности на симметричных щелевых

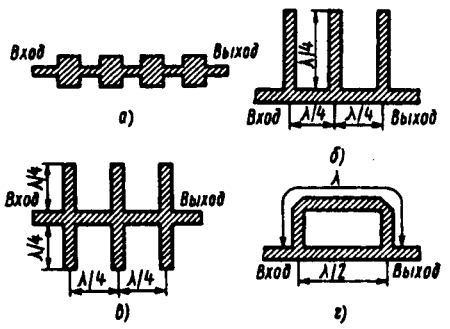


Рис. 3.16. Микрополосковые фильтры:

а — нижних частот; *б* — полосно-пропускающий; *в, г* — шлейфовый и петлевой полосно-заграждающие

линиях: закороченный и разомкнутый отрезки, скачок ширины щели, поворот линии на произвольный угол. Т-соединение и т. д.

Примеры реализации на микрополосковых линиях направленных ответвителей показаны на рис. 3.15, *а*, *б*; кольцевого моста — на рис. 3.15, *в*; фильтров — на рис. 3.16, *а...г*.

Основой создания современных средств приема, передачи и обработки информации на СВЧ стали новые поколения приборов и устройств на твердом теле в интегральном исполнении, в том числе в виде гибридных и полупроводниковых схем. Технология *полупроводниковых интегральных схем* СВЧ совместима с технологией цифровых микросхем ЭВМ и приборов с зарядовой связью, что позволяет разрабатывать многофункциональные устройства. Кроме того, для обеспечения малых потерь в некоторых случаях ИС непосредственно сочленяются с объемными волноводно-резонаторными устройствами.

Качественно новым этапом развития электроники является создание объемных интегральных схем (ОИС). ОИС представляет собой СВЧ-модуль, выполненный на разнотипных регулярных волноведущих линиях, пространственно расположенных в слоях диэлектрика и соединенных между собой емкостной, индуктивной, гальванической или электромагнитной связью. В реализованных устройствах выигрыш в габаритах и массе до трех раз больше, чем в планарных конструкциях. ОИС позволяют решить ряд проблем принципиального характера, которые не под силу планарным ИС, например, оптимизацию функциональных узлов по набору типов линий передач, обеспечивающему наилучшие конструктивно-электрические характеристики всех базовых элементов. Выборки в слоях диэлектриков дают возможность удобной конструктивной реализации ОИС с включением диэлектрических резонаторов и их дифракционно связанных систем, балансных резисторов, электрически управляемых шлейфов и др. ОИС позволяют выносить активные элементы на общую панель, представляющую крайние слои диэлектрика, что обеспечивает удобство охлаждения, экранирования, развязки, замены, подстройки элементов и т. п. Коммутация базовых элементов легко осуществляется с помощью взаимно пересекающихся волноводов с высокой степенью развязки по электромагнитному полю путем расположения элементов в раз-

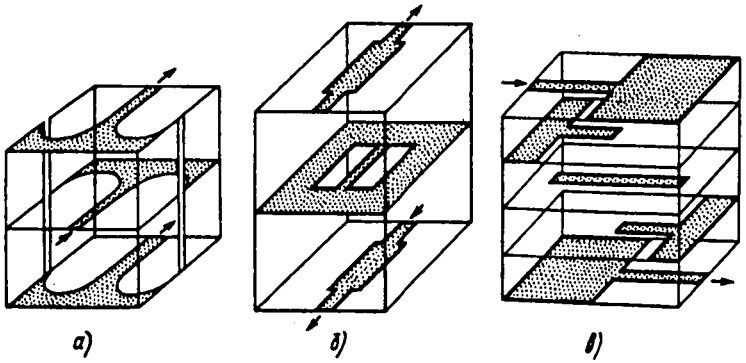


Рис. 3.17. Элементы и узлы объемных интегральных схем:

а — Т-соединение на симметричных полосковых линиях; *б* — направленный ответвитель; *в* — полосовой фильтр

личных слоях диэлектрика. Имеются и непрерывно создаются целые классы новых, присущих только ОИС, передающих линий и базовых элементов. Некоторые элементы и узлы ОИС представлены на рис. 3.17, *а...в*.

Типичный пример ОИС СВЧ приведен на рис. 3.18, где изображен микромодуль, работающий на частоте 26 ГГц. Он представляет собой законченную однокорпусную конструкцию, состоящую из нескольких слоев (слои диэлектриков на рисунке показаны прозрачными). Ортогональные стороны прямоугольных антенн *4* соединены с входными и выходными плечами *3* удвоителей частоты с помощью гальванических переходов через отверстия в слоях диэлектрика и металлизации *2*. Расположение двух независимых удвоителей частоты *1* вместе с цепями фильтрации и полупроводниковыми приборами в среднем слое решает задачу защиты схемы от электромагнитных помех. Выполнение антенны на противоположных гранях ИС позволяет получить диаграмму направленности с круговым обзором в трехмерном пространстве. Маяк отвечает на принимаемый сигнал мощностью 10 мкВт при отсутствии источника питания. Малая масса и высокая надежность свидетельствуют о перспективе использования маяка для поиска жертв снежных лавин.

Так как пока не создан общий метод анализа базовых элементов ИС различного типа на электродинамическом уровне, приходится подбирать в каждой задаче свой математический аппарат. Наиболее часто используется метод частичных областей и принцип декомпозиции, предполагающий предварительное разбиение анализируемой схемы интегрального устройства на ряд составных более простых элементов. Электродинамический анализ простого базового элемента позволяет получить его модель в виде матрицы рассеяния.

Любое электронное устройство, применяемое для обработки информации, можно рассматривать как некоторый преобразователь и трансформатор. В устройствах СВЧ всегда принимаются меры для исключения паразитной связи между базовыми элементами. Поэтому можно сказать, что ОИС есть волноводный трансформатор, который представляет собой структуру со многими входами и выходами, описываемыми матрицей рассеяния. Диагональные коэффициенты матрицы S_{ii} имеют простой физический смысл — они равны комплексным коэффициентам отражения при подаче сигнала со стороны i-го плеча, если в остальных плечах предполагаются включенными согласованные нагрузки, и волны из них не поступают. Недиагональные элементы матрицы S_{ik} равны коэффициентам передачи сигнала в i-е плечо от генератора, включенного в k-е плечо при условии, что ко всем остальным плечам многополюсника подключены согласованные нагрузки.

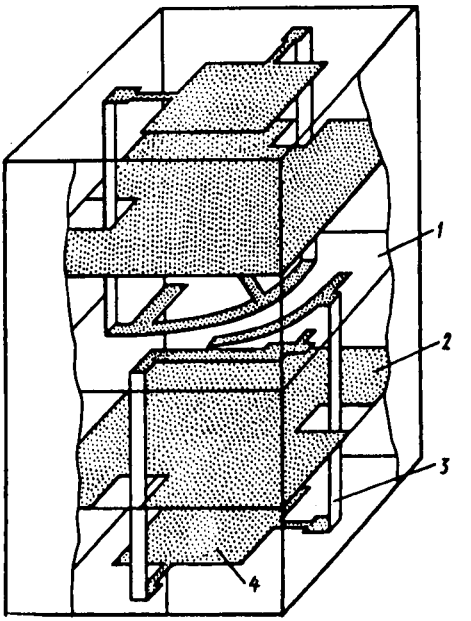


Рис. 3.18. Объемная интегральная схема модуля маяка-ответчика

Построение матрицы рассеяния всего интегрального устройства по набору известных матриц рассеяния базовых элементов проводят достаточно разработанными методами и алгоритмами. Эти методы можно разделить на четыре группы: 1) алгебраическая; 2) топологическая; 3) теоретико-множественная; 4) комбинированная. Первая из них основана на «сворачивании» матриц базовых элементов в общую матрицу. Наиболее рациональный алгоритм вычисления результирующей матрицы рассеяния — поочередное соединение элементов устройства с использованием рекурентной процедуры: матрицу рассеяния $\mathbf{S}$ соединения двух многополюсников рассчитывают по двум матрицам рассеяния — матрице многополюсника $\mathbf{S}'$, полученного на предыдущем этапе, и матрице очередного элемента $\mathbf{S}''$.

Матрицы рассеяния соединяемых элементов делятся на четыре блока

$$\mathbf{S}' = \left[\begin{array}{c|c} \mathbf{S}'_{11} & \mathbf{S}'_{12} \\ \hline \mathbf{S}'_{21} & \mathbf{S}'_{22} \end{array}\right], \quad \mathbf{S}'' = \left[\begin{array}{c|c} \mathbf{S}''_{11} & \mathbf{S}''_{12} \\ \hline \mathbf{S}''_{21} & \mathbf{S}''_{22} \end{array}\right],$$

которые соответствуют разделению входов многополюсников на свободные и участвующие в соединении. Например, $\mathbf{S}'_{11}$ — блок матрицы, который описывает взаимодействие по входам, оставшимся свободными в первом многополюснике; $\mathbf{S}''_{21}$ — блок матрицы, составленный из коэффициентов передачи между свободными входами первого многополюсника и его входами, присоединенными ко второму многополюснику; $\mathbf{S}''_{11}$ — блок матрицы второго многополюсника, соответствующий входам, соединенным с первым многополюсником, и т. д.

Четыре блока матрицы рассеяния соединения многополюсников

$$\mathbf{S} = \left[\begin{array}{c|c} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \hline \mathbf{S}_{21} & \mathbf{S}_{22} \end{array}\right]$$

находят по рекурентным формулам

$$\mathbf{S}_{11} = \mathbf{S}'_{11} + \mathbf{S}'_{12}(\mathbf{E} - \mathbf{S}''_{11}\mathbf{S}'_{22})^{-1}\mathbf{S}''_{11}\mathbf{S}'_{21};$$

$$\mathbf{S}_{12} = \mathbf{S}'_{12}(\mathbf{E} - \mathbf{S}''_{11}\mathbf{S}'_{22})^{-1}\mathbf{S}''_{12};$$

$$\mathbf{S}_{21} = \mathbf{S}''_{21}(\mathbf{E} - \mathbf{S}'_{21}\mathbf{S}''_{11})^{-1}\mathbf{S}'_{21};$$

$$\mathbf{S}_{22} = \mathbf{S}''_{22} + \mathbf{S}''_{21}(\mathbf{E} - \mathbf{S}'_{21}\mathbf{S}''_{22})^{-1}\mathbf{S}'_{22}\mathbf{S}''_{12}.$$

Рассмотренный алгебраический подход достаточно прост и обладает четкой физической интерпретацией. Однако при расчетах с использованием матриц высоких порядков сказывается их громоздкость, и при обращении матриц с неточным знанием элементов снижается точность. В этих условиях предпочтительнее *топологический метод*, в котором в качестве основного понятия используется граф схемы или матрицы. Метод позволяет получить искомую матрицу непосредственно по графу без дополнительных аналитических преобразований. Он наглядно отображает связи между переменными и параметрами моделируемого устройства. При этом требуется хранение в оперативной памяти ЭВМ матриц большого порядка, что заставляет пользоваться в ряде случаев комбинированным подходом, включающим различные группы методов.

Иногда эффективным оказывается и *теоретико-множественный метод*, основанный на отображении модели базового элемента или функционального узла частично упорядоченными множествами цифровых индексов, изоморфных элементам исходной модели, а также на строгой последовательности операций над этими множествами.

Математическое описание базовых элементов и взаимосвязей между ними лежит в основе работы реализованных автоматизированных систем проектирования ИС СВЧ.

§ 3.8. Диэлектрические волноводы

Диэлектрические пластины или стержни, не ограниченные металлическими поверхностями, называют *диэлектрическими волноводами*. Распространение волн по диэлектрическому волноводу обусловлено полным внутренним отражением, происходящим при падении волны на поверхность раздела двух диэлектриков со стороны среды, имеющей более высокую диэлектрическую проницаемость. Следовательно, передача энергии по диэлектрическим волноводам сопровождается поверхностными волнами, напряженность поля которых вне волновода убывает по мере удаления от граничной поверхности.

Если размеры поперечного сечения волновода одного порядка с длиной волны, то основная доля распространяющейся энергии сосредоточена в диэлектрике, а скорость распространения в несколько раз меньше скорости света. При этом волновод может подвергаться изгибам без появления заметного излучения. Затухание за счет тепловых потерь при использовании высококачественных диэлектриков (фторопласт, полиэтилен, полистирол) для длины волны около 1 мм оказывается в несколько раз меньшим, чем в соответствующих стандартных полых металлических волноводах. Стоимость и трудоемкость изготовления диэлектрических волноводов также ниже, чем у полых металлических. Диэлектрические волноводы не обладают электропроводностью, а теплопроводность их очень мала.

Если размеры сечения волновода в несколько раз меньше длины волны, бóльшая часть распространяющейся энергии сосредоточена во внешнем пространстве, а скорость направленной волны отличается от скорости света не более чем на несколько процентов. Затухание при этом в несколько десятков раз меньше, чем в стандартных полых металлических волноводах. Такие волноводы с успехом применяют в качестве линий передачи на значительные расстояния. В области дальнего инфракрасного и светового диапазонов диэлектрический волновод является единственной приемлемой регулярной линией передачи.

Одним из типичных представителей диэлектрических волноводов служит волновод в виде сплошного диэлектрического цилиндра. Для области, находящейся внутри диэлектрического стержня с $\varepsilon = \varepsilon_r \varepsilon_0$ и $\mu = \mu_0$ решение уравнения (3.27) с учетом конечного значения напряженности поля на оси ($r = 0$) имеет вид

$$\dot{E}_z = A_m J(\chi r) e^{jm\varphi} e^{-jk_z z},$$

а вне стержня ($\varepsilon = \varepsilon_0$, $\mu = \mu_0$)

$$\dot{E}_{z0} = C_m K_m(\chi_0 r) e^{jm\varphi} e^{-jk_z z},$$

где K_m — функция Макдональда m-го порядка.

Аналогичный вид имеют выражения для $\dot{H}_z$ и $\dot{H}_{z0}$. Поперечные составляющие электрического и магнитного полей могут быть найдены из (3.6). Приравнивая соответствующие тангенциальные составляющие напряженностей на границе раздела ($r = a$) и исключая из получающейся системы постоянные интегрирования (A_m, C_m и т. д.), можно получить следующее дисперсионное уравнение:

$$\left[\frac{1}{a\chi}\frac{J'_m(\chi a)}{J_m(\chi a)}+\frac{j}{\chi_0 a}\frac{K'_m(\chi_0 a)}{K_m(\chi_0 a)}\right]\left[\frac{\varepsilon_r}{\chi a}\frac{J'_m(\chi a)}{J_m(\chi a)}+\frac{j}{\chi_0 a}\frac{K'_m(\chi_0 a)}{K_m(\chi_0 a)}\right]=$$

$$=m^2\left[\frac{\varepsilon_r}{(\chi a)^2}-\frac{1}{(\chi_0 a)^2}\right]\left[\frac{1}{(\chi a)^2}-\frac{1}{(\chi_0 a)^2}\right]. \qquad (3.30)$$

Решение этого трансцендентного уравнения (численно или графически) позволяет определить продольный коэффициент распространения k_z, связанный с поперечными коэффициентами χ и χ_0 равенством (3.4).

Анализ структуры поля круглого диэлектрического волновода позволяет заключить, что внутри регулярного волновода может существовать бесконечно большое множество типов волн, имеющих различные числа (m и n) изменений направления напряженностей по координатам φ и r. При этом независимо друг от друга могут существовать только симметричные ($m = 0$) волны типов E_{0n} и H_{0n}*. Несимметричные волны здесь не могут быть разделены на электрические и магнитные, они всегда гибридные. Каждый тип волны имеет свою критическую частоту (3.8а). Основным типом волны является волна типа HE_{11}, представляющая собой суперпозицию волн типов H_{11} и E_{11} (см. рис. 3.6 и 3.7). К ее особенностям относится отсутствие критической частоты — она может распространяться на всех частотах и при любом диаметре стержня.

Чем больше a/λ и ε, тем бóльшая часть энергии распространяется внутри диэлектрического стержня. Энергия внутри стержня уменьшается по мере приближения к критической частоте, становясь равной нулю при $f_{кр}$.

В последние годы диэлектрические волноводы все шире применяют в качестве направляющих систем оптического диапазона.

§ 3.9. Пленочные волноводы

При передаче световых сигналов от полупроводникового оптического квантового генератора к фотодиоду целесообразно применение прямо-

* n — номер корня уравнения $J_m(\chi a) = 0$.

угольных световодов, поскольку излучающая зона генератора и приемная площадка фотодиода обычно представляют собой прямоугольники с малой высотой и большой шириной. Расчет и эксперименты показывают, что с помощью плоскопараллельных световодных пластин можно в некоторых случаях получить характеристики оптических систем более высокие, чем при применении линзовой, зеркальной и волоконной оптики.

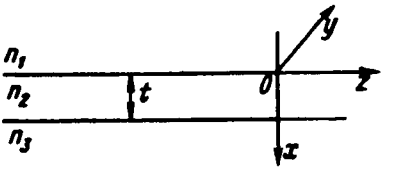


Рис. 3.19. Асимметричный плоский диэлектрический волновод

Плоскопараллельные световодные пластины составляют основу *интегральной оптики*, возникшей как самостоятельная область на стыке классической оптики, электродинамики, экспериментальных методов техники СВЧ, полупроводниковой электроники, материаловедения и квантовой электроники. К интегральной оптике относят теоретические и экспериментальные вопросы распространения света по пленочным световодам, ввода и вывода излучения, генерирования света в субминиатюрных интегрально-оптических лазерных элементах, преобразования излучения нелинейными элементами, в том числе детекторами, модуляторами, умножителями, смесителями и т. п.

Тонкопленочный оптический волновод образует нанесенная на подложку полоска тонкой пленки, показатель преломления которой n_2 больше показателей преломления подложки n_3 и наружного слоя n_1 (рис.3.19). Предполагается, что в направлении оси y волновод имеет бесконечную протяженность, т. е. $\partial/\partial y = 0$.

Для волн магнитного типа $E_z = 0$. В этом случае согласно (3.5),

$$\dot{E}_x = \dot{H}_y = 0;$$

$$\dot{E}_y = j\frac{\omega\mu}{\chi^2}\frac{\partial \dot{H}_z}{\partial x}; \qquad (3.31)$$

$$\dot{H}_x = -j\frac{k_z}{\chi^2}\frac{\partial \dot{H}_z}{\partial x}, \qquad (3.32)$$

т. е. отличны от нуля только три составляющие (E_y, H_x и H_z). Как и во всяком диэлектрическом волноводе, распространение бегущих волн в тонкопленочном оптическом волноводе сопровождается полным отражением от границ и возникновением поверхностной волны. Поэтому зависимость напряженности электрического поля от поперечной координаты можно представить в виде

$$\dot{E}_y(x)=\begin{cases} A\mathrm{e}^{qx}=\dot{E}_y^{(1)} & \text{при } x<0;\\ B\mathrm{e}^{jhx}+C\mathrm{e}^{-jhx}=\dot{E}_y^{(2)} & \text{при } 0\leqslant x\leqslant t,\\ D\mathrm{e}^{-p(x-t)}=\dot{E}_y^{(3)} & \text{при } x\geqslant t. \end{cases}$$

Из второго уравнения Максвелла

$$\dot{H}_z(x)=j\frac{1}{\omega\mu}\frac{\partial\dot{E}_y(x)}{\partial x}=\begin{cases} j\dfrac{qA}{\omega\mu}\mathrm{e}^{qx}=\dot{H}_z^{(1)};\\ j\dfrac{h}{\omega\mu}(B\mathrm{e}^{jhx}-C\mathrm{e}^{-jhx})=\dot{H}_z^{(3)};\\ -j\dfrac{pD}{\omega\mu}\mathrm{e}^{-p(x-t)}=\dot{H}_z^{(3)}. \end{cases}$$

Из граничных условий при $x=0$ ($\dot{E}_y^{(1)}=\dot{E}_y^{(2)}$, $\dot{H}_z^{(1)}=H_z^{(2)}$) и $x=t$ ($\dot{E}_y^{(2)}=\dot{E}_y^{(3)}$, $\dot{H}_z^{(2)}=\dot{H}_z^{(3)}$) следуют четыре соотношения:

$$A=B+C; \tag{3.33}$$

$$D=B\mathrm{e}^{jht}+C\mathrm{e}^{-jht}; \tag{3.34}$$

$$qA=jhB-jhC; \tag{3.35}$$

$$-pD=jhB\mathrm{e}^{jht}-jhC\mathrm{e}^{-jht}, \tag{3.36}$$

которые определяют постоянные A, B, C и D. Напряженность электрического поля выражается через амплитудную постоянную A следующим образом:

$$\dot{E}_y^{(1)}=A\mathrm{e}^{qx}; \tag{3.37}$$

$$\dot{E}_y^{(2)}=A\frac{q}{h}\left(\sin hx+\frac{h}{q}\cos hx\right); \tag{3.38}$$

$$\dot{E}_y^{(3)}=A\frac{q}{h}\left(\sin hx+\frac{h}{q}\cos hx\right)^{-p(x-t)} \tag{3.39}$$

Если приравнять нулю определитель системы уравнений (3.33)—(3.36), то получим характеристическое уравнение

$$\begin{vmatrix} -q & jh & -jh & 0\\ -1 & 1 & 1 & 0\\ 0 & jh\mathrm{e}^{jht} & -jh\mathrm{e}^{-jht} & p\\ 0 & \mathrm{e}^{jht} & \mathrm{e}^{-jht} & -1 \end{vmatrix}=0,$$

из которого

$$ht = \operatorname{arctg}(p/h) + \operatorname{arctg}(q/h) + m\pi, \tag{3.40}$$

где $m = 0, 1, 2, 3, \ldots$ — характеристические числа, определяющие множество решений (мод) характеристического уравнения. Дополнительные соотношения можно найти, если подставить выражения (3.37)—(3.39) в волновое уравнение (3.2):

$$q^2 + (2\pi/\lambda)^2 n_1^2 = -\gamma^2; \tag{3.41}$$

$$-h^2 + (2\pi/\lambda)^2 n_2^2 = -\gamma^2; \tag{3.42}$$

$$p^2 + (2\pi/\lambda)^2 n_3^2 = -\gamma^2. \tag{3.43}$$

После исключения γ получим

$$q^2 + h^2 = (2\pi/\lambda)^2 (n_2^2 - n_1^2), \tag{3.44}$$

$$p^2 + h^2 = (2\pi/\lambda)^2 (n_2^2 - n_3^2). \tag{3.45}$$

Соотношения (3.40), (3.44) и (3.45) образуют систему уравнений, связывающую p, h и q и определяющую дисперсию мод (зависимость β от ω). Связь каждого из параметров h, p и q с частотой и постоянной распространения следует из уравнений (3.41)—(3.43). Если все три среды прозрачны и потерями можно пренебречь, то $\gamma = j\beta$. В этом случае показатели преломления — действительные величины и уравнения (3.41)—(3.43) примут вид

$$q^2 = \beta^2 - k_1^2; \tag{3.46}$$

$$p^2 = \beta^2 - k_3^2; \tag{3.47}$$

$$h^2 = k_2^2 - \beta^2, \tag{3.48}$$

где

$$k_i = 2\pi n_i/\lambda = \omega n_i/c_0.$$

Подставив значения p, q и h из этих выражений в уравнение (3.40), получим дисперсионное уравнение

$$t\sqrt{k_2^2 - \beta^2} = \operatorname{arctg}\sqrt{\frac{\beta^2 - k_1^2}{k_2^2 - \beta^2}} + \operatorname{arctg}\sqrt{\frac{\beta^2 - k_3^2}{k_3^2 - \beta^2}} + m\pi. \tag{3.49}$$

Поскольку в средах без потерь постоянные p, q и h должны быть действительными величинами, то, согласно (3.46)—(3.48), для каждой моды допустимые значения ω и β ограничиваются неравенствами

$$\beta > \omega n_1/c_0, \quad \beta > \omega n_3/c_0, \quad \beta < \omega n_2/c_0. \tag{3.50}$$

На рис.3.20 штриховкой выделена область *3* существования волноводных мод, в которой $n_1 < n_3 < n_2$. Отмеченное ограничение означает, что фазовая скорость мод должна быть не больше фазовой скорости пло-

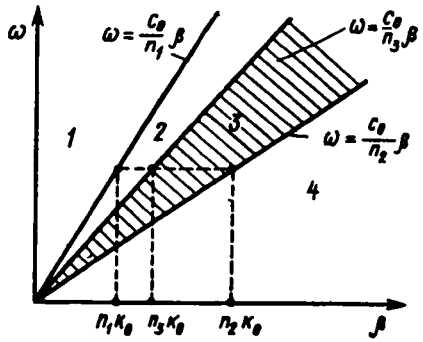


Рис. 3.20. Границы дисперсии мод пленочного волновода

ских волн, распространяющихся в материале подложки с показателем преломления n_3, и не меньше скорости плоских волн в волноводном слое с показателем преломления n_2.

Дисперсионные кривые для первых десяти мод приведены на рис. 3.21 при $n_1 = 1{,}0$; $n_2 = 2{,}23$; $n_3 = 1{,}393$ и $t = 3{,}34$ мкм. Их ход согласуется с тем, что по мере увеличения частоты глубина проникновения поля в области *1* и *3* на рис. 3.20 уменьшается (p и q возрастают). Поэтому поле все в большей степени сосредоточивается в области с показателем преломления n_2, а в предельном случае ($\omega \to \infty$) поле ограничено только волноводным слоем и распространяется в виде плоских волн в среде с показателем преломления n_2. При этом все дисперсионные кривые асимптотически приближаются к нижней границе, наклон которой определяется скоростью распространения света в волноводном слое.

По мере уменьшения частоты экспоненциально спадающее поле поверхностной волны все в большей степени простирается вне волновода (p и q уменьшаются) и при $\omega/c_0 \to k_3$ поле в подложке простирается до бесконечности ($p \to 0$). Поэтому наклон верхней границы волноводных мод (см. рис. 3.21) определяется скоростью распространения света в подложке с показателем преломления n_3.

Области *1* и *2* (см. рис. 3.20) за граничными прямыми ($\omega/c_0 = \beta/n_3$, $\omega/c_0 = \beta/n_2$) соответствуют непрерывному спектру *излучательных мод*, причем области *4* ниже нижней граничной прямой моды называют *затухающими* (β принимает мнимые значения).

Дисперсные кривые для волн электрического

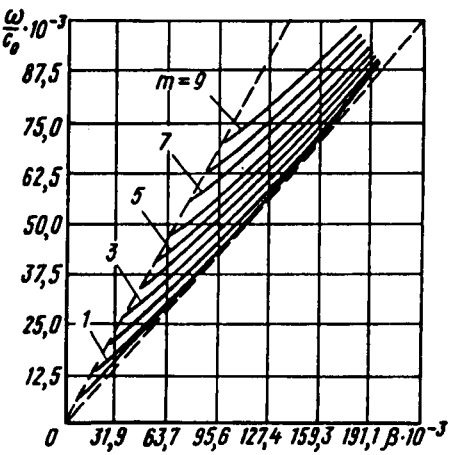


Рис. 3.21. Дисперсионные кривые волноводных мод

типа качественно подобны кривым волн типа Н. Они подчиняются уравнению

$$t\sqrt{k^2-\beta^2}=\operatorname{arctg}\left(\frac{n_2^2}{n_1^2}\sqrt{\frac{\beta^2-k_1^2}{k_2^2-\beta^2}}\right)+\operatorname{arctg}\left(\frac{n_2^2}{n_3^2}\sqrt{\frac{\beta^2-k_3^2}{k_2^2-\beta^2}}\right)+m\pi. \quad (3.51)$$

Решение волновых уравнений для волн типа Е приводит к следующим значениям их компонент:

$$\dot{H}_y^{(1)}=Ae^{\varphi x};$$

$$\dot{H}_y^{(2)}=A\frac{q}{h}\frac{n_2^2}{n_1^2}\left[\sin hx+\frac{h}{q}\left(\frac{n_1^2}{n_2^2}\right)\cos hx\right];$$

$$\dot{H}_y^{(3)}=A\frac{q}{h}\frac{n_2^2}{n_1^2}\left[\sin ht+\frac{h}{q}\left(\frac{n_1^2}{n_2^2}\right)\cos ht\right]e^{-p(x-t)}; \quad (3.52)$$

$$\dot{E}_z=\frac{1}{\omega n^2\varepsilon_0 t}\frac{\partial \dot{H}_y}{\partial x};\quad \dot{E}_x=\frac{\gamma}{j\omega n^2\varepsilon_0}\dot{H}_y.$$

Реальные оптические волноводы обычно не имеют скачка показателя преломления на границе раздела диэлектриков даже в том случае, когда исходные материалы однородны. Это связано с диффузией материалов при нагреве или ионной имплантацией в процессе изготовления. Изменение показателя преломления в направлении, перпендикулярном оси волновода, характеризуют функцией, которую называют *профилем показателя преломления*. Если показатель преломления изменяется плавно, то такие волноводы называют *градиентными*. Они позволяют существенно уменьшить искажение формы сигналов и повысить скорость передачи информации по волноводу по сравнению с волноводом со ступенчатым профилем изменения показателя преломления.

Рассмотрим симметричное изменение показателя преломления

$$n^2(x)=n^2(0)\left(1-2\Delta\frac{x^2}{a^2}\right). \quad (3.53)$$

Это выражение описывает изменение диэлектрической проницаемости по параболическому закону. Однако при не слишком больших значениях величины x, когда $2\Delta(x/a)^2 \ll 1$, параболическому закону подчиняется также и показатель преломления:

$$n(x)\approx n(0)\left(1-\Delta\frac{x^2}{a^2}\right). \quad (3.54)$$

Для *параболического распределения* (3.54) в среднем слое толщиной $2a$, окруженном прилегающими слоями с однородным распределением показателя преломления $n_2=n_3$, согласно условию $n(a)=n_2$

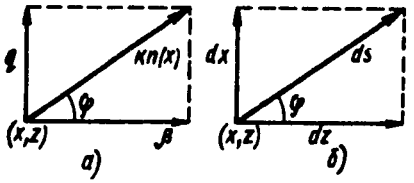


Рис. 3.22. Компоненты локального волнового вектора (*а*) и элемента длины луча (*б*) в плоскости распространения

$$\Delta = (n_1^2 - n_2^2)/n_1^2, \qquad (3.55)$$

где $n_1 = n(0)$.

В слабонаправляющих волноводах $\Delta \approx (n_1 - n_2)/n_1$. В этом приближении параметр (3.55) является относительным приращением показателя преломления при переходе от сердцевины к оболочке плоского волновода.

Рассмотрим траекторию светового луча, предполагая, что все линейные размеры волновода существенно превышают длину волны, и показатель преломления изменяется достаточно медленно в поперечном направлении x. Тогда локальное направление луча можно задать волновым вектором $kn(x)$. Обозначим его продольную компоненту (вдоль z) через β. Тогда его поперечная компонента (вдоль x) $q^2 = k^2 n^2(x) - \beta^2$. Для элемента длины луча $\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}z^2$. Сравнив рис. 3.22, *а* и *б*, получим

$$\frac{\mathrm{d}s}{\mathrm{d}x} = \frac{kn(x)}{\sqrt{k^2 n^2(x) - \beta^2}}; \qquad \frac{\mathrm{d}z}{\mathrm{d}x} = \frac{\beta}{\sqrt{k^2 n^2(x) - \beta^2}}. \qquad (3.56)$$

Приращению $\mathrm{d}x = 0$ в некоторой точке x соответствует $q = 0$, когда $\beta = kn(x)$. Этим значениям отвечают точки поворота луча («каустики»), между которыми параметр q принимает действительные значения, соответствующие существованию распространяющегося луча.

После подстановки показателя преломления (3.53) в уравнение (3.56), оно примет вид

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \frac{\beta}{\sqrt{k^2 n_1^2 - k^2 n_1^2 2\Delta (x/a)^2 - \beta^2}} = \frac{a\beta}{\sqrt{u^2 - v^2 (x/a)^2}}, \qquad (3.57)$$

где $v = akn_1\sqrt{2\Delta}$; $u = a\sqrt{k^2 n_1^2 - \beta^2}$ — *нормированные частоты*. Если исходить из граничного условия $x = x_0$ при $z = 0$, то интегрирование (3.56) приведет к выражению

$$\frac{zv}{a^2 v} = \arcsin\frac{vx}{ua} - \arcsin\frac{vx_0}{ua}, \qquad (3.58)$$

которое можно преобразовать к более удобному виду

$$x = x_0 \cos\frac{vz}{a^2\beta} + \frac{a^2\beta}{v}\sin\frac{vz}{a^2\beta}, \qquad (3.59)$$

причем

$$\operatorname{tg}\varphi_0 = \frac{1}{a\beta}\sqrt{u^2 - v^2\left(\frac{x_0}{a}\right)^2} \qquad (3.60)$$

— начальный наклон луча $\mathrm{d}x/\mathrm{d}z$ при $x = x_0,\ z = 0$.

Из выражения для смещения луча (3.59) как функции расстояния вдоль оси z, а также (3.58) следует, что траектория луча имеет периодический характер с периодом $L = 2\pi a^2\beta/v$ и максимальной амплитудой осцилляции ua/v.

В параксиальном приближении, т. е. для малых углов φ на рис. 3.22 и $\beta \approx kn_1$ выражение (3.59) преобразуется к виду

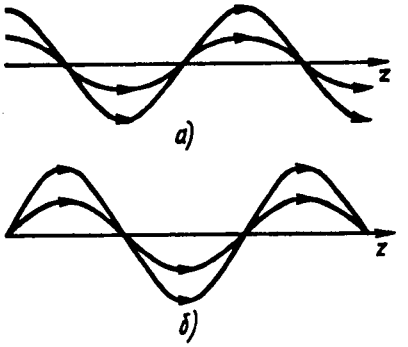


Рис. 3.23. Траектории параксиальных лучей в среде с параболическим профилем:

а — лучи, параллельные на входе ($\varphi_0 = 0$); *б* — лучи, пересекающие на входе ось волновода ($x_0 = 0$)

$$x \approx x_0 \cos\frac{z}{a}\sqrt{2\Delta} + \frac{a\varphi_0}{\sqrt{2\Delta}}\sin\frac{z}{a}\sqrt{2\Delta}. \qquad (3.61)$$

Из этого выражения следует, что все лучи имеют одинаковый период осцилляции (рис. 3.23), в результате чего волновод приобретает свойства линзы.

Эффект фокусировки в среде с параболическим профилем приводит к тому, что на расстояниях, кратных периоду осцилляции, формируются исходные изображения. Действительно, после прохождения волновода длиной L лучи, входящие в волновод параллельно оси z ($\varphi_0 = 0$), выходят из волновода согласно (3.59) с поперечным смещением $x = x_0 \cos\frac{L}{a}\sqrt{2\Delta}$ и наклоном (в параксиальном приближении) $\rho \approx -\frac{n_1 x_0}{n_0 a}\sqrt{2\Delta}\sin\frac{L}{a}\sqrt{2\Delta}$, где n_0 — показатель преломления среды вне волновода (рис.3.24).

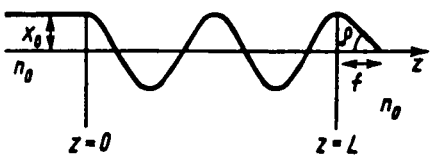


Рис. 3.24. Траектория луча в волноводе, эквивалентном линзе

Фокусное расстояние

$$f \approx \left|\frac{x_0}{\rho}\right| \approx \frac{n_0 a}{n_1\sqrt{2\Delta}}\operatorname{ctg}\frac{L}{a}\sqrt{2\Delta}$$

оказалось не зависящим от начального смещения x_0 входного луча. Поэтому среда с параболическим профилем для всех лучей

с небольшим смещением относительно оси световода способна работать как настоящая линза.

Волноводы с такими свойствами получили название *сельфоков*. Фокусирующие свойства сельфоков не зависят от длины волн. Поэтому их можно использовать и для передачи цветного изображения (без нарушения распределения цветов на выходе).

Путь, проходимый лучом вдоль оси волновода за полупериод от $x_1 = -ua/v$ до $x_2 = ua/v$, можно найти путем интегрирования (3.57):

$$\int_{z_1}^{z_2} \mathrm{d}z = \pi \frac{a^2 \beta}{v}.$$

Оптическая длина луча на этом участке

$$\int_{s_1}^{s_2} n(x)\,\mathrm{d}s = \int_{-ua/v}^{ua/v} \frac{kn^2(x)\,\mathrm{d}x}{\sqrt{k^2 n^2(x) - \beta^2}} =$$

$$= \int_{-ua/v}^{ua/v} akn_1^2 \frac{1 - 2\Delta(x/a)^2}{\sqrt{u^2 - v^2(x/a)^2}} \mathrm{d}x = \frac{\pi a^2 n_1^2}{v}\left(1 - \Delta\frac{u^2}{v^2}\right).$$

Следовательно, время распространения луча

$$\tau = \frac{L}{c} \frac{\int_{s_1}^{s_2} n(x)\,\mathrm{d}s}{\int_{z_1}^{z_2} \mathrm{d}z} = \frac{L}{c}\frac{kn_1^2}{\beta}\left(1 - \Delta\frac{u^2}{v^2}\right), \qquad (3.62)$$

если период луча много меньше L и не учитывается дисперсия показателя преломления. Таким образом, в параксиальном приближении ($\beta \approx kn_1$) время распространения (3.62) оказывается одинаковым и равным Ln_1/c для всех лучей, несмотря на их разные геометрические длины (см. рис. 3.23). Это связано с тем, что лучи с большим отклонением от оси волновода распространяются в среде с меньшим показателем отклонения, т. е. быстрее лучей с меньшим поперечным сечением. Незначительное искажение формы, расширение фронтов импульсов и отсутствие взаимодействия световых лучей, распространяющихся в сельфоке по разным путям, позволяет увеличивать длину волоконно-оптической линии и осуществлять уплотнение каналов линии связи.

Расчет поля в градиентных волноводах базируется на уравнениях Максвелла

$$\operatorname{rot}\mathbf{H} = n^2(x)\varepsilon_0\frac{\partial \mathbf{E}}{\partial t};\ \ \operatorname{div}\left[n^2(x)\varepsilon_0\mathbf{E}\right] = 0;$$

$$\operatorname{rot}\mathbf{E} = -\mu\frac{\partial \mathbf{H}}{\partial t};\ \ \operatorname{div}\mathbf{H} = 0.$$

Решив систему этих уравнений для плоского волновода с параболическим распределением (3.53), для магнитной волны получим

$$\frac{\mathrm{d}^2 H_x}{\mathrm{d}x^2} + \left[k^2 n_1^2\left(1 - 2\Delta x^2/a^2\right) - \beta^2\right]H_x = 0.$$

Решение этого уравнения имеет вид

$$H_x = \frac{2^{1/4}}{\pi^{1/4}(2^N N! w_0)^{1/2}} H_N\left(\frac{x\sqrt{2}}{w_0}\right)\mathrm{e}^{-x^2/w_0}, \qquad (3.63)$$

где $N = 0, 1, 2, \ldots$ — индекс (номер) моды; $w_0 = \dfrac{2a}{kn_1\sqrt{2\Delta}} = \dfrac{2a^2}{v}$ — ширина гауссова пучка; H_N — функции Гаусса-Эрмита, первые из которых задаются выражениями $H_0(y) = 1;\ H_1(y) = 2y;\ H_2(y) = 4y^2 - 2$.

Постоянная интегрирования в (3.63) получена из условия нормировки $\int\limits_{-\infty}^{\infty} |H_x|^2\,\mathrm{d}x = 1$.

Характеристическое уравнение имеет вид $2N = (k^2 n_1^2 - \beta^2)w_0^2/2 - 1$, т. е. нормированная частота $u^2 = v(2N+1)$ и нормированная постоянная распространения

$$b = 1 - \frac{u^2}{v^2} = 1 - \frac{2N+1}{v}. \qquad (3.64)$$

Результаты расчета по формулам (3.63) и (3.64) приведены на рис. 3.25 и 3.26.

Равенство нулю нормированной постоянной распространения $b\,(\beta = kn_2)$ при $v = 2N + 1 = v_{\text{кр}}$ не определяет режим отсечки в обычном смысле, поскольку исходное распределение показателя преломления (3.53) не ограничено в поперечном направлении и поле сохраняет свойства, характерные для направляемых волн при всех значениях v.

Распределение напряженности E_x в электрической волне описывается выражением, отличающимся в общем случае от (3.63) только постоянным множителем. При $2\Delta \ll k^2 n_1^2 a^2$ эти распределения совпадают. В этом приближении дисперсионные уравнения и ширина гауссовского

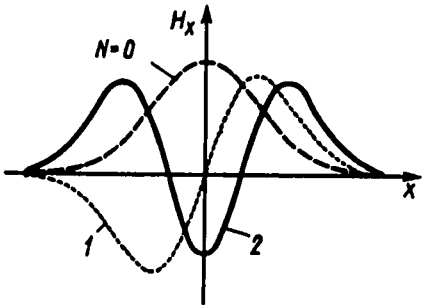


Рис. 3.25. Распределение поля магнитной волны для первых трех типов колебаний в волноводе с параболическим профилем

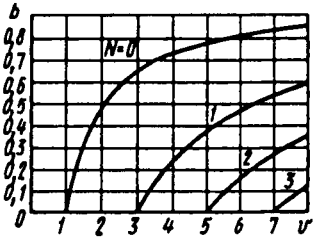


Рис. 3.26. Дисперсионные характеристики низших типов магнитных волн в волноводе с параболическим профилем

пучка для волн типа E и H практически совпадают. Поэтому E- и H-моды оказываются квазивырожденными. Расчеты показывают, что для H- и E-мод групповые задержки сигналов практически не зависят от номера моды и, кроме нескольких мод низшего порядка, почти одинаковы.

Рассмотрение волн и лучей в среде с параболическим профилем позволило познакомиться с основными свойствами и достоинствами градиентных волноводов. Необходимо, однако, учитывать, что параболическое распределение показателя преломления (3.54) применимо только для симметричных волноводов и его нельзя использовать в режиме отсечки.

В рассмотренных пленочных волноводах планарного типа ограничение светового канала имело место только в одном направлении. Эта особенность затрудняет применение таких волноводов в интегральной оптике, поскольку ухудшает согласование с другими оптическими компонентами и световодами (в частности, с волоконными), осложняет локализацию оптических связей, снижает эффективность волноводных устройств. Поэтому в интегральной оптике получили основное распространение *трехмерные волноводные структуры*, в которых ограничение светового канала имеет место в обоих направлениях поперечного сечения.

Трехмерные структуры пленочных волноводов условно можно разделить на два типа: канальные (рис. 3.27, *а*) и полосковые (рис. 3.27, *б*). Значительный скачок показателя преломления в поперечном направлении в канальных волноводах приводит к многомодовому режиму.

Полосковый волновод представляет собой тонкопленочный или диффузионный волновод с узкой диэлектрической областью, нанесенной на его поверхность. В таких волноводах легко реализуется одномодовый и маломодовый режим работы. К их достоинствам также относятся слабая чувствительность к нарушениям и неровностям границ волноводного слоя и увеличение допуска на его формат в одномодовом режиме. В ин-

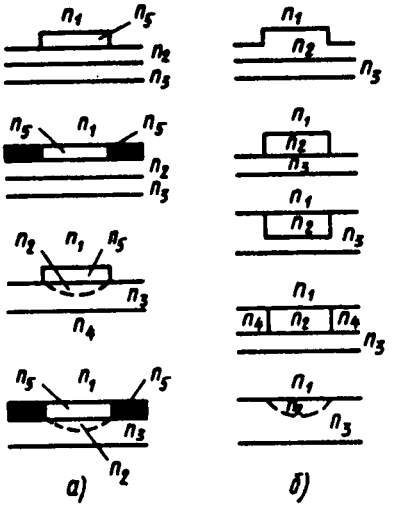


Рис. 3.27. Трехмерные волноводы

тегрально-оптических схемах ширина полоски может достигать 10 мкм и более, что упрощает стыковку волноводов.

Обычные маломодовые полосковые волноводы с однородным, ступенчатым профилем показателя преломления при некоторых условиях оказываются пригодны для передачи изображений точно так же, как линзоподобные волноводы с параболическим профилем показателя преломления (сельфоки). Действительно, если на входном торце симметричного планарного волновода ($z = 0$) когерентным монохроматическим светом сформировано изображение с некоторым распределением амплитуд, то это распределение можно представить в виде разложения $E_0(x) = \sum_m a_m F_m(x)$ по волновым функциям волновода $F_m(x)$ со своими амплитудными коэффициентами a_m. В произвольном сечении ($z > 0$) поле

$$\dot{E}(x) = e^{-j\beta_0 z} \sum_m a_m F_m(x) e^{-j(\beta_m - \beta_0)z},$$

где β_m — коэффициент фазы моды. Прямое изображение воспроизводится в сечении $z = L$ для четных мод m, у которых $e^{j(\beta_m - \beta_0)L} = (-1)^m$. Для нечетных мод воспроизводится перевернутое изображение. Если выразить коэффициенты фазы через параметры волновода и длину волны в вакууме λ, то последнее условие примет вид

$$L_m = 4mn_2 w_{эф}^2 / \lambda,$$

где

$$w_{эф} = w + \frac{\lambda}{\pi}\left(\frac{n_1}{n_2}\right)^{2\sigma} \left(n_2^2 - n_1^2\right)^{-1/2}$$

— эффективная ширина волновода с шириной w ($n_2 > n_1 = n_3$). Для волн типа H $\sigma = 0$; для волн типа E $\sigma = 1$.

Излучение, распространяющееся по планарному или волоконному световодам, может быть использовано для записи и восстановления голо-

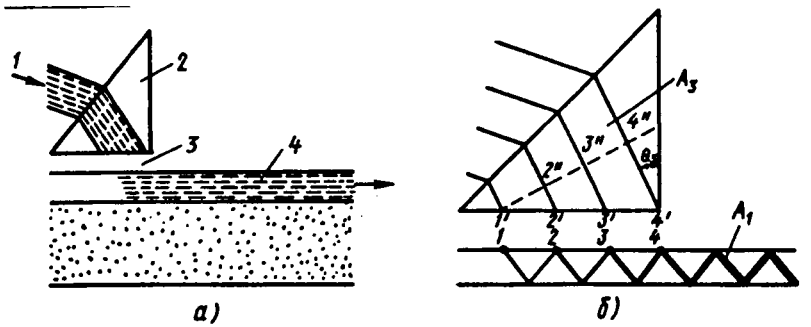


Рис. 3.28. Призменный ввод излучения в пленку

грамм. Подобная *волноводная голограмма* единственная из видов голограмм, совместимая с интегрально-оптическими устройствами.

Волноводная голограмма может быть сформирована двумя пересекающимися поверхностными волнами в слое фоторезиста, в фотографическом слое, нанесенном на поверхности волновода, либо в объеме волновода, если он обладает светочувствительностью.

Волноводная голограмма может быть записана при помощи поверхностной опорной и объемной предметной волн. Для записи используют также «зону связи» в области нарушенного полного внутреннего отражения на границе светочувствительного волновода с призмой связи, хотя призма вносит дополнительные искажения и шумы.

Методы возбуждения пленочных волноводов носят специфический характер. Одним из основных является метод *призменного ввода* лазерного излучения в пленку. Призму *2* помещают при этом сверху на пленку *4* таким образом, чтобы между ними оставался небольшой зазор *3* — порядка нескольких долей длины волны света (рис. 3.28, *а*). Лазерный луч *1*, испытывая от основания призмы полное внутреннее отражение, создает под ним поле бегущей волны, спадающее по экспоненте. Хотя спадающее поле не передает энергию в направлении оси x в отсутствие волновода, оно может быть использовано для передачи энергии из призмы в пленку и обратно.

Принцип действия призменного ввода излучения можно наглядно объяснить, предполагая, что поле в волноводе образовано зигзагообразной волной, испытывающей многократные полные внутренние отражения от границ пленки. На рис.3.28, *б* изображен зигзагообразный луч, соответствующий некоторой волноводной моде. Цифрами *1*, *2*, *3* и *4* обозначены точки поворота луча на верхней поверхности пленки. В относительно широком пучке A_3 лазерного излучения выделены четыре параллельных луча *1′*, *2′*, *3′* и *4′*, падающие точно напротив точек *1*, *2*, *3* и *4*. Световой импульс лазера одновременно достигает волновой фронт в точ-

ках *1'*, *2''*, *3''* и *4''*. В этот момент луч *1'* возбуждает в пленке в точке *1* волну A_1. Если волна A_1 достигнет точки *2* точно в тот момент, когда луч *2'* достигнет воздушного зазора *3* (см. рис. 3.28, *а*), то луч *2'* возбудит в пленке новую волну A_1, которая сложится в фазе с уже существующей в пленке волной A_1. Поэтому результирующая волна в точке *2* имеет удвоенную амплитуду, а ее интенсивность в четыре раза больше чем в точке *1*. Интенсивность результирующей волны в точке *3* по этой же причине в девять раз, а в точке *4* — в шестнадцать раз больше. Если связь между пленкой и призмой охватывает большое количество точек поворота луча, то интенсивность поля в пленке возрастает вдоль оси *z* очень быстро.

Для увеличения интенсивности волноводной моды необходимо выполнение *условия синхронизма*, согласно которому первая волна A_1 в пленке и луч в призме должны одновременно достигать соответственно точек *2* и *2'*. Если лучи A_1 и A_3 составляют с нормалью к поверхности пленки углы θ_1 и θ_3, а расстояние между точками *1'* и *2'* равно *d*, то запаздывание волны в пленке между точками *1* и *2* составит $k_1 d \sin\theta_1 = \varphi_1$, для волны в призме между точками *1'* и *2'* — $k_3 d \sin\theta_3 = \varphi_3$. Тогда условие синхронизма можно записать в одном из следующих видов: $\varphi_1 = \varphi_3$; $k_1 \sin\theta_1 = k_3 \sin\theta_3$; $n_1 \sin\theta_1 = n_3 \sin\theta_3$.

Показатель преломления призмы n_3 обычно больше, чем показатель преломления пленки n_1. Поэтому всегда можно выбрать такой угол падения θ_3, при котором условие синхронизма выполняется. Следовательно, важным преимуществом призменного ввода излучения в пленку является его способность возбуждать в пленке любую заданную волноводную моду (с любым углом θ_1).

Связь между полями в призме и пленке приводит, таким образом, к новому эффекту, который может быть назван *нарушенным полным внутренним отражением*. Связь между призмой и пленкой осуществляется на длине в несколько сотен длин волн света вдоль зазора, что обеспечивает высокую эффективность метода. Экспериментально была достигнута эффективность передачи, превышающая 90%. Кроме призменного ввода излучения в пленку в интегральной оптике используют связи через сужающийся край пленки и связь через дифракционную решетку, рассматриваемую в следующей главе.

§ 3.10. Оптические волокна

Оптические волноводы (световоды) в виде прозрачных волокон находят основное применение в системах связи в качестве передающей линии, которая должна в первую очередь передавать широкополосные сиг-

налы на большие расстояния с низкой дисперсией, малыми потерями и шумами. Требования к этим характеристикам оптических волноводов существенно отличаются от требований к пленочным, полосковым и другим планарным волноводным структурам, поскольку последняя группа волноводов используется в качестве коротких соединительных линий оптических схем. Затухание направляемой моды в волноводных отрезках длиной, не превышающей нескольких миллиметров, около нескольких децибел вполне допустимо, а малая дисперсия в коротких секциях не будет вызывать заметного искажения сигнала. Требования к этим характеристикам оптических волноводов, длина которых между центрами связи и оконечными станциями или ретрансляторами может достигать сотен километров, совершенно иные.

Несмотря на единство исходных положений и ряда основных выводов теории диэлектрических волноводов оптического и СВЧ-диапазонов, особенности оптического диапазона требуют решения дополнительных задач. Они связаны, в частности, с малыми поперечными размерами оптических волноводов (мкм), соизмеримостью возможных неоднородностей с длиной волны, большим разнообразием типов волноводов и характеристиками излучателей. Размеры излучателей в оптическом диапазоне соизмеримы с поперечными размерами волновода, поэтому их можно рассматривать в ряде случаев как элементарные излучатели или их ограниченную совокупность. Многие излучатели оптического диапазона не строго монохроматичны.

Волоконный волновод обычно состоит из сердечника и оболочки, оптические свойства которой принимают во внимание лишь в случаях поглощения ею лучей, вышедших из волокна. Диаметр оболочки выбирается так, чтобы при передаче сигналов сохранялись все необходимые свойства основной моды и не происходили паразитные эффекты из-за ограниченных размеров оболочки.

Свойства волноводов с круглым поперечным сечением определяются профилем показателя преломления. Световоды со ступенчатым профилем имеют постоянный по радиусу показатель преломления. Световоды с градиентным профилем обладают плавно изменяющимся показателем преломления. Возможен кусочно-градиентный световод, у которого концентрические зоны отличаются от соседних ступенчатым изменением показателя преломления на границе зон при непрерывном изменении внутри зоны.

В зависимости от количества распространяющихся на рабочей частоте типов волн (мод) световоды подразделяют на *одномодовые* и *многомодовые*.

В зависимости от размеров поперечного сечения оптические волноводы можно разделить на световоды с размерами поперечного сечения, значительно большими длины волны света, и на световоды с размерами

поперечного сечения, сравнимыми с длиной волны. В последнем случае явления, происходящие в световодах, анализируются с помощью законов волновой электромагнитной теории света и описываются уравнениями Максвелла. Такой подход был изложен при рассмотрении круглого диэлектрического волновода. В первом случае распространение света можно рассматривать с достаточной точностью, исходя из законов геометрической оптики, трактующей прохождение света как распространение, отражение и преломление световых лучей.

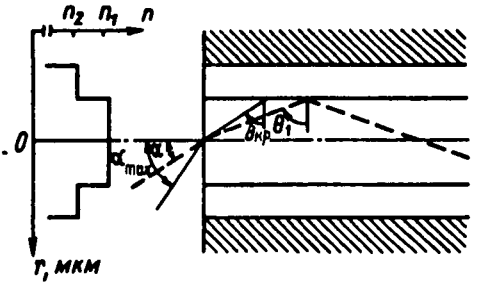


Рис. 3.29. Траектория меридианального луча в световоде со ступенчатым профилем

Лучи в оптических волокнах делят на *меридианальные* и *косые*. Меридианальные лучи лежат в плоскости, проходящей через ось световода, остальные лучи считают косыми. Последние распространяются по ломаным спиральным линиям, причем их проекции на поперечное сечение представляют собой хорды, равные друг другу и имеющие одинаковое расстояние от оси световода. Прохождение меридианального луча в световоде показано иа рис. 3.29. Каждому значению угла падения α на торец световода соответствует свой угол падения θ_1 на границу сердечника и оболочки. Максимальное значение угла $\alpha = \alpha_{max}$, при котором луч еще распространяется в волноводе и $\theta_1 = \theta_{кр} = \arcsin(n_2/n_1)$ называют *апертурным* или *входным углом световода.*

Одним из параметров волновода является его *числовая апертура* $NA = n_0 \sin\alpha_{max} = \sqrt{n_1^2 - n_2^2}$, где n_0 — показатель преломления среды, из которой падает луч на торец световода. От числовой апертуры зависят эффективность ввода излучения в световод, дисперсия, потери на микроизгибах и количество распространяющихся мод N.

Для ступенчатого волоконного световода

$$N = 2\pi^2 \frac{a^2}{\lambda^2}(n_1^2 - n_2^2),$$

где a — радиус сердечника; λ — длина волны применяемого излучения.

У световода с профилем, близким к параболическому, при максимальном значении показателя преломления луча в сердечнике $n_{max} = n_1$ количество направляемых мод в два раза меньше, чем у ступенчатого с такими же параметрами. В градиентных световодах различают три вида возможных путей лучей: 1) центральный (рис. 3.30, *а*), 2) меридианаль-

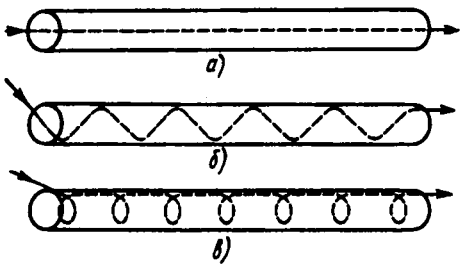


Рис. 3.30. Траектории лучей в градиентных световодах

ный (рис. 3.30, *б*), 3) спиральный (рис.3.30, *в*), если луч входит под косым углом к цилиндрической поверхности.

В системах оптической связи в диапазоне длин волн 0,8...1,6 мкм обычно для одномодовых волокон $2 < a < 5$ мкм и $0{,}006 < \Delta < 0{,}02$, а для многомодовых $12{,}5 < a < 100$ мкм и $0{,}02 < \Delta < 0{,}06$. Числовая апертура одномодовых волноводов мала (0,14...0,20). Поэтому, несмотря на их потенциально высокие характеристики, приходится часто использовать многомодовые световоды, обладающие существенно бóльшими размерами сердцевины, числовой апертурой, эффективностью ввода излучения, допусками при юстировке. Эти преимущества достигаются некоторым ухудшением передаточных характеристик. Необходимо, в частности, считаться с той спекловой структурой картины поля на выходе многомодового волокна, которая возникает в результате интерференции отдельных мод в выходной плоскости. Пространственная частота спекловой структуры пропорциональна диаметру световода и зависит от количества направляемых мод.

Основные причины затухания направляемых мод в световодах связаны с поглощением и рассеянием света. При малых длинах волн преобладает релеевское рассеяние (обратно пропорциональное λ^4) на субмикроскопических оптических неоднородностях материала, вызванных флуктуациями его плотности и состава. При больших длинах волн (в ИК-диапазоне) увеличивается резонансное поглощение света ионами ОН и молекулами, образующими стекло.

Коэффициент затухания моды $\alpha = \alpha_1 \Gamma + \alpha_2(1 - \Gamma)$, где α_1, α_2 — коэффициенты потерь в сердцевине и оболочке; Γ — доля энергии распространяющейся волны, сосредоточенная в сердечнике.

В многомодовых волокнах основной вклад в потери вносит затухание в сердцевине. В одномодовых волокнах значительная часть энергии может распространяться в оболочке, потери в которой при этом становятся существенными.

Кроме указанных основных потерь существуют и другие, возникающие в процессе изготовления в полевых условиях. К ним относятся потери на макро- и микроизгибах волокон, а также потери из-за вытекания, преобразования и различного затухания мод. *Вытекающие моды* — это моды с частотой, меньшей критической. Они соответствуют косым лучам и могут распространяться на значительные расстояния с различ-

ными коэффициентами затухания. Световоды могут быть выполнены из стекла или полимеров. В световоде из кварца затухание может составлять 2 дБ/км в видимом диапазоне (0,85 мкм) и 0,2 дБ/км — в ИК-диапазоне (1,55 мкм). Для производства волокон с малыми потерями необходимо максимальное снижение концентрации примесей в исходном материале. В волоконных световодах из поликристалла бромистого и бромиодистого таллия получено затухание 0,01 дБ/км в длинноволновой инфракрасной области (4...5 мкм). Теоретические расчеты показывают, что в этой области спектра материалы на основе некоторых галоидных соединений, халькогенидных стекол, могут иметь потери в тысячи раз меньшие, чем потери в лучшем кварцевом стекле. В этом перспективном направлении уже достигнуты некоторые успехи, хотя экспериментально полученные результаты еще далеки от теоретических прогнозов.

Полимерные волноводы отличаются высокой стойкостью к действию ионизирующих излучений и высокой пластичностью. Однако в системах связи они пока не получили широкого распространения из-за значительного затухания, а также нестабильности их характеристик (увеличение затухания при старении). При достаточно хорошей технологии устранения колеблющихся С-Н-групп могут быть созданы полимерные оптические волокна с затуханием 20 дБ/км.

Для защиты волоконных световодов от механических повреждений, влаги и других внешних воздействий, их покрывают слоем полимера. *Оптический кабель* представляет собой систему волноводов, предназначенных для передачи оптических сигналов и защищенных внутренними и внешними оболочками от вредных влияний и разрушающих усилий. Примером может служить восьмиволоконная конструкция кабеля с фигурным сердечником и полиэтиленовой оболочкой (рис. 3.31), используемая на оптических линиях связи в некоторых городах нашей страны. Кабели применяют в качестве соединительных линий между АТС и ЭВМ, а также в местных сетях для передачи широкополосной информации (кабельное телевидение, передача данных и др.).

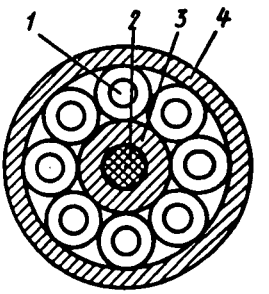


Рис. 3.31. Оптический кабель:

1 — волокно; *2* — армирующий элемент; *3* — пластмассовое покрытие; *4* — полиэтиленовая оболочка

Стыки световодов вносят дополнительные потери и увеличивают дисперсию сигналов в линии передачи. Потери в стыках однотипных световодов могут быть вызваны радиальным смещением осей, угловым рассогласованием, зазором между торцами, царапинами и микронеровностями на торцах. При соединении разнотипных световодов появляются дополнительные потери, связанные с различием числовых апертур, диаметров сердечников, их эл-

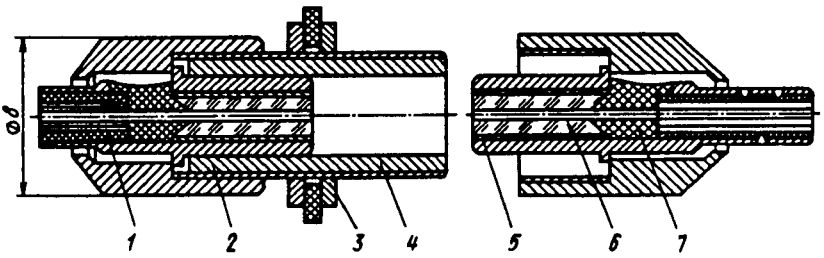


Рис. 3.32. Оптический разъем:

1 — одноволоконный кабель; *2, 3* — фиксирующие гайки; *4* — втулка; *5* — наконечник; *6* — капилляр; *7* — оптическая смола

липтичностью и френелевским отражением. Конструкция разъема, разработанного в институте Радиотехники и электроники РАН приведена на рис. 3.32. Разъемное соединение вносит потери 0,3 дБ.

При соединении различных типов световодов, отобранных по наружному размеру в пределах 140...150 мкм без дополнительного контроля сердечников, потери на соединение составили 1...2 дБ.

К преимуществам волоконно-оптических линий относится невосприимчивость к электромагнитным помехам, отсутствие риска «короткого замыкания» или «замыкания на землю», безопасность работы в окружении горючих веществ (топлива), исключение радиоперехвата или подслушивания, большая информационная способность, малые размеры и масса, использование доступных материалов.

Все расширяющееся техническое использование оптического диапазона сопровождается заменой электрических сигналов на оптические как в линиях связи, так и в местах преобразования информации. Изменение параметров светового потока может происходить в световодах при нарушении полного внутреннего отражения на границе между световодной структурой и внешней средой. Это дает возможность создания на основе световодных структур большой группы малогабаритных *волоконно-оптических преобразователей*, реагирующих на изменение физических параметров внешней среды. Такие преобразователи не имеют электрических контактов, отличаются высоким быстродействием, помехозащищенностью, эффективностью использования в агрессивных и взрывоопасных средах. Для этой цели применяют световоды, поверхность которых может контактировать с внешней средой.

Преобразователи с нарушенным полным внутренним отражением реагируют на изменения относительного показателя преломления на границе световода, площади оптического контакта и микрозазора между световодом и перемещающейся пластиной. Чувствительные элементы преобразователей выполняют на основе волоконных, плоских и профи-

лированных световодов с торцевой или поверхностной чувствительной частью.

Оптический волновод может выполнять функции преобразователя, если вызываемые в нем физические возмущения способны модулировать интенсивность, фазу, поляризацию или длину светового луча. Амплитудные датчики отличаются простотой конструкции и возможностью применения обычного многомодового волокна.

Сенсорные устройства с датчиками с фазовой модуляцией используют интерферометры. Поэтому их конструкция сложнее, но чувствительность значительно выше, чем у амплитудного датчика. Фазовые датчики на основе одномодового волокна в 1000 раз более чувствительны, чем датчики на основе многомодового волокна. Модуляция света по частоте используется в волоконно-оптических сенсорных устройствах, регистрирующих эффект Доплера от движущихся частиц или использующих зависимость фосфоресценции (длины волны) от измеряемого параметра, например, температуры.

В уровнемере применяется протяженный световод, в котором при контакте с контролируемой жидкостью нарушается условие полного внутреннего отражения. При этом зависимость тока фотоприемника от уровня жидкости описывается показательной функцией. Волоконно-оптический сигнализатор уровня жидкости имеет чувствительность около 1 мм. В индикаторе вида жидкости используется различие показателей преломления контролируемых жидкостей (для воды — 1,33; для керосина — 1,41), приводящее к изменению светопропускания оптического канала.

В гидроакустике световодные датчики с амплитудной модуляцией позволяют измерять давление. Для этой цели применяются мембраны с профилированными выступами *2*, создающие периодические изгибы световода *3* (рис. 3.33). Потери в волокне, вызванные изгибами, модулируются действующим на диафрагму *1* давлением. Порог чувствительности датчика удалось понизить до 60 дБ/мкПа.

В устройствах для измерения температуры используют эффекты изменения размеров стеклянного световода и показателя преломления стекла при изменении температуры (ΔT). Вызванный этим сдвиг фазы $\Delta\varphi = -\frac{2\pi L}{\lambda}\left[\left(n + \frac{a\lambda}{2\pi}\frac{\partial\beta}{\partial a}\right)\alpha + \frac{\partial n}{\partial T}\right]\Delta T$, где L — длина волокна, являющегося датчиком; n — среднее значение показателя преломления сердечника;

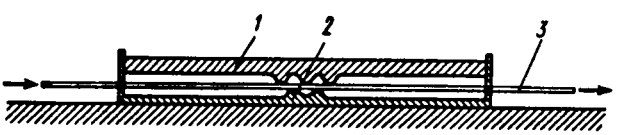


Рис. 3.33. Гидроакустический датчик с амплитудной модуляцией

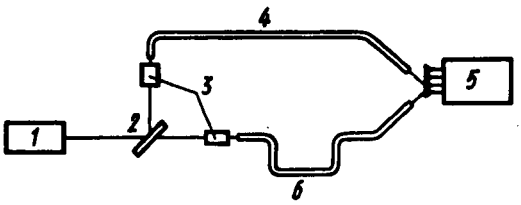


Рис. 3.34. Световодный измеритель температуры с фазовой модуляцией:

1 — лазер; *2* — полупрозрачное зеркало; *3* — объективы микроскопов; *4* — опорное одномодовое волокно (репер); *5* — матрица фотоприемников; *6* — измерительное одномодовое волокно (датчик)

λ — длина волны света (измеренная в вакууме); a — радиус сердечника; β — коэффициент фазы; α — коэффициент линейного расширения стекла. Датчик включают в оптический интерферометр. В некоторых конструкциях сенсорного устройства об изменении температуры судят по перемещению интерференционных полос (рис. 3.34). Чувствительность подобных устройств доходит до 10^{-6} °С/см длины световода в измерительном плече, а инерционность — менее 10^{-4} с.

Для дистанционного контроля температуры флуоресцирующих объектов, поверхностей объектов, подвергаемых реактивно-ионному травлению, а также участков ИС свечение передается к фотоприемнику по волоконному световоду. При этом учитывается линейная зависимость длительности флуоресценции европия от температуры и связанное с этим запаздывание поступающего от световода света фазы относительно фазы модулятора возбуждающих световых импульсов. Точность измерений может быть доведена до 0,01 °С.

Световодные системы в настоящее время используют в оптоэлектронных счетно-решающих устройствах, электронно-оптических преобразователях и усилителях яркости, высокоскоростной фотографии, оптических системах для исправления аберраций и для разделения участков спектра в спектрофотометрах, в медицине — для визуального наблюдения полостей внутри организма, подкожных микроисследований и т. д. *Жгуты* из цилиндрических световодов применяют для передачи изображений.

Волокна можно перепутать так, чтобы объект, наблюдаемый через беспорядочный жгут, на выходном торце был неузнаваем (рис. 3.35). Такие жгуты могут быть применены в криптографии для кодирования оптической информации. Декодирование может быть осуществлено только при использовании того же или идентичного жгута. Кодирующие и декодирующие элементы можно также получить в виде слоя волокон, сканирующего оптическое изображение в одном направлении.

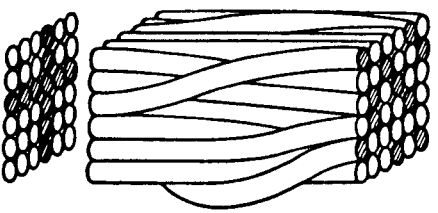


Рис. 3.35. Волоконное кодирующее устройство

Изготовление отдельных кодирующих слоев волокон значительно проще, чем получение больших жгутов.

При бесконтактном и неразрушающем контроле и дефектоскопии все большее применение находят оптические методы с использованием *оптоэлектронных измерительных преобразователей*. Замена традиционных гальванических и магнитных связей оптическими делает такие устройства более простыми, обеспечивает гальваническую развязку цепей и минимизирует схемные решения, а также снижает порог чувствительности, повышает помехозащищенность и быстродействие. Аналого-цифровые преобразователи составляют самую многочисленную группу оптоэлектронных измерительных преобразователей.

Аналого-цифровые преобразователи перемещений с волоконной оптикой делят на две группы. В первой группе волоконно-оптические узлы выполняют простейшую функцию передачи энергии излучения от кодирующего элемента к фотоэлектрическому для образования электрического выходного сигнала. Во второй группе волоконно-оптические узлы осуществляют сложные функции по квантованию, кодированию и передаче измерительной информации в форме излучения. При этом геометрическое кодирование является основной операцией. В процессе оптического кодирования входное оптическое изображение преобразуется в оптическое изображение, наилучшим образом соответствующее возможностям изменения и передачи информации, содержащейся во входном изображении.

В аналого-цифровых преобразователях перемещений используют дискретное оптическое кодирование, при котором на выходе возникает конечная совокупность оптических сигналов, определенным образом размещенных в пространстве. Квантование координат может быть осуществлено, например, двухкоординатной матрицей *2* (рис. 3.36), образованной из полированных торцов жгутов гибких светопроводящих волокон. Размеры торцов каждого жгута определяются шагом квантования.

Кодирование уровней квантования, соответствующих координатам x и y светового штриха *1*, проводится в волоконно-оптическом кодирующем устройстве (ВОКУ) *3*. При этом образование выходного оптического кода *4* заданных функций преобразования $f_1(x)$ и $f_2(y)$ обеспечивается расщеплением

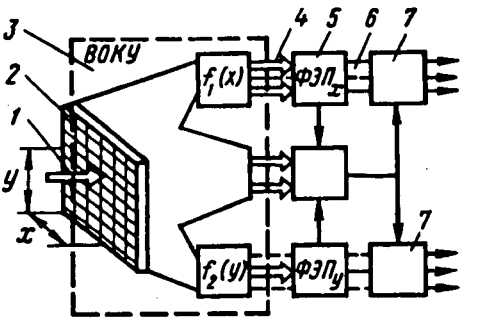


Рис. 3.36. Схема волоконно-оптического аналого-цифрового преобразователя с квантующей матрицей

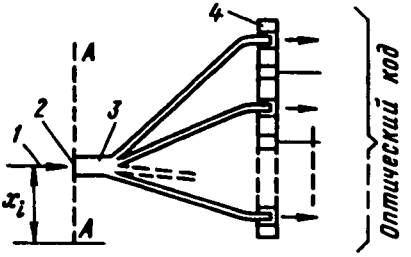


Рис. 3,37. Схема оптического кодирования в аналого-цифровом преобразователе

светового штриха гибкими жгутами и соответствующей раскладкой выходных торцов отдельных волокон этих жгутов по разрядным зрачкам выходного торца ВОКУ. Преобразование оптического кода *4* в электрический код *6* производится блоком фотоэлектрических преобразователей ФЭП *5*. Электронно-логические схемы *7* имеют целью исключение неоднозначности аналого-цифрового преобразования, которое может возникнуть при одновременной засветке штрихом *1* двух соседних квантов на матрице *2*. Схема *7* также выполняет функцию блока памяти, необходимую для сохранения выходной информации преобразователя при остановке светового штриха на границе смежных квантов или кратковременном его исчезновении.

В устройствах с визуальным считыванием результатов измерений к оптическому выходу ВОКУ подключено непосредственно отсчетное устройство.

В современных информационно-измерительных системах требуется применение функциональных аналого-цифровых преобразователей, в которых использование волоконной оптики также дает свои преимущества. При этом коды заданных функций координат входного луча образуют ВОКУ в виде комбинации излучающих и неизлучающих зрачков на выходных торцах, причем каждый из зрачков соответствует определенному разряду кода. Примером может служить схема кодирования, приведенная на рис. 3.37, где на прямой *AA* в точке, соответствующей координате x_i, расположен сформированный входной торец *2* жгута *3* гибких светопроводящих волокон; *n* выходных торцов этих волокон разведены по *p* отверстиям матрицы *4*, каждому из которых приписан определенный разряд выходного позиционного кода, например двоичного. Когда же перемещающийся штрих *1* окажется в точке x_i, то энергия излучения штриха будет «расщеплена» на *n* частей и передана к выходным торцам, образуя определенную комбинацию выходных лучей. Полученную комбинацию излучающих и неизлучающих отверстий матрицы можно рассматривать как *p*-разрядный оптический код, приписывая единицу кода наличию излучения и нуль — его отсутствию. Закон разводки выходных торцов кодирующих элементов ВОКУ определяется типом выбранного кода, числом его разрядов и заданной функцией преобразования $y = \varphi(x)$.

Для осуществления аналого-цифрового преобразования достаточно выходные торцы жгутов разложить по отверстиям матрицы так, чтобы излучающие отверстия ее соответствовали единицам кодовой комбинации значений функции $y_i = \varphi(x_i)$.

§ 3.11. Дисперсия в световодах

Пропускную способность световода или скорость передачи информации ограничивает дисперсия, в результате которой происходит искажение и уширение светового импульса. Дисперсия и связанная с ней ширина полосы пропускания определяются областью частот, в которой могут по волокну распространяться световые импульсы при уменьшении их амплитуды, более чем в два раза. Ширину полосы пропускания характеризуют частотой f_b или соответствующей ей дисперсией Δt_f (нс/км). Если рассматривать световод как гауссовский фильтр, то $\Delta t_f = 0{,}44/f_b$. В общем случае для перехода от временно́й к частотной характеристике полосы пропускания необходимо применять преобразование Фурье.

Общая дисперсия возникает в результате наложения дисперсий двух видов: модовой и хроматической, причем последняя обусловлена материальной и волноводной дисперсиями.

Уширение импульса $\Delta t_{мод}$ в результате *модовой дисперсии* связано с различием скоростей мод, формирующих сигнал и появляющихся на выходе многомодового световода неодновременно. Например, в ступенчатом волоконном световоде длиной L это уширение можно оценить разностью времен распространения основной t_{max} и предельной t_{min} моды, попадающей в волновод под апертурным углом:

$$\Delta t = t_{max} - t_{min} = \frac{L}{c}\frac{n_1^2}{n_2} - \frac{L}{c}n_1 = \frac{L}{c}n_1\sqrt{1-\Delta}.$$

Так как $\Delta \ll 1$, то

$$\Delta t_{мод} = \Delta t/L \approx n_1 \theta_{кр}^2/2c.$$

В одномодовом световоде и градиентном световоде с профилем, близким к параболическому, $\Delta t_{мод} \approx 0$.

Хроматическая дисперсия связана с немонохроматичностью излучения, вводимого в световод. Для одномодовых инжекционных лазеров она доходит до 1 нм, для многомодовых полупроводниковых лазеров — до 3 нм, а для светодиодов — до 300 нм. При этом сказывается зависимость показателя преломления материала световода (материальная дисперсия) и изменение модовой структуры в световоде (волноводная дисперсия) от длины волны. В одномодовом световоде волноводная дисперсия обусловлена зависимостью скорости распространения единственной способной к распространению моды от длины волны.

В одномодовых волокнах отсутствует многомодовая дисперсия и устраняются биения между модами. Длина передачи информации ограничена только дисперсией групповой скорости единственной моды. По-

этому перспективы высокоскоростной оптической связи связывают с одномодовым волоконным световодом.

Остановимся несколько подробнее на *волноводной дисперсии* одномодового волокна, пренебрегая пока потерями и полагая $k_z = \beta$. Тогда поперечную компоненту напряженности электрического поля (3.1) с частотой ω можно представить в виде $\dot{E} = \dot{A}(z)B(x,y)$, причем $\partial\dot{A}/\partial z = -j\beta\dot{A}$, где зависимость $\beta(\omega)$ определяется дисперсионным уравнением распространяющейся моды.

Для сигнала, спектр которого занимает узкий диапазон частот вблизи ω_0, при разложении зависимости $\beta(\omega)$ в ряд можно ограничиться учетом только нескольких членов ряда:

$$\frac{\partial}{\partial z}\dot{A}(z,\omega) = -j\left[\beta(\omega_0) + \left(\frac{\partial\beta}{\partial\omega}\right)_{\omega=\omega_0}(\omega-\omega_0) + \frac{1}{2}\left(\frac{\partial^2\beta}{\partial\omega^2}\right)_{\omega-\omega_0}(\omega-\omega_0)^2\right]\dot{A}(z,\omega) \tag{3.65}$$

и представить комплексную амплитуду в виде произведения двух сомножителей:

$$\dot{A}(z,\omega) = \dot{A}_0(z,\omega-\omega_0)\mathrm{e}^{-j\beta(\omega_0)z}. \tag{3.66}$$

Первый из них характеризует комплексную огибающую, медленно меняющуюся в зависимости от пространственной координаты, второй соответствует сильной зависимости от z.

Пространственно-временна́я зависимость сигнала может быть найдена обратным преобразованием Фурье:

$$\dot{a}(z,t) = \int_{-\infty}^{\infty}\dot{A}(z,\omega)\mathrm{e}^{j\omega t}\,\mathrm{d}\omega = \int_{-\infty}^{\infty}\dot{A}_0(z,\omega-\omega_0)\mathrm{e}^{-j[\beta(\omega_0)z-\omega_0 t]}\mathrm{e}^{j(\omega-\omega_0)t}\,\mathrm{d}(\omega-\omega_0) =$$

$$= \dot{a}_0(z,t)e^{-j[\beta(\omega_0)z-\omega_0 t]}.$$

Спектр огибающей сигнала согласно (3.65) и (3.66) должен удовлетворять уравнению

$$\frac{\partial}{\partial z}\dot{A}_0(z,\omega-\omega_0) = -j\left[\left(\frac{\partial\beta}{\partial\omega}\right)_{\omega_0}(\omega-\omega_0) + \frac{1}{2}\left(\frac{\partial^2\beta}{\partial\omega^2}\right)_{\omega_0}(\omega-\omega_0)^2\right]\dot{A}_0(z,\omega-\omega_0), \tag{3.67}$$

а пространственно-временна́я зависимость огибающей сигнала — уравнению

$$\left(\frac{\partial}{\partial z} + \frac{1}{u}\frac{\partial}{\partial t}\right)\dot{a}_0(z,t) = j\frac{1}{2}\left(\frac{\partial^2\beta}{\partial\omega^2}\right)_{\omega_0}\frac{\partial^2}{\partial t^2}\dot{a}_0(z,t), \tag{3.68}$$

которое получено обратным преобразованием Фурье выражения (3.67) с учетом свойства преобразования Фурье

$$\int_{-\infty}^{\infty} \frac{\partial^n \dot{a}_0}{\partial t^n} e^{-j(\omega-\omega_0)t} dt = [j(\omega-\omega_0)]^n \dot{A}_0.$$

Уравнение (3.68) при $\partial^2\beta/\partial\omega^2 = 0$ превращается в однородное волновое уравнение, т. е. в этом случае импульс распространяется без искажения (изменения формы огибающей a_0) с групповой скоростью $u = (\partial\omega/\partial\beta)_{\omega_0}$. Чем больше вторая производная $\partial^2\beta/\partial\omega^2$, тем больше искажение формы импульса, обусловленное изменением групповой скорости

$$\Delta\left(\frac{\partial\beta}{\partial\omega}\right) = \Delta\left(\frac{1}{u}\right) = \frac{\partial^2\beta}{\partial\omega^2}\Delta\omega$$

на протяжении всей спектральной полосы $\Delta\omega \gg \tau^{-1}$ импульса длительности τ. После прохождения по волокну длиной l компоненты импульса на крайних точках спектральной полосы окажутся сдвинутыми на интервал

$$\Delta\tau \approx l\Delta\left(\frac{1}{u}\right) = \left(\frac{\partial^2\beta}{\partial\omega^2}\right)_{\omega_0} \Delta\omega l =$$

$$= \frac{\omega_0}{\lambda_0}\left(\frac{\partial^2\beta}{\partial\omega^2}\right) l\Delta\lambda = \frac{\omega_0^2}{2\pi c} l\left(\frac{\partial^2\beta}{\partial\omega^2}\right)_{\omega_0} \Delta\lambda. \qquad (3.69)$$

Производная $k_2 = \partial^2\beta/\partial\omega^2$ получила название *дисперсии*, характеризуя в данном случае волноводную дисперсию (здесь групповой скорости) одномодового волокна (рис. 3.38). Поскольку правая часть уравнения (3.68) мнимая, то дисперсия воздействует только на фазовую характеристику огибающей.

В одномодовых волокнах на основе кварца минимум дисперсии лежит в области длин волн 1,3 мкм. Именно эта спектральная область обеспечивает минимум искажений, максимум быстродействия и пропускной способности ($2\Delta\tau/l$) волоконно-оптических линий связи. Соотношение (3.69) характеризует влияние на искажение сигнала также спектральной ширины $\Delta\lambda$ источника и объясняет стремление свести ее к минимуму.

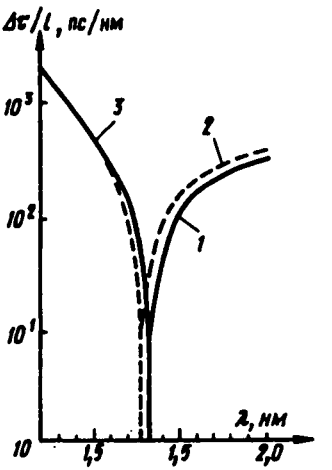


Рис. 3.38. Дисперсии одномодового волокна:

1 — волноводная; *2* — диэлектрической проницаемости, *3* — полная

Приведенная ранее качественная оценка влияния дисперсии на распространение импульсов может удовлетворительно объяснить искажение сигнала с относительно широким спектром ($\tau\Delta\omega \gg 1$), подобных тем, которые получают при модуляции светодиодов. Процесс уширения узкополосных импульсов ($\tau\Delta\omega \approx \pi$), формируемых в некоторых случаях лазерами, требует более подробного рассмотрения уравнения распространения импульса (3.68). Если перейти к системе координат $\eta = t - \xi/u$, движущейся с групповой скоростью u, то это уравнение можно переписать в виде

$$j\frac{\partial \dot{a}_0}{\partial \xi} + \frac{1}{2}\left(\frac{\partial^2 \beta}{\partial \omega^2}\right)_{\omega_0}\frac{\partial^2 \dot{a}_0}{\partial \eta^2} = 0. \qquad (3.70)$$

Здесь переменная η линейно возрастает со временем t, т. е. характеризует время (в единицах расстояния), за которое происходит наблюдение за движением данной точки огибающей импульса, причем $d\eta = 0$ при $dt = d\xi/u$, $\xi = z - z_0$ и z_0 — произвольная постоянная.

Уравнение (3.70) имеет тот же вид, что и волновое уравнение (2.41) для параксиальных лучей. Если его переписать для одномерного варианта в виде $j\frac{\partial f}{\partial z} - \frac{1}{2k}\frac{\partial^2 f}{\partial x^2} = 0$, то нетрудно составить таблицу величин-аналогов (табл. 3.1) и, используя решение уравнения (2.41), найти решение, описывающее распространение импульса гауссовой формы вдоль оптического волокна с дисперсией

$$\dot{a}_0(z,t) = \frac{A}{\sqrt{\delta_0 + j\alpha - z(d^2\beta/d\omega^2)}}\exp\left[\frac{j(t - z/u)^2/2}{z(\partial^2\beta/\partial\omega^2) - j\alpha - \delta_0}\right].$$

Таблица 3.1

Гауссовский пучок	*Гауссовский импульс*
$f(x, z)$	$\dot{a}_0(z,t)$
k	$-\partial^2\omega/\partial\beta^2$
z	ξ
x	η

При $z = z_0$ частота импульса

$$\omega = \omega_0 - \delta_0 t/(\alpha^2 + \delta_0^2)$$

линейно зависит от времени, т. е. импульс модулирован по частоте. Параметр $\delta_0 = z_0(\partial^2\beta/\partial\omega^2)$ называют *параметром линейно-частотной мо-*

дуляции (ЛЧМ). Для $\delta_0 > 0$ высокие частоты располагаются на переднем фронте импульса ($t < 0$), а низкие — на заднем фронте ($t > 0$). Для $\delta_0 < 0$ имеет место обратное поведение. Импульсы с линейно-частотной модуляцией иногда называют *чирп-импульсами*.

После распространения импульса в оптическом волокне длиной l его амплитуда

$$\dot{a}_0(l,t) = \frac{A}{\sqrt{\delta + j\alpha}} \exp\left[-\frac{(\alpha - j\delta)(t-\tau)^2}{2\sqrt{\alpha^2 + \delta^2}}\right],$$

где временна́я групповая задержка $\tau = l/u = (\partial\beta/\partial\omega)l$, а модифицированный параметр модуляции

$$\delta = \delta_0 - (\partial^2\beta/\partial\omega^2)l.$$

Модуляция частоты приводит к изменению длительности импульса. В частности, при $\delta_0 \neq 0$ и $\delta = 0$ происходит укорочение импульса. Это можно объяснить тем, что при «нормальной дисперсии» ($\partial^2\beta/\partial\omega^2 > 0$) групповая скорость высокочастотных фурьер-компонент меньше скорости низкочастотных компонент. При $\delta_0 > 0$ первые расположены в передней части импульса, движутся медленнее низкочастотных составляющих и настигаются ими. При достаточно большой длине волокна, когда $\delta_0 < 0$, импульс снова расширяется и для его сжатия необходима система с аномальной дисперсией ($\partial^2\beta/\partial\omega^2 < 0$). Сжатие удобно реализовать с помощью пары дифракционных решеток. Как видно из рис. 3.38, в оптических волокнах нормальная дисперсия возникает на частотах выше критической, а аномальная — ниже критических.

В общем случае решение уравнения (3.70) имеет вид

$$a_0(\eta, z) = \int_{-\infty}^{\infty} \frac{\dot{a}_0(\eta, 0)}{\sqrt{j2\pi k_2 z}} e^{j(\eta - t)^2/(2k_2 z)}\, dt. \qquad (3.71)$$

Если в начале координат ($z = 0$) гауссовский импульс имеет огибающую

$$a_0(\eta, 0) = a_0(t, 0) = a_0(t) = \rho_0 e^{-t^2/(2\tau_0^2)}, \qquad (3.72)$$

то решение (3.71) приводит на выходе световода ($z = l_1$) к комплексной огибающей

$$a_1(t) = \rho_1 e^{-t^2/(2\tau_1^2)} e^{j(\delta_1 t^2 + \varphi_0)/2}, \qquad (3.73)$$

где

$$\tau_1 = \tau_0\sqrt{1 + l_1^2/L_1^2} = \sqrt{\tau_0 + (k_2 l_1/\tau_0)^2} = \Delta\omega_1^{-1}; \qquad (3.73а)$$

$$\delta_1 = l_1 k_2^{-1}(l_1^2 + L_1^2)^{-1}; \qquad (3.73б)$$

$$\rho_1 = \rho_0(1 + l_1^2/L_1^2)^{-1/4}; \tag{3.73в}$$

$$L_1 = \tau_0^2/|k_2|^{-1}; \tag{3.73г}$$

$$\varphi_0 = \operatorname{arctg}(k_2 l_1 \tau_0^{-2}). \tag{3.73д}$$

Полная фаза импульса (3.66) с огибающей (3.73)

$$\Phi_1(t) = \omega_0 t + \delta_1 t^2/2 + \varphi_0/2. \tag{3.73е}$$

Мгновенная частота

$$\omega_1(t) = \partial\Phi_1/\partial t = \omega_0 + \delta_1 t \tag{3.74и}$$

изменяется во времени линейно со скоростью $\partial\omega_1/\partial t = \delta_1$, определяемой параметром линейно-частотной модуляции (3.73б). Ширина импульса (3.73а) увеличилась, а амплитуда (3.73в) уменьшилась по сравнению с начальными значениями ($\tau_0 = \Delta\omega_0^{-1}$ и ρ_0), однако форма огибающей сохранилась.

Таким образом, в световоде с дисперсией спектрально ограниченный импульс преобразуется в чирп-импульс с линейно-частотной модуляцией. Знак ЛЧМ определяется знаком дисперсии $k_2 = (\partial^2\beta/\partial\omega^2)_{\omega=\omega_0}$. При $k_2 > 0$ низкие частоты оказываются расположенными на переднем фронте импульса ($t < 0$), а высокие — на его хвосте ($t > 0$). В оптических волокнах нормальная дисперсия ($k_2 > 0$) возникает на частотах выше, а аномальная ($k_2 < 0$) — ниже критических. Если последовательно соединить два таких световода, то будет происходить сжатие («компрессия») импульса.

Действительно, если на вход оптического волновода с аномальной дисперсией $k_2' < 0$ поступает импульс (3.73), причем начало отсчета расстояний совмещено с его входом, а отсчет времени — с моментом, когда $\varphi_0 = 0$, то на выходе с помощью (3.71) получим

$$\dot{a}_2(t,z) = \frac{\rho_1}{\sqrt{1 + k_2'\delta_1 z + jk_2'\tau_1^{-2}z}} \exp\left(-\frac{\tau_1^{-2} - j\delta_1}{1 + k_2'\delta_1 z + jk_2'\tau_1^{-2}z}\frac{t^2}{2}\right).$$

Длительность этого импульса

$$\tau = \tau_1\sqrt{(1 + k_2'\delta_1 z)^2 + (k_2'\tau_1^{-2}z)^2}.$$

Следовательно, по мере распространения импульс сначала сжимается (при $k_2'\delta_1 < 0$), а затем расширяется, не меняя своей формы. Максимальное сжатие достигается на расстоянии

$$l_2 = \frac{\delta_1 \tau_1^4 |k_2''|^{-1}}{1 + (\delta_1 \tau_1^2)^2},$$

где длительность импульса

$$\tau_2 = \tau_1 [1 + (\delta_1 \tau_1^2)^2]^{-1} = \Delta\omega_2^{-1};$$

параметр линейно-частотной модуляции

$$\delta_2 = 2\partial\omega_2 / \partial t = 0,$$

т. е. модуляция частоты во втором световоде (компрессоре) полностью компенсировала начальную ЛЧМ на входе (выходе модулятора). При этом мгновенная частота $\omega_2 = \partial\Phi_2 / \partial t = \omega_0 - \delta_2 t = \omega_0$ гармонических составляющих во всех точках огибающей импульса становится одинаковой и равной частоте несущей. Следовательно, все составляющие на выходе компрессора распространяются с одинаковой фазовой скоростью и оказываются сфазированными во всех точках огибающей импульса (низкочастотные компоненты догнали высокочастотные).

Для оптимальных условий коэффициент компрессии

$$s = \tau_1 / \tau_2 = \sqrt{1 + (\delta_1 \tau_1^2)^2} = \Delta\omega_2 / \Delta\omega_1,$$

т. е. импульс может быть сжат во столько раз, во сколько произошло расширение его спектра в компрессоре. При этом пиковая мощность увеличивается во столько же раз.

Общая дисперсия в обычном световоде

$$\Delta t_f = \sqrt{\Delta t_{\text{мод}}^2 + \Delta t_{\text{хром}}^2} = \sqrt{\tau_2^2 - \tau_1^2},$$

где τ_1, τ_2 — ширина гауссовских импульсов на входе и выходе волокна.

В качестве примера приведем параметры одной из световодных систем, в которой модовая дисперсия составляет 0,3 нс/км и используется лазер с длиной волны 0,85 мкм и полушириной спектра излучения 3 нм. При этом хроматическая дисперсия равна 300 пс/км, полная дисперсия — 424 пс/км, а ширина полосы — 1,02 ГГц.

Дисперсия определяет и скорость передачи информации по световодам. В волоконно-оптических линиях связи (ВОЛС) обычно применяют импульсные методы модуляции сигналов. Например, в импульсно-кодовой системе модуляции (ИКМ) используют группы стандартных импульсов. При этом весь амплитудный диапазон квантуется на ступени и каждому отсчету присваивается двоичный код в соответствии с его амплитудной ступенью. В позиционно-импульсной системе модуляции (ПИМ) каждому отсчету исходного сообщения соответствует один импульс, положение которого во времени зависит от амплитуды сообщения. Каждому информационному каналу выделяется своя последовательность

временны́х интервалов. Чем более узкими делаются эти интервалы, тем точнее определяется их положение во времени и тем больше пропускная способность системы. Чем выше полоса пропускания световода, тем короче можно сделать и точнее определить его положение и, следовательно, тем выше получить скорость передачи информации.

Системы передачи информации на световодах обладают целым рядом преимуществ: широкополосность, достигающая десятков гигагерц; универсальность, поскольку за счет широкополосности возможно применение унифицированной цифровой системы передачи любых видов информации при совмещении их в одном канале; малая масса и размеры; возможность быстрого и точного определения места повреждения; высокая помехозащищенность вследствие нечувствительности к электромагнитным помехам и наводкам ввиду полной электрической изоляции; гибкость; низкая стоимость, обусловленная экономией дефицитных материалов и существенным уменьшением энергозатрат на производство.

Волоконно-оптические линии передач информации применяют для телевизионной и телефонной связи на большие и малые расстояния; соединения блоков контрольно-измерительной аппаратуры в системах управления технологическими процессами; сопряжения периферийных устройств ЭВМ с ее центральным процессором; соединения авиационной бортовой аппаратуры; систем передачи данных замкнутых систем телевидения, в том числе кабельных, а также соединения печатных системных плат.

§ 3.12. Периодические структуры

Неоднородные периодические структуры образуются однородной направляющей линией, в которую на одинаковых расстояниях включены идентичные неоднородности.

Периодические структуры применяют часто как замедляющие системы. В этом отношении они являются аналогами низкочастотных линий задержки, фазовая скорость в которых $v_ф = 1/\sqrt{L_0 C_0} < c_0$, где L_0 и C_0 — сосредоточенные индуктивность и емкость. При анализе направляющих волновых систем с периодическими неоднородностями методами теории цепей всю систему заменяют эквивалентной схемой и анализируют с помощью известных свойств фильтров с сосредоточенными параметрами.

Любая компонента электромагнитного поля в рассматриваемой направляющей системе должна удовлетворять условию периодичности

$$\dot{F}(z+L) = \dot{F}(z)\mathrm{e}^{-j\varphi},$$

где L — период структуры.

Условие периодичности выполняется даже при постоянной частоте для любого угла $\varphi_n = \varphi_0 + 2\pi n$, если n — целое число ($\pi \geqslant \varphi_0 \geqslant -\pi$). Каждому значению n соответствует *пространственная гармоника* (гармоника Хартри) со своей постоянной распространения

$$\beta_n = \varphi_n / L = (\varphi_0 + 2\pi n)/L = \beta_0 + 2\pi n/L$$

и фазовой скоростью

$$v_{\phi n} = \omega/\beta_n = \omega L/(\varphi_0 + 2\pi n) = v_{\phi 0} L/(L + n\Lambda_0).$$

Таким образом, в периодической структуре в общем случае может распространяться при заданной частоте множество волн

$$\dot{F} = \sum_{n=-\infty}^{\infty} \dot{F}_n e^{j(\omega t - \beta_n z)}.$$

Гармонику с наибольшей фазовой скоростью $v_{\phi 0}$ называют основной. Она соответствует числу $n = 0$. Положительные значения n соответствуют волнам, распространяющимся в сторону положительных значений z, т. е. прямым волнам. Обратными волнами называют волны, распространяющиеся в сторону отрицательных значений z (с отрицательной фазовой скоростью) и соответствующие числу $n < 0$.

Неизбежность появления группы пространственных волн обусловлена необходимостью удовлетворить граничным условиям на сложном профиле периодической структуры с углами, ребрами и щелями.

Наличие пространственных гармоник характерно для периодических направляющих систем. Их не надо смешивать с временны́ми гармониками, соответствующими несинусоидальным колебаниям и волнам высших типов разных частот. Кроме того, волны высших типов могут существовать в волноводах независимо друг от друга. Все пространственные гармоники обязательно существуют одновременно, имеют одинаковую частоту ω и групповую скорость $v_{гр} = d\omega/d\beta_0$.

Поскольку удовлетворить граничным условиям в периодической системе с помощью одной гармоники нельзя, то не имеет физического смысла говорить о передаче всей энергии электромагнитного поля с помощью одной волны. Все пространственные гармоники обязательно существуют одновременно, образуя в сумме единый волновой процесс, сопровождаемый переносом энергии от генератора к нагрузке. Именно поэтому групповая скорость всех пространственных гармоник

$$v_{гр\,n} = (d\beta_n/d\omega)^{-1} = (d\beta_0/d\omega)^{-1} = v_{гр\,0}$$

одинакова и не зависит от номера n, хотя фазовые скорости гармоник совершенно различны.

Однако отдельная n-я пространственная гармоника может иметь реальный физический смысл, если вдоль волновода будет двигаться поток зарядов со скоростью $v_{фn}$. Поскольку в этом случае в синхронизме с зарядами будет одна n-я гармоника, только она будет определять силу, действующую на заряды со стороны поля. Силы, действующие со стороны других пространственных гармоник, не могут изменить энергии зарядов, поскольку они имеют знакопеременный быстроизменяющийся характер.

К типичным периодическим волноводным структурам относят цилиндрические волноводы с металлическими диафрагмами. Они являются основным узлом линейных волноводных ускорителей и сепараторов частиц высоких энергий.

§ 3.13. Машинные методы анализа и синтеза волноводных устройств

Машинные метода расчета волноводных устройств приобретают особую актуальность в связи со становлением нового научного направления — созданием автоматизированных и диалоговых систем проектирования. В основе работы современных систем машинного проектирования лежит электродинамический расчет параметров базовых элементов с определением, например, полной матрицы рассеяния или эквивалентной схемы каждого базового элемента. Анализ матрицы рассеяния или схемы всего устройства, скомпонованного из базовых элементов, производят по правилам линейной алгебры, используемой в теории цепей.

Разработка и применение численной модели включают в себя несколько этапов: постановку задачи, аппроксимацию, численную реализацию, счет и контроль правильности расчета.

Конечная цель использования численной модели состоит в обеспечении точности и детальности результатов, соответствующих требованиям задачи при минимальных условиях и стоимости вычислений.

Математическая формулировка задачи — краеугольный камень всего процесса. Без этого самая мощная вычислительная машина не способна решить даже самую простейшую задачу. Каковы бы ни были особенности постановки задачи, она в конце концов должна включать адекватные описания физического процесса, а именно: интегральные и дифференциальные соотношения для поля плюс соответствующие граничные условия. Увеличение аналитической работы на этой стадии обычно уменьшает вычислительную работу. Экономия аналитической работы достигается ценой увеличения затрат на численный счет и наоборот. Следует напомнить, что одна из особенностей аналитических решений состоит в том, что их можно получить только для ограниченного набора

геометрических конфигураций. На этапе постановки задачи следует установить, какой метод или комбинация методов являются наилучшими для той или иной совокупности задач. Приводимые далее методы служат далеко не полной иллюстрацией тех многочисленных вычислительных приемов, которые используются при расчете полосковых линий и волноводов.

Метод конечных разностей (МКР). Данный метод или метод сеток основан на замене дифференциальных уравнений поля уравнениями в конечных разностях, которые получают заменой производных их приближенными выражениями через разностные отношения или значения функции в отдельных точках координатной сетки. Таким образом, решение дифференциального уравнения сводится к решению системы алгебраических уравнений. МКР относится к наиболее разработанным методам решения краевых задач математической физики. Он широко используется для решения уравнений Лапласа и Пуассона. На его основе была реализована также программа решения двухмерного уравнения Гельмгольца для расчета поля в регулярных волноводах и резонаторах. Например, для квадратной сетки и пятиточечной схемы, применяемой в большинстве работ, уравнение $\nabla_{\perp}^2\varphi + k^2\varphi = 0$ заменяется разностным аналогом

$$\varphi_1 + \varphi_2 + \varphi_3 + \varphi_4 - 4\varphi_0 + (k\Delta)^2\varphi_0 = 0,$$

где Δ — шаг сетки. Если N — число неизвестных значений φ_i в узлах, то составляется система из N уравнений, которая в матричной форме может быть записана в виде

$$\mathbf{L\Phi} + k^2\mathbf{\Phi} = 0,$$

где $\mathbf{\Phi}$ — вектор-столбец, являющийся приближенным представлением функции φ; $\mathbf{L}$ — квадратная матрица порядка N. Таким образом, решение уравнения Гельмгольца в результате дискретизации сводится к решению матричного уравнения $\mathbf{B\Phi} = 0$, где $\mathbf{B} = \mathbf{L} + k^2\mathbf{I}$; $\mathbf{I}$ — единичная матрица.

Метод конечных разностей может найти применение в системах автоматизированного проектирования, так как с его помощью удобно формализовать сложные границы и неоднородные среды. МКР является одним из самых точных методов решения краевых задач. Погрешность расчета можно сделать очень малой, правда, ценой значительного машинного времени. Трудности, связанные с применением данного метода к решению волноводных задач, связаны, в частности, со сложным построением эффективных итерационных схем для оператора Гельмгольца.

Метод конечных элементов (МКЭ). Этот метод можно рассматривать как комбинацию конечно-разностного (МКР) и вариационного (ВМ) методов. Так же, как и в МКР при применении МКЭ расчетная область

разбивается на подобласти (ячейки), однако здесь они могут иметь вид любых многоугольников, в том числе криволинейных. Наиболее распространенной формой элемента является треугольник, поскольку треугольниками можно плотно покрыть любую полигональную область. Для распределения искомой функции внутри k-го элемента выбирается аппроксимирующая функция. В качестве нее можно использовать разложение по полной системе функций, а в некоторых случаях — билинейную функцию. Локальная аппроксимирующая функция позволяет получить зависимость приближенного значения искомой функции от координат и значений функции φ_{ik} в i-х узлах (например, вершинах треугольников). Дальнейшее решение сводится к определению значений φ_{ik}, минимизирующих функционал для соответствующего уравнения искомого поля. Например, для уравнения Гельмгольца минимизируется функционал $F = \iint_D |\nabla\varphi|^2 \, \mathrm{d}S \Big/ \int_D \varphi^2 \, \mathrm{d}S$ на расчетной области D. После приравнивания нулю всех частных производных по коэффициентам разложения образуется система линейных алгебраических уравнений, неизвестными в которых являются значения φ_{ik} в узлах. На границе расчетной области они должны удовлетворять заданным граничным условиям. Значения φ_{ik} на границах между элементами обеспечивают «сшивание» φ_k соседних элементов.

Сеточный подход проявляется в том, что интегрирование ведется только по элементам. Вклад в i-ю вариацию функционала дают только интегралы по площадям элементов, прилегающих к i-у узлу. Поэтому в i-й строке матрицы системы линейных алгебраических уравнений ненулевыми будут только элементы, соответствующие узлам прилегающих элементов. Это значит, что матрица будет иметь ленточную структуру, однако ширина ленты может быть существенно больше, чем в МКР.

Метод моментов. Данный метод используется в том случае, когда для описания электромагнитного поля применяют интегральные соотношения между характеристиками поля и их источниками — токами и зарядами:

$$\mathbf{A}(R) = \mu \int \mathbf{J} G(kR) \, \mathrm{d}S, \quad \varphi(R) = \varepsilon^{-1} \int \rho G(kR) \, \mathrm{d}V,$$

где G — функция Грина для уравнения Гельмгольца $G = (4j)^{-1} H_0^2(kR)$; H_0^2 — функция Ганкеля второго рода. В этом случае для решения задачи необходимо найти распределение источников по заданным граничным условиям. Численные методы решения сводят интегральное уравнение к системе линейных алгебраических уравнений с N неизвестными, которые обычно представляют собой коэффициенты разложения искомой функ-

ции в ряд по базисным функциям в области определения интегродифференциального оператора L. Например, при расчете плотности тока $\mathbf{J}$ операторное уравнение формируется из выражения для напряженности электрического поля $\dot{\mathbf{E}} = -j\omega\dot{\mathbf{A}} - \nabla\dot{\varphi}$. На металлической поверхности напряженность не имеет касательной составляющей, и операторное уравнение приобретает вид

$$L\dot{\mathbf{J}} = (-j\omega\dot{\mathbf{A}} + \nabla\dot{\varphi})_t = -\mathbf{E}_t = 0. \qquad (3.71)$$

В общем случае оператор L осуществляет отображение некоторой области пространства функций $\mathbf{J}$ на область пространства функций $\mathbf{E}$. Если решение единственно, то это отображение взаимно однозначно ($\mathbf{J} = -L^{-1}\mathbf{E}$).

Метод моментов предусматривает решение задачи в четыре этапа.

1. Искомый вектор разлагается по базисным функциям $\{J_n;\ n = 1, 2, ..., N\}$ в области определения оператора: $J = \sum_n I_n J_n$. Выбор системы базисных функций J_n относится к одной из основных проблем при численном решении. Теоретически имеется бесконечное число систем базисных функций, однако лишь малое число базисов адекватно рассматриваемой задаче. Удачный выбор системы базисных функций обычно позволяет добиться более быстрой сходимости решения, что приводит к меньшему объему расчетов при заданной точности.

2. Вводится понятие внутреннего произведения тангенциальных векторных функций на поверхности S:

$$\langle PQ\rangle = \int_S PQ\,\mathrm{d}S$$

и выбирается система весовых функций $\{W_m;\ m = 1, 2, ..., M\}$, принадлежащих области значений оператора L. В случае, когда весовые функции совпадают с базисными ($W_m = J_n$), метод моментов называют методом Галеркина, который эквивалентен вариационному методу Рэлея—Ритца.

3. Операторное уравнение приводят к матричному виду путем вычисления внутреннего произведения пробных функций и левой части операторного уравнения после подстановки в него разложения неизвестной функции по системе базисных функций. Из уравнения (3.71) при этом образуется система уравнений

$$\sum_n I_n \langle W_m, LJ_n\rangle = 0,\ m = 1, 2, ..., M,$$

которую можно записать в более компактном матричном виде

$$ZI = 0,$$

где при использовании метода Галеркина

$$Z = \begin{pmatrix} \langle J_1, LJ_1\rangle & \langle J_1, LJ_2\rangle & \dots \\ \langle J_2, LJ_1\rangle & \dots & \dots \\ \vdots & \vdots & \vdots \end{pmatrix},\quad I = \begin{pmatrix} I_1 \\ I_2 \\ \vdots \end{pmatrix}.$$

Элементы матрицы импеданса

$$Z_{mn} = \int_S W_m(-j\omega\dot{A} + \nabla\dot{\varphi})\,\mathrm{d}S.$$

4. Находится решение матричного уравнения.

Нетривиальность решения матричного уравнения требует $|\det Z| = 0$. Однако для нахождения критических чисел k волновода при конечном числе членов разложения ($N < \infty$) приходится исходить из условия $|\det Z(k)| = \min$.

Декомпозиционный метод. Сложность и протяженность объектов электродинамических задач ставит предел прямой формализации, при котором моделируемому объекту в целом сопоставляется краевая задача. Одним из путей преодоления этого препятствия служит декомпозиция сложного объекта на независимо анализируемые части с последующей рекомпозицией на основании предварительно рассчитанных параметров автономных блоков.

Декомпозиция начинается с выделения пространственных подобластей, рассматриваемых сначала как независимые электродинамические объекты («автономные блоки»). Результатом анализа автономного блока становится такой описательный аппарат, который позволяет находить отклик на всякое граничное возбуждение и именуется *дескриптором*. Существование дескриптора следует из теоремы единственности. Он может формироваться по-разному, например, в виде матриц рассеяния, проводимости и сопротивлений.

В интегральных схемах СВЧ на горизонтальном металлическом листе размещается диэлектрический слой, называемый *подложкой*, на котором расположены плоские проводники. Все устройство помещено в прямоугольный экран; внешние каналы могут быть коаксиальными линиями, подводимыми извне. Декомпозицию удобно сначала проводить вертикальными плоскостями, которые разбивают полный объем на параллелепипеды. Каждый такой физический автономный блок, являющийся элементом реального устройства СВЧ, соприкасается с остальными четырьмя гранями «окнами», которые можно трактовать как поперечные сечения волноводов. Это означает, что блок анализируется совершенно так же, как если бы он находился внутри общей области крестообразного соединения волноводов.

Расчет матрицы рассеяния автономного блока относится к дифракционным задачам. В общем случае по внешним каналам к блоку приходят волны $\mathbf{c}^+$, которые рассматриваются как заданные и нужно определить «отклик» устройства в виде волн $\mathbf{c}^-$, уходящих по тем же каналам и связанных с приходящими соотношением

$$\mathbf{c}^+ = \mathbf{S}\mathbf{c}^-, \tag{3.72}$$

где $\mathbf{S}$ — матрица рассеяния.

Полная матрица рассеяния линейной области получается путем решения серии дифракционных задач, в каждой из которых рассматривается падение на один (α-й или β-й) из p входов одной (n-й или k-й) из собственных волн, свойственных данному присоединенному волноводу.

Рекомпозиция начинается с попарного соединения блоков. Ее конечным результатом является формирование математической модели устройства в целом в виде матрицы рассеяния относительно ее входов.

Краевые задачи для уравнении Максвелла, решение которых непосредственно приводит к нахождению дескриптора, называют ключевыми. Для нахождения всех элементов матрицы рассеяния при решении уравнений Максвелла должны выполняться условия излучения (в дифрагированном поле отсутствуют падающие волны) на всех входах и краевые условия на оставшихся границах (если они имеются).

Решение такой задачи приводит к следующему выражению для элемента $n(\beta)$ столбца матрицы рассеяния:

$$S_{kn}^{\alpha\beta} = \frac{1}{2N_{k(\alpha)}} \int_S \left\{ [E_{(\alpha)} \dot{h}_{k(\alpha)}^{*}] + [\dot{e}_{k(\alpha)}^{*} H_{(\alpha)}] \right\} \mathrm{d}S, \qquad (3.73)$$

где $k = -1, -2, \ldots$; $\alpha = 1, 2, \ldots, P$; N_k — некоторые константы нормировки e и h; $e_{k,n}$, $h_{k,n}$ — поперечные части векторов $E_{k,n}$, $H_{k,n}$; n, k — номера собственных волн, $\mathrm{d}s$ указывает положительное направление, совпадающее с осью z; индексами со звездочкой обозначены волны, коэффициенты распространения которых комплексно сопряжены по отношению к коэффициентам волн без звездочки. Элементы любого другого столбца могут быть найдены из (3.73) изменением n и β.

Общий подход к декомпозиции интегральных схем (ИС) СВЧ приводит к расчленению исходного пассивного устройства на параллелепипеды, включающие относительно простые фрагменты структуры ИС, например, обычную экранированную полосковую линию, две параллельные связанные линии и просто экран (прямоугольный волновод) с диэлектрическим слоем (роль экрана может быть только методической; если моделируемая система открытая, то для моделирующего алгоритма можно выбрать поперечное сечение экранированного волновода настолько большим, что его влияние будет малым).

В случае линейной декомпозиции ключевая задача сводится к серии простых дифракционных задач. В каждой из них задается падение на стык одной из собственных волн каждой из двух стыкуемых линий. Поле дифракции находят в виде суммы всех расходящихся от стыка волн обеих линий. Комплексные амплитуды этих волн непосредственно дают элементы одного из столбцов матрицы рассеяния. При замене типа падающей на стык волны меняется номер столбца. Суммы собственных волн слева и справа от стыка подлежат «сшиванию» по тангенциальным компонентам напряженностей поля.

Решение ключевых задач опирается на предварительный анализ регулярных систем, который должен привести к нахождению собственных волн.

Пространственные объекты в однородной среде, ограниченные координатными поверхностями, называют автономными *многомодовыми блоками* (АМБ). В простейшем варианте каждая грань АМБ выступает как поперечное сечение вол-

новода с идеально проводящими «электрическими» стенками ($E_t = 0$). Многомодовый дескриптор такого АМБ строится на основе собственных волн полубесконечного волновода, продолжающего боковые стенки, начиная от каждой грани. Возможны и иные типы граничных условий на оболочке присоединенных волноводов, например, типа «магнитной стенки» ($H_t = 0$) и смешанных, а также удовлетворяющих условиям повторяемости. Тип используемых граничных условий на оболочке присоединенных волноводов в широких пределах безразличен к свойствам физического объекта, подвергаемого декомпозиции на АМБ, так как волноводы виртуальные.

Для многих типичных конфигураций полосковых структур декомпозицию можно довести до уровня, при котором остаются только блоки в виде параллелепипедов (АМБ). Их размеры определяются конфигурацией расчленяемого объекта. Такую декомпозицию иногда называют «предельной», поскольку исчезает необходимость решать краевые задачи для получения дескрипторов блоков. При предельной декомпозиции все элементы матричных дескрипторов имеют замкнутое выражение (в виде формул); однако их многомодовость может привести к значительной громоздкости алгоритма.

Дескрипторная матрица *минимального автономного блока* (МАБ) оставляет лишь две волны на канал.

Поскольку ключевая задача решена с учетом граничных условий и условия излучения, то полученные выражения должны точно удовлетворять закону сохранения энергии. При соответствующей нормировке матрицы рассеяния становятся унитарными, если проницаемости вещественны.

Анализ электродинамической системы методом МАБ начинается с представления ее в виде совокупности МАБ.

Грани МАБ, совпадающие с границами раздела сред, необходимо соединять *элементом перехода*, необходимым для согласования разных базисов граничных МАБ. Матрица рассеяния элемента перехода для двухмерного варианта

$$S_{\text{п}}^{(2)} = \begin{pmatrix} \rho & \tau_1 \\ \tau_2 & -\rho \end{pmatrix},$$

где

$$\rho = \frac{Z_2 - Z_1}{Z_2 + Z_1}, \qquad \tau_1 = \frac{2Z_1}{Z_2 + Z_1}, \qquad \tau_2 = \frac{2Z_2}{Z_2 + Z_1},$$

если $Z_1 = \sqrt{\mu_1/\varepsilon_1}$ и $Z_2 = \sqrt{\mu_2/\varepsilon_2}$. Матрица рассеяния элемента перехода для трехмерных МАБ

$$S_{\text{п}}^{(3)} = \begin{pmatrix} \rho & 0 & \tau_1 & 0 \\ 0 & \rho & 0 & \tau_1 \\ \tau_2 & 0 & -\rho & 0 \\ 0 & \tau_2 & 0 & -\rho \end{pmatrix},$$

если учитывать сначала волны первой и второй поляризации в первой среде, а затем — во второй.

Дискретизацию металла проводить алгоритмически невыгодно, так как размеры МАБ пришлось бы брать чрезвычайно малыми. В этом случае для каждого подходящего к металлу МАБ присоединяется блок «короткого замыкания», который моделирует металлическую границу полубесконечным каналом и описывается матрицей рассеяния

$$S_{кз} = \begin{pmatrix} R & 0 \\ 0 & R \end{pmatrix},$$

где коэффициент отражения

$$R = \frac{(1+j)\sqrt{\omega\varepsilon_0/2\sigma} - 1}{(1+j)\sqrt{\omega\varepsilon_0/2\sigma} + 1}.$$

Энергетическую изоляцию обеспечивают «электрическая» и «магнитная» стенки. Первая соответствует поверхности идеального проводника ($\sigma \to \infty$) и моделируется блоком идеального короткого замыкания с матрицей рассеяния

$$S^0_{кз} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Вторая моделируется элементом холостого хода

$$S^0_{хх} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Импедансный аналог электромагнитного поля. Метод предусматривает два этапа расчета. На первом этапе объем устройства разбивается на объемы с малыми по сравнению с длиной волны размерами. Элементарный объем рассматривается как 12-входовый многополюсник, стыкуемый с условно вводимыми шестью регулярными волноводами с Т-волнами. На основе приближенного решения уравнений Максвелла (пренебрегают вторыми производными) составляется эквивалентная схема многополюсника, содержащая сосредоточенные индуктивности, емкости, сопротивления и трансформаторы. Из схем элементарных объемов получают полную эквивалентную схему устройства, которая называется *RLC-схемой импедансного аналога электромагнитного устройства.* Граничные условия и условия возбуждения учитываются подключением к общей эквивалентной схеме граничных двухполюсников и трансформаторов. Эквивалентная схема из сосредоточенных элементов (*RLC-аналог*) используется при возбуждении синусоидальным источником в стационарном режиме.

Для исследования переходного процесса эквивалентная схема многополюсника составляется из отрезков соответствующих идеальных Т-

линий и сопротивлений (*Rт-аналог*). Условия возбуждения, в частности, синусоидальные, аппроксимируются прямоугольными импульсами напряжения длительностью, равной времени задержки волны в элементарной линии.

На втором этапе расчета производится прямое (неитерационное) численное определение параметров полей и внешних параметров устройства, например, его матрицы рассеяния.

Импедансный аналог электромагнитного поля, как численный метод, можно считать дальнейшим развитием метода конечных разностей. Как и МКР он отличается универсальностью, и погрешность расчета может быть доведена до заданного значения путем выбора соответствующего уровня дискретизации.

Принципиальное отличие данного метода от МКР состоит в устранении сложного и длительного итерационного процесса решения краевой задачи.

Машинное проектирование. Проектирование и, в частности, машинное проектирование, есть задача синтеза. Однако решение обратных задач электродинамических устройств основывается, главным образом, на высокоэффективных алгоритмах численного решения соответствующих прямых задач (задач анализа).

Решение любой конструкторской задачи, в том числе и проектирование волноводных устройств, сочетает в себе два основных этапа. Первый этап — выбор одного или нескольких эскизных вариантов конструкции. На этом этапе эвристического плана используются некоторые общие представления о принципе работы, известные из литературы или опыта разработки аналогичных устройств, приближенные расчеты, справочный графический или табличный материал.

Второй этап заключается в окончательном выборе варианта и уточнении его параметров в процессе оптимизации конструкции. При этом используются более точные эквивалентные схемы и расчеты, а также экспериментальные модели или макеты, причем экспериментальная оптимизация обычно бывает продолжительной и дорогостоящей. Применение ЭВМ в проектно-конструкторских изысканиях позволяет наиболее эффективно проводить именно этот этап оптимизации. На первом этапе проектирования роль машины обычно является второстепенной ввиду трудностей формализации творческой деятельности. На втором этапе сложности связаны только с трудоемкостью расчетов, и поэтому естественно стремление автоматизировать в первую очередь этот этап, используя быстродействующую вычислительную технику.

Система автоматизированного проектирования интегральных схем СВЧ на уровне *параметрического синтеза* осуществляет полный или квазиполный параметрический синтез проектируемого модуля по необходимому количеству показателей и возможных пределов их изменения.

Общая эффективность модуля, как и всего устройства СВЧ, оценивается целевой функцией $F = f(x_1, x_2, ..., x_n) = f(\mathbf{X})$, где элементы x_i множества $\mathbf{X}$ являются частными целевыми функциями, определяющими количественные и качественные параметры устройства в соответствии с техническим заданием на проектирование. К таким показателям относятся массогабаритные данные, энергопотребление, диапазон частот, быстродействие, чувствительность, коэффициент усиления, ширина полосы пропускания, степень подавления сигнала в полосе заграждения, дальность действия, точность, безотказность, долговечность, ремонтопригодность, сохраняемость, ударная и вибрационная стоимость, влагостойкость, обеспечение кондуктивной и конвекционной передачи теплоты, радиационная стойкость и баростойкость, технологичность, безопасность, экономичность и т. д.

Первым этапом машинного проектирования является расчленение заданной конструкции на базовые элементы. Система автоматизированного проектирования интегральных схем СВЧ представляет собой комплекс программ для быстродействующей ЭВМ или вычислительного комплекса, в который входят: библиотека (архив) вспомогательных алгоритмов (набор программ оптимизации, алгоритмы матричной алгебры и т. п.); программа-транслятор для оперативного моделирования всех типов волноводных трактов и различных встречающихся неоднородностей в них; программа-транслятор для автоматизированного изготовления фотооригиналов, конструктивной и технологической документации.

Системы автоматизированного проектирования позволяют проектировать сложные СВЧ-структуры, компилируя их посредством программ-трансляторов волноводных устройств. Обращение к программам-трансляторам производят с помощью формализованных заданий, которые являются вводимыми данными в систему. Оно составляется разработчиком на основе выбранной эскизной схемы проектируемого устройства.

Формализованное задание включает в себя следующую информацию: топология (описание состава тракта и схемы соединения базовых элементов); числовые данные (геометрические и электромагнитные характеристики базовых элементов); директивы (набор констант, управляющих процессом проектирования, к которым относится, в частности, набор базисных функций для волноводных устройств).

Наиболее перспективными являются *диалоговые методы* машинного синтеза устройств СВЧ. Они позволяют разрабатывать устройства на электродинамическом уровне путем машинных экспериментов и параметрической оптимизации геометрии удачных вариантов сложных элементов, не расчленяемых на более простые. Подобный новый этап в автоматизации проектирования, сохраняя все положительное, что накоплено техникой машинного проектирования с декомпозицией сложных уст-

ройств, ориентирует инженеров-разработчиков на поиск в диалоге с ЭВМ новых нетрадиционных решений, позволяющих повысить компактность устройств СВЧ, уменьшить их массу и улучшить электрические параметры.

Диалоговое проектирование устройств СВЧ с использованием ЭВМ предполагает решение двух вычислительных задач. Во-первых, должны быть разработаны алгоритмы и универсальные электродинамические программы для расчета отдельных классов устройств. Во-вторых, требуется создание универсальных программ для автоматической оптимизации параметров по заданному критерию. Так как время анализа волноводных устройств на электродинамическом уровне велико, то оптимизационные программы удается применить пока только для относительно простых устройств. Сложные недекомпозируемые устройства оптимизирует разработчик в процессе диалога, используя минимаксную технологию синтеза.

Структурный синтез волноводных устройств проводится на основе эвристических приемов поиска новых технических решений (ТР) и конечного множества метапроцедур (МП), определяющих любую интеллектуальную деятельность человека. При разработке оптимальных волноводных устройств используются следующие метапроцедуры, обобщающие правила формирования дерева вариантов: генерирование геометрических образов и ассоциаций (МП1), формирование множества элементов и установление связей между ними (МП2), использование аналогий (МП3), изменение исходных допущений (МП4), декомпозиция на независимые задачи (МП6), декомпозиция на подзадачи (МП5), расширение функций элементов (МП7), изменение целей (МП8), формирование поэтапного эвристического алгоритма поиска оптимальных решений (МП9).

Правила отбрасывания малоперспективных вариантов, т. е. усечения дерева вариантов, обобщают следующие метапроцедуры: установление физической реализуемости (МПА), выявление предельных возможностей и инвариантов (МПБ), выявление топологии, упрощающей математический анализ (МПВ), выявление ключевых факторов (МПГ).

Метопроцедура МП1 формирует понятия, образы, гипотезы, которые далее используются для формирования множества элементов и правил их взаимодействия (МП2). Важным моментом в МП1 является «осознание», «понимание» проблемы, причем часто до детального рассмотрения и математического анализа. Подобное «понимание» может возникнуть в результате как сопоставлений и аналогий (МП3), так и на основе ассоциаций.

Метапроцедура МП2 лежит в основе обоснования математических моделей и алгоритмических процедур анализа и синтеза устройств. МП1 и МП2 регулируют процесс формирования ряда множеств элементов, на которые можно расчленить проектируемое устройство, а также выявляют закономерности взаимодействия элементов в множествах друг с другом так, чтобы множества элементов позволяли бы моделировать процессы и устройства с приемлемой точностью. МП1 и МП2 требуют периода адаптации, накопления новых знаний в качестве ориентировочной основы действий для последующего создания способа поиска оптимальных решений. Как правило, МП1 и МП2 не исследуют конкретные устройства, их основная задача — сформировать методологическую и математиче-

скую основы исследования и проектирования. МП2, МПГ, МПЖ и МПЕ используются также и в алгоритмах анализа и параметрического синтеза.

Множеств, сформированных на основе МП1, МП2, МП3, может оказаться недостаточным для проведения анализа и еще в большей мере для синтеза сложных волноводных устройств. Для таких устройств определяющими могут оказаться МП5 и МП6. Задачи усечения дерева вариантов могут при этом контролироваться МПБ, МПД, МПЕ и МПЖ. МП5 используется при проектировании наиболее широко, как более простая метапроцедура и соответствующая подходу «от частного к общему».

Метапроцедура МП6 является, по-видимому, одной из наиболее сложных, так как выявляет интегральные закономерности, которые могут использоваться в бесформульном диалоговом синтезе, формировать условия физической реализуемости (МПА), инварианты предельных возможностей (МПБ), идеи поиска решений (МИД) и, наконец, математический аппарат синтеза.

Для решения задач машинного проектирования волноводных устройств, как правило, необходимо применение мощных машин с большим быстродействием и объемом оперативной памяти. В то же время вычерчивание чертежей и графиков, выпуск текстовой информации невыгодно поручать полностью высокопроизводительной машине. Эти процедуры отнимают много ресурсов ЭВМ, и делают использование дорогостоящих ЭВМ неэффективным. Их выгоднее выполнять на комплексе мини-ЭВМ, включающем управляемые графические и другие периферийные устройства, выполняющие относительно простые задачи.

Глава 4
Излучение и дифракция электромагнитных волн

Излучением электромагнитных волн называют процесс передачи энергии электромагнитного поля от источника в окружающее его пространство. Излучающую систему, возбуждающую в этом пространстве волновое поле, называют *антенной*. Основная задача теории излучения заключается в установлении связи между токами и зарядами излучающей системы и возбужденным в пространстве электромагнитным полем.

§ 4.1. Электродинамические потенциалы

Связь между электромагнитными волнами и их источниками — токами и зарядами — наиболее отчетливо выражена в уравнениях для потенциалов электромагнитного поля:

$$\nabla^2\mathbf{A} - \frac{1}{v^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = -\mu J, \qquad \nabla^2\varphi - \frac{1}{v^2}\frac{\partial^2\varphi}{\partial t^2} = -\frac{\rho}{\varepsilon}, \tag{4.1}$$

где $v^{-2} = \varepsilon\mu$.

Эти *неоднородные волновые уравнения* (уравнения Даламбера) получены из совместного решения уравнений Максвелла в предположении, что

$$\mathbf{B} = \operatorname{rot}\mathbf{A}, \qquad \operatorname{div}\mathbf{A} = -\frac{1}{v^2}\frac{\partial\varphi}{\partial t}. \tag{4.2}$$

Если токи и заряды меняются по гармоническому закону, то уравнения (4.1) переходят в уравнения Гельмгольца

$$\nabla^2\dot{\mathbf{A}} + k^2\dot{\mathbf{A}} = -\mu\dot{\mathbf{J}}, \qquad \nabla^2\dot{\varphi} + k^2\dot{\varphi} = -\dot{\rho}/\varepsilon. \tag{4.3}$$

Решение любого линейного дифференциального уравнения можно найти, если известна *функция Грина*. Так называют функцию, удовлетворяющую дифференциальному уравнению, когда в его правой части стоит δ-функция. Для уравнения (4.3) с учетом условия излучения функция Грина

$$G(p,q) = \frac{e^{-jkR_{pq}}}{4\pi R_{pq}}, \tag{4.4}$$

где R_{pq} — расстояние между источником q и точкой наблюдения p.

В общем случае, когда правая часть уравнения Гельмгольца равна $M(q)$ и источники находятся в объеме V_q решение уравнения записывают в виде

$$f(p)=\int_{V_q} M(q)G(p,q)\,\mathrm{d}V_q. \tag{4.5}$$

Для скалярного потенциала $M(q)=\dot{\rho}/\varepsilon$, для векторного потенциала $M(q)=\mu\dot{\mathbf{J}}$. Тогда решения (4.3) принимают вид

$$\varphi(p)=\frac{1}{\varepsilon}\int\frac{\rho(q)\,\mathrm{e}^{-jkR_{pq}}}{4\pi R_{pq}}\mathrm{d}V_q;$$

$$\mathbf{A}(p)=\mu\int_{V_q}\frac{\dot{\mathbf{J}}(q)\,\mathrm{e}^{-jkR_{pq}}}{4\pi R_{pq}}\mathrm{d}V_q. \tag{4.6}$$

В частном случае точечного излучателя ($V_q\to 0$, $\lambda\gg d\ll R_{pq}$) с зарядом и «током»

$$Q(p)=\int_{V_q}\rho(q)\,\mathrm{d}V_q,\quad I\mathrm{d}=\int_{V_q}\dot{\mathbf{J}}\,\mathrm{d}V_q;$$

$$\dot{\varphi}(p)=\frac{\dot{Q}(q)\,\mathrm{e}^{-jkR_{pq}}}{4\pi\varepsilon R_{pq}};\quad \mathbf{A}=\frac{\mu\dot{I}\mathrm{d}\,\mathrm{e}^{-jkR_{pq}}}{4\pi R_{pq}}. \tag{4.7}$$

Из уравнений (4.4) и (4.7) следует, что функция Грина здесь имеет простой физический смысл — она описывает (с точностью до постоянного множителя) распределение в пространстве поля уединенного *точечного излучателя.*

В общем случае распределенных излучателей решения волновых уравнений в виде (4.5) и (4.6), выражают, очевидно, суперпозицию или суммирование потенциалов всех точечных зарядов $\rho\,\mathrm{d}V$ и токов $\mathbf{J}\,\mathrm{d}V$.

Важно отметить, что как в процессе суммирования (4.5), так и при расчете поля (4.7) точечного заряда для каждого из них учитываются те отставания фазы (kR_{pq}), которые связаны с конечным временем (R_{pq}/v) распространения волн от излучателя q до расчетной точки p со скоростью v: при вычислении потенциалов в момент t необходимо взять значения заряда и тока в элементе $\mathrm{d}V_q$ в момент его излучения $t-R_{pq}/v$. Ввиду этого потенциалы (4.6) получили название *запаздывающих.*

Использование запаздывающих потенциалов значительно упрощает расчет поля излучения многих видов антенн. В этом можно убедиться на примере расчета элементарного электрического излучателя в виде короткого диполя, поле которого на больших расстояниях эквивалентно полю источника с исчезающе малыми размерами.

§ 4.2. Элементарный электрический излучатель

Диполь постоянной длины d, в котором заряды меняются по гармоническому закону с частотой ω, обладает электрическим моментом

$$\mathbf{p}(t) = q(t)\mathbf{d} = q_m \mathbf{d} \sin \omega t = \mathbf{p}_m \sin \omega t,$$

где q_m — максимальный заряд; $\mathbf{p}_m$ — максимальный электрический момент вибратора.

Такой диполь эквивалентен линейной короткой антенне длиной d, возбуждаемой переменным током

$$i = \frac{\mathrm{d}q}{\mathrm{d}t} = \frac{1}{d}\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{\omega p_m}{d}\cos \omega t = \mathbf{JS} = \frac{\mathbf{J}}{\mathbf{d}}V,$$

где S — площадь поперечного сечения антенны; $V = \mathbf{Sd}$ — ее объем.

Комплексная амплитуда эквивалентного тока

$$\dot{I}_m = j\frac{\omega}{d}\dot{p}_m = j\frac{\omega}{d}p_m. \tag{4.8}$$

Векторный потенциал короткого тонкого провода в среде без потерь на расстояниях, много больших его длины ($R \gg d$) согласно уравнениям (4.7) и (4.8),

$$\mathbf{A} = \frac{\mu \dot{I}_m \mathbf{d}}{4\pi R}\mathrm{e}^{-jkR} = j\frac{\mu\omega\dot{\mathbf{p}}_m}{4\pi R}\mathrm{e}^{-jkR}.$$

В соответствии с уравнением (4.2) в сферической системе координат, когда дипольный момент находится в начале координат и направлен вдоль оси z,

$$H_R = H_\theta = 0;$$

$$\dot{H}_\alpha = -\frac{1}{\mu}\frac{\partial \dot{\mathbf{A}}_z}{\partial R}\frac{\partial R}{\partial r} = \frac{k^2 d\dot{I}_m \sin\theta}{4\pi}\left(\frac{1}{k^2R^2} + j\frac{1}{kR}\right)\mathrm{e}^{-jkR}, \tag{4.9}$$

где r — радиус-вектор цилиндрической системы координат.

Из уравнения Максвелла для комплексных амплитуд ($\mathrm{rot}\,\dot{\mathbf{H}} = j\omega\varepsilon\dot{\mathbf{E}}$)

$$\dot{E}_R = \frac{k^3 d\dot{I}_m \cos\theta}{j\omega 2\pi\varepsilon}\left(\frac{1}{k^3R^3} + j\frac{1}{k^2R^2}\right)\mathrm{e}^{-jkR},$$

$$\dot{E}_\theta = \frac{k^3 d\dot{I}_m \sin\theta}{j\omega 4\pi\varepsilon}\left(\frac{1}{k^3R^3} + j\frac{1}{k^2R^2} - \frac{1}{kR}\right)\mathrm{e}^{-jkR}, \tag{4.10}$$

$$\dot{E}_\alpha = 0.$$

Ближней зоной излучающего устройства называют область, соответствующую расстояниям, значительно меньшим длины волны: $R/\lambda \ll 1$. Поле ближней зоны электрического диполя определяется членами с высшими степенями kR в уравнениях (4.9) и (4.10):

$$H_\alpha = \frac{dI_m \sin\theta}{4\pi R^2};\ E_R = \frac{p_m \cos\theta}{2\pi\varepsilon R^3};\ E_\theta = \frac{p_m \sin\theta}{4\pi\varepsilon R^3}.$$

Таким образом, поле в ближней зоне не имеет волнового характера. Пространственное распределение электрического поля совпадает с распределением статического поля диполя с моментом p_m, а распределение магнитного поля подчиняется закону Био—Савара. Поток энергии, переносимой ближним полем, в среднем равен нулю, так как электрическое и магнитное поля сдвинуты по фазе на 90°. Колебательный характер движения энергии около источника объясняется тем, что в последних уравнениях не учитывается излучение (поле излучения незначительно по сравнению с квазистационарным полем ближней зоны).

Дальней зоной излучателя называют область, соответствующую расстояниям, значительно превышающим длину волны: $R \gg \lambda$. Например, в дальней зоне диполя поле определяется только членами уравнений (4.9) и (4.10), которые пропорциональны R^{-1}:

$$H_\alpha = j\frac{kdI_m \sin\theta}{4\pi R}e^{-jkR};\quad E_\theta = j\frac{k^2 dI_m \sin\theta}{4\pi\omega\varepsilon R}e^{-jkR}. \qquad (4.11)$$

Эти выражения описывают поле излучения. Векторы **E** и **H** перпендикулярны направлению распространения излучения и друг другу, связаны между собой волновым сопротивлением (3.12), совпадают по фазе и образуют *сферическую волну* типа Т, распространяющуюся радиально. Напряженности поля уменьшаются пропорционально $1/R$, а энергия поля — $1/(4\pi R^2)$, так же как для любого точечного источника излучения. Напряженности поля пропорциональны второй производной электрического момента, т. е. первой производной или скорости изменения тока вибратора. Излучение является направленным: интенсивность его максимальна в направлении, перпендикулярном колеблющемуся диполю, и равна нулю в направлении дипольного момента. Плотность энергии, излучаемой в единицу времени, в различных направлениях неодинакова.

§ 4.3. Диаграмма направленности

Свойство антенны концентрировать излучаемую энергию в какомлибо одном направлении называют *направленным действием* антенны и характеризуют пространственной *диаграммой направленности*. Про-

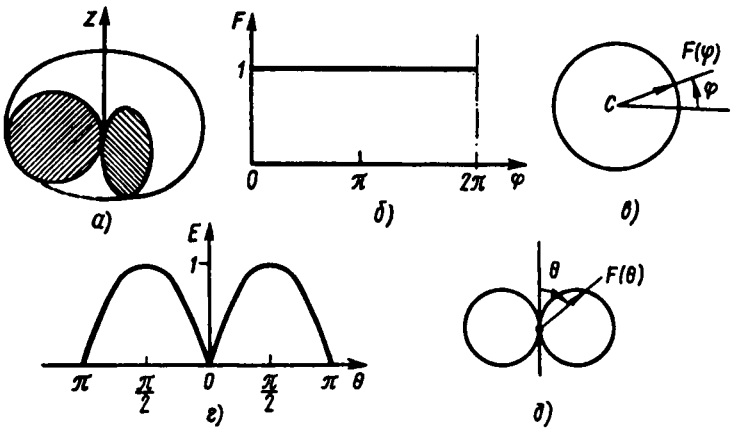


Рис. 4.1. Пространственная диаграмма направленности (*а*) электрического диполя и ее сечение в декартовых (*б*, *г*) и сферических (*в*, *д*) системах координат

странственная диаграмма направленности (по напряженности) изображается поверхностью, описываемой концом радиус-вектора, длина которого соответствует амплитуде напряженности излучения в данном направлении при условии измерения напряженности поля на одном и том же расстоянии от антенны. Обычно строят *нормированные* диаграммы направленности, принимая максимальный радиус-вектор равным единице. Диаграммы направленности строят в сферической, цилиндрической и декартовой системах координат.

Диаграмма направленности элементарного электрического излучателя в сферической системе координат может быть построена по уравнениям (4.11). Амплитуда напряженности электрического поля

$$E_m = \frac{k^2 dI_m}{4\pi\omega\varepsilon R}\sin\theta.$$

При фиксированном расстоянии $R \gg \lambda$ максимальное значение напряженности соответствует направлению $\theta = 90°$: $E_{max} = k^2\, dI_m/(4\pi\omega\varepsilon R)$. Следовательно, в данном случае нормированная пространственная диаграмма направленности выражается функцией $F(\theta,\varphi) = |E_m(\theta,\varphi)|/E_{max} = |\sin\theta|$, которая зависит только от одной координаты θ и дает в пространстве объемную фигуру в виде тора (рис. 4.1, *а*). Сечения пространственной диаграммы плоскостями, проходящими через начало координат, называют диаграммами направленности излучателя в соответствующих плоскостях. На рис. 4.1 приведены диаграммы направленности электрического диполя в экваториальной (*б*, *в*) и меридиональной (*г*, *д*) плоскостях.

§ 4.4. Радиационное «трение»

Мгновенное значение вектора Пойнтинга в дальней зоне электрического диполя, согласно (4.8) и (4.11)

$$\Pi(t) = E_\theta H_\alpha = \frac{\mu\omega^4 \sin^2\theta}{(4\pi)^2 \upsilon R^2} p^2 \sin^2(\omega t - kR). \qquad (4.12)$$

Его среднее значение за период $\Pi = \dfrac{\mu\omega^4 p_m^2 \sin^2\theta}{2(4\pi)^2 \upsilon R^2}$.

Средняя мощность, излучаемая во внешнее пространство через сферический слой площадью $\mathrm{d}S = 2\pi R^2 \sin\theta \mathrm{d}\theta$, равна $\Pi \mathrm{d}S$. Среднее значение полной мощности, излучаемой вибратором

$$P = \oint \Pi \, \mathrm{d}S = \int_0^\pi \frac{\mu\omega^4 p_m^2}{2(4\pi)^2 \upsilon R^2} 2\pi R^2 \sin^3\theta \, \mathrm{d}\theta = \frac{\mu\omega^4 p_m^2}{12\pi\upsilon}. \qquad (4.13)$$

Мгновенные значения всей излучаемой мощности без учета запаздывания

$$P(t) = \frac{\mu\omega^4}{6\pi\upsilon} p^2(t). \qquad (4.14)$$

Среднюю мощность можно характеризовать с помощью так называемого эквивалентного сопротивления излучения $R_{изл}$, в котором поглощается и расходуется на нагрев та же мощность

$$P = I_m^2 R_{изл}/2.$$

По уравнениям (4.8), (4.13) и (4.14) сопротивление излучения дипольного вибратора

$$R_{изл} = 80\pi^2 (d/\lambda)^2. \qquad (4.15)$$

Мощность излучения дипольного вибратора можно также связать с эквивалентной *силой радиационного трения* F, ослабляющей колебания диполя. При этом обычно предполагается, что дипольный момент изменяется не вследствие вариации его заряда, а вследствие изменения его длины $d = d_m \sin\omega t$. Тогда в уравнении (4.14) можно произвести следующую замену:

$$\omega^2 p(t) = \omega^2 q d(t) = q \frac{\mathrm{d}^2}{\mathrm{d}t^2} d(t).$$

Работа, произведенная при изменении длины диполя против действия силы F, равна энергии, отданной за интервал времени от 0 до t:

$$\int F\,\mathrm{d}(d) = \int_0^t F\frac{\mathrm{d}}{\mathrm{d}t}d\,\mathrm{d}t = \int_0^t P(t)\,\mathrm{d}t = \int_0^t \frac{\mu q^2}{6\pi v}\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2}d\right)^2 \mathrm{d}t. \qquad (4.16)$$

Проинтегрировав (4.16) по частям, получим

$$\int F\,\mathrm{d}(d) = -\int_0^t \frac{\mu q^2}{6\pi v}\left(\frac{\mathrm{d}^3}{\mathrm{d}t^3}d\right)\left(\frac{\mathrm{d}}{\mathrm{d}t}d\right)\mathrm{d}t + \left[\frac{\mu q^2}{6\pi v}\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2}d\right)\right]_0^t.$$

При малом затухании второе слагаемое обращается в нуль. Тогда сила радиационного трения

$$F = -\frac{\mu q^2}{6\pi v}\frac{\mathrm{d}^3}{\mathrm{d}t^3}d. \qquad (4.17)$$

В противоположность обычным силам трения, которые изменяются пропорционально первой производной радиус-вектора (пропорционально скорости движения), сила радиационного трения изменяется пропорционально третьей производной длины диполя (пропорционально скорости изменения поля).

Процессы электромагнитного излучения обусловлены переходом микрочастиц с одного энергетического уровня на другой. Для строгой количественной оценки излучения требуется суммирование излучения всех микрочастиц по законам квантовой статистики. Однако, когда число фотонов достаточно велико, эту задачу можно решить значительно проще, если, следуя предложению А. Эйнштейна, представить атомы и молекулы вещества в виде дискретного набора осцилляторов, каждый из которых обладает собственной частотой колебаний. При этом создаваемое этими осцилляторами электромагнитное поле можно считать классическим, т. е. представляющим волновой процесс со сплошным неограниченным спектром, плотность которого подчиняется закону Планка. Осуществляя статистическое усреднение и суммирование энергии излучения одного осциллятора, можно найти излучение тела или антенны.

Излучение атома или молекулы можно условно представить как излучение, вызванное колебаниями электронов относительно ядра. Так как вплоть до ближайшей ультрафиолетовой области длины излучаемых волн ($\lambda \geqslant 10^{-6}$ см) значительно больше размеров атома (порядка 10^{-8} см) и меньше расстояний, на которых обычно наблюдается электромагнитное поле, то соотношения (4.9)—(4.15) можно применить с достаточной точностью и для атомного диполя, трактуемого как классический осциллятор.

Электромагнитное излучение на основе квантовой теории и связь между классическим и квантовым осциллятором рассмотрено в гл. 11.

§ 4.5. Спектральное представление поля излучения

Ранее были рассмотрены различные виды электромагнитных волн (плоские, цилиндрические, сферические), каждой из которых соответствует частный случай решения волнового уравнения. Для получения решения, соответствующего волнам с произвольной формой фазовой поверхности, можно воспользоваться в качестве исходного плоской волной, распределение которой в пространстве характеризует экспонента (1.1). Эта экспонента входит в выражение комплексных амплитуд напряженностей как электрического, так и магнитного поля. Для большей определенности будем исходить из комплексной амплитуды вектора электрического поля (1.2):

$$\dot{E}(x, y, z) = \dot{E}_0 e^{jkR} = E_0 e^{j(\varphi + k_x x + k_y y + k_z z)} = \dot{E}_0 e^{j(k_x x + k_y y)} e^{\pm j\sqrt{k^2 - k_x^2 - k_y^2}}. \quad (4.18)$$

Комплексное выражение (4.18) несет полную информацию о плоской волне. Его модуль имеет физический смысл *амплитуды волны*, а аргумент — *фазы колебаний. Длина волны* может быть найдена из (2.1). Три проекции волнового вектора ($k_x = k\cos\alpha$, $k_y = k\cos\beta$, $k_z = k\cos\gamma$) определяют *направление* распространения волны (α, β, γ). Однако если известна длина волны, то для задания направления достаточно знать только две проекции k_x и k_y, так как $k_z = \sqrt{k^2 - k_x^2 - k_y^2}$.

Существенно, что выражение (4.18) представляет собой решение волнового уравнения при любых k_x и k_y, т. е. они могут иметь значения от $-\infty$ до ∞. Физический смысл проекции волнового вектора однородной плоской волны очевидно будут иметь только при

$$\chi^2 = k_x^2 + k_y^2 \leqslant k^2. \quad (4.19)$$

При других значениях k_x и k_y решению волнового уравнения будет соответствовать неоднородная плоская волна, затухающая по крайней мере вдоль одной из координат. Чтобы подчеркнуть более широкий смысл этих компонент волнового вектора, целесообразно перейти в дальнейшем к их новым обозначениям: $u = k_x$ и $v = k_y$.

В линейной среде волновое уравнение линейно и для него справедлив принцип суперпозиции. Это позволяет утверждать, что поскольку поле плоской волны с произвольными параметрами описывает решение волнового уравнения, то и сумма полей плоских волн вида (4.18), также будет решением волнового уравнения. Если включить в эту сумму волны со всеми возможными направлениями распространения и значениями u и v, то получим общее решение в виде интеграла

$$\dot{E}(x, y, z) = \frac{1}{4\pi^2} \int_{-\infty}^{+\infty}\int \dot{\mathscr{E}}(u, v) e^{j\sqrt{k^2-u^2-v^2}} e^{j(ux+vy)} du\, dv, \qquad (4.20)$$

где $\dot{\mathscr{E}}(u, v)\,du\,dv/(4\pi^2)$ — начальная комплексная амплитуда (при $x = y = z = 0$) семейства плоских волн, направление распространения которых определяется значениями двух действительных переменных, находящихся в интервалах $(u, u + du)$ и $(v, v + dv)$.

Следовательно, $\dot{\mathscr{E}}(u, v)$ является функцией распределения или спектральной функцией комплексных амплитуд плоских волн с различными направлениями ($u = k\cos\alpha$, $v = k\cos\beta$). Ее называют *спектром волнового поля* или *угловым спектром поля*. Название «угловой спектр» отражает связь аргументов u и v с углами распространения (α и β) соответствующих парциальных плоских волн. Отметим, что комплексные амплитуды парциальных волн, хотя и не равны спектральной функции, прямо пропорциональны ей.

Замечательной особенностью общего решения волнового уравнения является открываемая им возможность описать поле (4.20) во всех точках (x, y, z) по известному его решению $\dot{E}(x, y, 0)$ в плоскости $z = 0$.

Это следует из того, что, согласно (4.20), на поверхности $z = 0$ напряженность электрического поля

$$\dot{E}(x, y, 0) = \frac{1}{4\pi^2} \int_{-\infty}^{+\infty}\int \dot{\mathscr{E}}(u, v) e^{j(ux+vy)} du\, dv \qquad (4.21)$$

представляет собой интеграл Фурье от функции распределения $\dot{\mathscr{E}}(u, v)$. Таким образом, эта функция может быть получена из известной функции $\dot{E}(x, y, 0)$ с помощью обратного интегрального преобразования, т. е. преобразования Фурье

$$\dot{\mathscr{E}}(u, v) = \int_{-\infty}^{+\infty}\int \dot{E}(x, y, 0) e^{-j(ux+vy)} dx\, dy. \qquad (4.22)$$

Итак, зная $\dot{E}(x, y, 0)$ и определив (4.22), можно по (4.20) найти напряженность поля в любой точке $\dot{E}(x, y, z)$. Изложенный метод решения широко используется при решении многих вопросов дифракции. В ряде случаев он дает удовлетворительные результаты и при приближенном анализе антенн СВЧ сложной конфигурации, подобных, например, рупорной, зеркальной и щелевой.

В электродинамике встречаются задачи, в которых параметры поля на начальной плоскости не зависят от одной из координат, например y. В

этом случае $E(x, y, 0) = E(x, 0)$ и согласно (4.22)

$$\mathscr{E}(u, v) = 2\pi\delta(v)\mathscr{E}(u), \tag{4.23}$$

где

$$\mathscr{E}(u) = \int_{-\infty}^{\infty} E(x, 0)^{-jux}\, dx \tag{4.24}$$

— функция только переменной u.

Подставив выражение (4.23) в (4.21) и проинтегрировав полученное соотношение по v, нетрудно убедиться, что в этом случае напряженность не зависит от координаты y и при любом z,

$$E(x, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathscr{E}(u) e^{jz\sqrt{k^2-u^2}} e^{jux}\, du. \tag{4.25}$$

В частности, при $z = 0$

$$E(x, 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathscr{E}(u) e^{jux}\, du. \tag{4.26}$$

Выражения (4.25) и (4.26) представляют собой интегралы Фурье для переменных u и x. Переменная x является координатой точек пространства и имеет размерность длины. Переменная u имеет размерность, обратную длине.

§ 4.6. Аналогия между передачей электрических сигналов в цепях и свободном пространстве

Аналогия между колебательными процессами в цепях с сосредоточенными параметрами и распространением волн в свободном пространстве основывается на том, что как те, так и другие описываются линейными дифференциальными уравнениями и происходят в так называемых *линейных системах*. В этих системах выполняются принципы суперпозиции и транспозиции, благодаря которым реакцию или отклик системы на некоторое входное воздействие можно представить в виде суммы реакций на какие-либо простые входные воздействия.

Если исходить из простого *гармонического* воздействия в виде функции $e^{j\omega t}$, то реакцию линейной системы на такое входное воздействие называют *частотной характеристикой системы* $G(\omega)$ *. Если она

---

* Если $\dot{X}(t) = e^{j\omega t}$, то $\dot{Y}(t) = G(\omega) e^{j\omega t}$.

известна, то при произвольном воздействии $x(t)$ на входе реакция линейной системы

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) G(\omega) e^{j\omega t} d\omega, \tag{4.27}$$

где

$$F(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \tag{4.28}$$

— *спектр входного воздействия*, называемый также спектральной плотностью или спектральной характеристикой функции $x(t)$.

Спектр входного воздействия определяет амплитуды и фазы тех простейших воздействий $e^{j\omega t}$, сумма которых составляет входной сигнал

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega. \tag{4.29}$$

Каждое из этих единичных простейших воздействий вызывает отклик $G(\omega) e^{j\omega t}$. Изменение их амплитуды и фазы, определяемых (4.28), передается на выходной сигнал в виде таких же изменений $F(\omega) G(\omega) e^{j\omega t}$. Выражение (4.27) получено в результате суммирования этих выходных сигналов согласно принципам суперпозиции и транспозиции. Четырехполюсник, свойства которого описывает уравнение (4.27), можно представить в виде функциональной схемы, изображенной на рис. 4.2, *а*.

Сравнение выражений (4.27) и (4.28) с (4.24) и (4.25) делает очевидным их аналогию. Поскольку функция $\dot{E}(x, y)$ представляет собой комплексную амплитуду напряженности поля, прошедшего путь z в свободном пространстве, то естественно ее считать *откликом* свободного пространства на входной сигнал $\dot{E}(x, 0)$, действующий в начале («на входе») этого участка при $z = 0$. При этом функция $\dot{\mathcal{E}}(u)$, аналогично $F(\omega)$, является спектром входного сигнала. Переменная интегрирования u, входящая в качестве аргумента в разложение в интеграле Фурье (4.26) входно-

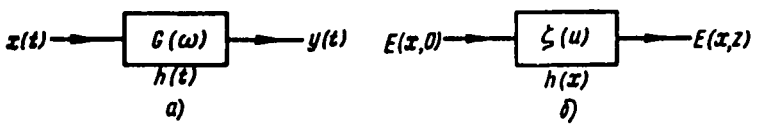


Рис. 4.2. Функциональные схемы линейных систем:
а — цепи с сосредоточенными параметрами; *б* — ее волнового аналога

го пространственного распределения $\dot{E}(x, 0)$, является аналогом частоты ω в разложении (4.28). Поэтому она получила название *пространственной частоты.*

Из сравнения выражений (4.25) и (4.27) следует, что функция $\zeta(u) = e^{jz\sqrt{k^2-u^2}}$ есть аналог $G(\omega)$, т. е. является *частотной характеристикой* волнового поля в свободном пространстве (рис. 4.2, *б*).

Роль входного сигнала в этой линейной системе играет распределение в плоскости $z = 0$ комплексных амплитуд в зависимости только от пространственных координат (x), а роль выходного сигнала — распределение комплексных амплитуд в некоторой плоскости z. Особенностью такого «фильтра» пространственных частот u является то, что его частотная характеристика $\zeta(u)$ зависит от координаты z. Особенностью волновых полей является также двумерный характер их спектральных характеристик. Так двумерный фурье-образ (4.22) амплитудно-фазового распределения поля (4.21) представляет собой угловой спектр волнового пучка в плоскости x, y или его угловую спектральную плотность, или его пространственно-частотный спектр.

Комплексный угловой спектр (4.22) пучка (4.21) может быть записан также в виде двумерного преобразования Лапласа

$$E(p, q) = \int_{-\infty}^{+\infty}\int \dot{E}(x, y, 0)\, e^{-px} e^{-qy}\, dx\, dy,$$

где $p = -ju$; $q = -jv$ — комплексные поперечные волновые числа.

Двумерная угловая фильтрация может быть названа и двусторонней, поскольку связана с сужением как положительных, так и отрицательных составляющих углового спектра.

Изложенное позволяет сделать вывод о том, что преобразование формы зависимости комплексных напряженностей волнового поля от поперечных координат можно описать аналогично тому, как описывается преобразование сигналов сосредоточенными линейными системами в теории связи, электронике и теории колебаний.

Важнейшим параметром рассматриваемого преобразования является частотная характеристика свободного пространства. Рассмотрим ее основные свойства.

Принципиальной особенностью преобразования сигналов в свободном пространстве является прежде всего двумерность этого преобразования. Согласно (4.20) и (4.27), частотная характеристика свободного пространства

$$\zeta(u, v) = e^{jz\sqrt{k^2-u^2-v^2}} \qquad (4.30)$$

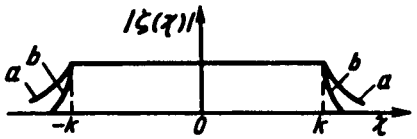


Рис. 4.3. Амплитудно-частотная характеристика свободного пространства

в общем случае представляет собой двумерную функцию от u и v.

Амплитуда частотной характеристики

$$|\zeta(u,v)| = \begin{cases} 1 & \text{при} \quad \chi^2 = u^2 + v^2 \leqslant k^2 \\ e^{-z\sqrt{\chi^2 - k^2}} & \text{при} \quad \chi^2 > k^2 \end{cases} \tag{4.31}$$

имеет постоянное значение $|\zeta| = 1$ в области однородных волн (полная прозрачность фильтра) и экспоненциально затухает в области неоднородных волн (рис. 4.3). Затухание тем больше, чем больше пространственная частота. В пределе для достаточно больших пространственных частот ($\chi \to \infty$) амплитуда частотной характеристики стремится к нулю. Скорость затухания частотной характеристики зависит также от расстояния z. При увеличении z форма амплитудной характеристики приближается к прямоугольной.

Нетрудно оценить расстояние $z_п$, при котором амплитудная характеристика становится практически прямоугольной. При этом обычно исходят из условия, что при $z_п$ спад характеристики происходит в e раз и соответствует интервалу поперечных частот $\Delta\chi = \chi_п - k \leqslant k$; т. е. $z_п\sqrt{\chi_п^2 - k^2} = z_п\sqrt{2k\Delta\chi + \Delta\chi^2} \approx z_п\sqrt{2k\Delta\chi} = 1$, откуда $\Delta\chi \approx 1/(2kz_п^2) \gg k$ или $z_п \gg 1/(k\sqrt{2}) = \lambda/(2\pi\sqrt{2})$.

Таким образом, *при расстояниях, больших длины волны, свободное пространство ведет себя как прямоугольный фильтр, не пропускающий неоднородные волны и прозрачный для однородных плоских волн.*

Фаза частотной характеристики

$$f(u,v) = z\sqrt{k^2 - \chi^2} = kz\left(1 - \frac{1}{2}\frac{\chi^2}{k^2} + \frac{1}{2^3}\frac{\chi^4}{k^4} - \ldots\right) \tag{4.32}$$

также зависит от пространственной (угловой) частоты и расстояния z.

Наиболее простой вид частотная характеристика приобретает в том случае, когда можно ограничиться только первым членом разложения (4.32). Для этого необходимо, чтобы второй член был много меньше π, т. е.

$$z\chi^2/(2k) = z(u^2 + v^2)/(2k) \ll \pi. \tag{4.33}$$

В этом случае частотная характеристика

$$\zeta(u,v) = e^{jkz} = \zeta_0 \tag{4.34}$$

вообще не зависит от переменных u и v и, согласно (4.20) и (4.21),

$$\dot{E}(x,y,z) = \dot{E}(x,y,0)e^{jkz}.$$

Таким образом, независимо от характера распределения поля в начальной плоскости $z = 0$ поле переносится в любое другое сечение $z > 0$ без изменений, все точки поперечной структуры волны распространяются в одном направлении с одинаковой скоростью ($v = \omega/k$) (параллельно оси z). Следовательно, данный случай характеризует то прямолинейное распространение всех лучей, проходящих через начальную плоскость, которое соответствует геометрической оптике, т. е. условие (4.33) определяет область применимости геометрической оптики.

§ 4.7. Импульсная характеристика свободного пространства

В теории цепей реакцию линейной системы $y(t)$ на произвольное воздействие $x(t)$ описывают не только с помощью частотной характеристики $G(\omega)$, но и с помощью импульсной характеристики $h(t)$:

$$y(t) = \int_{-\infty}^{\infty} x(t)h(t-\tau)\,d\tau. \qquad (4.35)$$

Если частотная характеристика является реакцией системы на элементарное воздействие вида $e^{j\omega t}$, то импульсная характеристика представляет собой отклик системы (4.27) и на δ-функцию

$$h(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F_\delta(\omega)G(\omega)e^{j\omega t}\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\omega)e^{j\omega t}\,d\omega, \qquad (4.36)$$

так как спектр (4.28) δ-функции

$$F_\delta(\omega) = \int_{-\infty}^{\infty}\delta(t)e^{-j\omega t}\,dt = 1.$$

Аналогия между передачей сигналов в электрических цепях и в свободном пространстве позволяет реакцию волнового поля в свободном пространстве на входное воздействие (4.21) рассматривать как суперпозицию откликов на все импульсные составляющие входного сигнала:

$$\dot{E}(x, y, z) = \int_{-\infty}^{+\infty}\int \dot{E}(x', y', 0)h(x-x', y-y', z)\,dx'\,dy', \qquad (4.37)$$

где функция

$$h(x, y, z) = \frac{1}{4\pi^2}\int_{-\infty}^{+\infty}\int \zeta(u, v)e^{j(ux+vy)}\,du\,dv \qquad (4.38)$$

— импульсная характеристика свободного пространства или функция рассеяния точки.

В волновом поле δ-функция представляет собой математическое описание абстрактного нормированного точечного излучателя, т. е. излучателя конечной постоянной мощности $P_\Sigma = 1$, когда его излучающая поверхность Σ стягивается к началу координат и интенсивность (плотность) излучения

$$I(x, y, z) \doteq \dot{E}\dot{E}^* = E^2 = \lim_{\Sigma \to 0} P_\Sigma / \Sigma = \delta(x, y, z).$$

Очевидно, что точечный излучатель образует в окружающем свободном пространстве поле сферической волны, описываемой с точностью до постоянного множителя функцией Грина (4.4)

$$\dot{E}(x, y, z) \doteq \frac{1}{R} e^{jkR}, \tag{4.39}$$

а в начальной плоскости $z = 0$, где $R = r = \sqrt{x^2 + y^2}$,

$$\dot{E}(x, y, 0) \doteq \frac{1}{r} e^{jkr}. \tag{4.40}$$

Угловой спектр такого излучателя можно получить из (4.22) и (4.40) с помощью замены переменных $-u$ на $q\cos\psi$, $-v$ на $q\sin\psi$, x на $r\cos\varphi$, y на $q\sin\varphi$:

$$\dot{\mathfrak{E}}(u, v) = \int_{-\infty}^{+\infty}\int e^{jkr} e^{-j(ux+vy)} \frac{dx\, dy}{r} = j\frac{2\pi}{\sqrt{k^2 - u^2 - v^2}} = j\frac{2\pi}{k_z}. \tag{4.41}$$

Согласно (4.20) и (4.41), раложение сферической волны по плоским волнам имеет вид

$$\frac{1}{R} e^{jkR} = \frac{j}{2\pi} \int_{-\infty}^{\infty}\int \frac{\exp[j(ux + vy + z\sqrt{k^2 - u^2 - v^2})]}{\sqrt{k^2 - u^2 - v^2}} du\, dv, \tag{4.42}$$

откуда

$$\frac{\partial}{\partial z}\left(\frac{1}{R} e^{jkR}\right) = -\frac{1}{2\pi} \int_{-\infty}^{+\infty}\int e^{j(ux+vy+z\sqrt{k^2-u^2-v^2})} du\, dv.$$

Следовательно, импульсную характеристику свободного пространства с учетом (4.30), (4.38) и (4.42) можно представить в общем случае в виде

$$h(x, t) = \frac{1}{4\pi^2} \int_{-\infty}^{+\infty}\int e^{jz\sqrt{k^2-u^2-v^2}} e^{j(xu+yv)} du\, dv =$$

$$= -\frac{1}{2\pi}\frac{\partial}{\partial z}\left(\frac{1}{R} e^{jkR}\right) = \frac{1}{2\pi R}\left(\frac{1}{R} - jk\right) e^{jkR} \frac{\partial R}{\partial z}. \tag{4.43}$$

Интересен тот частный случай, когда выполнены условия $R \gg 1/k = \lambda/2\pi$, $\gamma \ll 1$, т. е.

$$z \approx R \text{ и } \partial R/\partial z = z\Big/\sqrt{x^2+y^2+z^2} \approx 1, \tag{4.44}$$

и в выражении (4.43) можно пренебречь первым слагаемым. Тогда

$$h(x, y) \approx \frac{1}{j\lambda R} e^{jkR} \tag{4.45}$$

и, согласно соотношению (4.37),

$$\dot{E}(x, y, z) = \frac{1}{j\lambda} \int_{-\infty}^{\infty}\int \frac{\dot{E}(x', y', 0)\exp[jk\sqrt{(x-x')^2+(y-y')^2+z^2}]}{\sqrt{(x-x')^2+(y-y')^2+z^2}} dx'dy'. \tag{4.46}$$

Полученное выражение представляет собой математическую запись принципа Гюйгенса—Френеля, согласно которому распространение волны можно рассматривать как процесс суммирования волн, излучаемых точечными источниками, расположенными на некоторой поверхности; их амплитуды и фазы совпадают с амплитудами и фазами волн, проходящих через эту поверхность. В данном случае эта поверхность совмещена с плоскостью $z = 0$. Поле в рассматриваемой точке можно представить, согласно (4.46), как сумму сферических волн с амплитудами и фазами, определяемыми функцией $E(x', y', 0)$, причем каждая из этих волн излучается источником, расположенным в точке $(x', y', 0)$.

Частотную характеристику свободного пространства в условиях применимости принципа Гюйгенса—Френеля можно получить, исходя из (4.45) и выполняя преобразование, обратное преобразованию (4.38):

$$\zeta_{\text{ГФ}}(u, v) = \int_{-\infty}^{+\infty}\int h(x, y) e^{-j(ux+vy)} dx\, dy \approx$$

$$\approx \frac{k}{k_z} e^{jk_z z} = \frac{k}{\sqrt{k^2-u^2-v^2}} e^{jz\sqrt{k^2-u^2-v^2}}. \tag{4.47}$$

Из сравнения (4.30) и (4.41) следует, что для того, чтобы точное выражение частотной характеристики (4.30) перешло в приближенное (4.47), необходимо и достаточно

$$\chi^2 = u^2 + v^2 \ll k^2. \tag{4.48}$$

Это означает, что условие (4.48) эквивалентно введенным ранее двум условиям (4.44).

В заключение отметим, что расчет реакции линейной системы с помощью импульсной характеристики совпадает с интегральной операцией (4.35), называемой *сверткой,* и сокращенно обозначается с помощью операторного символа $\circledast$: $y(t) = x(t) \circledast h(t)$.

Следовательно, реакция свободного пространства (4.37) на входной сигнал

$$E(x, y, z) = E(x, y, 0) \circledast h(x, y, z). \qquad (4.49)$$

§ 4.8. Дифракция

Результаты анализа общего решения волнового уравнения находят широкое применение при решении дифракционных задач. Типичная дифракционная задача предполагает, что в поле волны помещено некоторое препятствие, геометрические параметры которого и электрические свойства материала известны. Требуется определить электромагнитное поле во всем пространстве.

При спектральном подходе дифракция может быть представлена совокупностью двух конкурирующих явлений: пространственной модуляции (изменения начального амплитудно-фазового распределения поля), приводящей к расширению углового спектра начального пучка, и угловой фильтрации, в результате которой происходит сужение углового спектра.

Точное решение задачи в общем случае затруднено, поскольку оно связано с решением волновых уравнений при сложных граничных условиях на поверхности препятствий. Однако в теории дифракции во многих случаях рассматриваются задачи, допускающие приближенное решение. При решении подобных задач должны выполняться два условия: 1) размеры препятствий, экранов и отверстий в них, а также расстояния между ними, точками излучения и наблюдения должны быть велики по сравнению с длиной волны; 2) отклонения от направления лучей в геометрической оптике должны быть небольшими. В тех задачах, где рассматривается поле на больших расстояниях от препятствия, перечисленные условия обычно выполняются. Дальнейшее изложение будет посвящено именно таким задачам.

Главная трудность при точном решении дифракционных задач заключается в определении неизвестного распределения электромагнитного поля на поверхности препятствия и вблизи него. В приближенных решениях используется то обстоятельство, что при длине волны, много меньшей линейных размеров препятствия и радиусов кривизны его поверхности, электромагнитное поле подчиняется законам, близким к законам геометрической оптики. В этом случае любой малый элемент фронта

первичной волны, падающий на граничную поверхность произвольной формы, ведет себя приблизительно так же, как и элемент фронта волны, падающей на плоскую поверхность. Это обстоятельство дало основание положить в основу большинства волновых методов приближенных расчетов дифракционных задач допущение, что при определении значений векторов вторичного поля на граничной поверхности (и только для них) можно в порядке первого приближения использовать законы геометрической оптики, рассчитывая поле в остальном пространстве уже на основании волновой теории.

Наиболее отчетливо такой ход решения выражает *приближение Кирхгофа*. Оно состоит в том, что поле на препятствии считается равным нулю, а вне препятствия — таким же, как и в отсутствие препятствия.

Решение выполняют обычно в такой последовательности. Сначала проводят плоскость на возможно близком расстоянии от препятствия со стороны прошедшей волны. Затем на этой плоскости находят поле волны, исходя из приближения Кирхгофа. С этой плоскостью совмещают плоскость $z = 0$. Далее определяют спектр пространственных частот (4.22) и, наконец, искомое поле (4.20) во всем пространстве.

Приближение Кирхгофа является грубым, так как вводимое им в рассмотрение поле не удовлетворяет граничным условиям в плоскости $z = 0$. Тем не менее оказывается, что получающиеся результаты всегда находятся в удовлетворительном согласии и с экспериментом, и со строгим решением на достаточно большом расстоянии.

Это можно объяснить фильтрующим свойством свободного пространства. Частотная характеристика (4.30) этого фильтра при больших z такова, что он пропускает только те пространственные частоты, которые попадают в некоторую узкую полосу ($\chi^2 \ll k^2$). При этом оказывается, что фильтр не пропускает как раз те частоты, которые учитывают отличие приближения Кирхгофа от истинных параметров полей. Следовательно, имеют место два эффекта: во-первых, отсекаются частоты исходного поля, на которых оно явно не соответствует действительности, и, во-вторых, в той области частот, где приближение оказывается приемлемым, происходит преобразование поля без искажения.

Различают два вида дифракции — на сравнительно близких и далеких расстояниях.

Дифракцией Френеля называют явления, которые возникают, когда источник света и точка наблюдения находятся на конечном расстоянии от экрана. При этом поле в точке наблюдения формирует лишь небольшие участки волновой поверхности вблизи прямой, которая соединяет источник с точкой наблюдения.

Дифракция Френеля соответствует тому диапазону расстояний и пространственных частот, при котором в выражении фазы частотной ха-

рактеристики свободного пространства (4.30) можно ограничиться первыми двумя членами разложения (4.32), т. е. когда

$$\zeta(u, v) \approx e^{jkz} e^{-j\chi^2 z/(2k)} = \zeta_0 e^{-j\chi^2 z/(2k)}. \quad (4.50)$$

Таким образом, условием дифракции Френеля, согласно (4.32), является $z\chi^4/(8k^3) \ll \pi$ или

$$z \ll 8\pi k^3/\chi^4 = 8\pi k^3/(u^2+v^2)^2 = 4\lambda/\sin^4\gamma. \quad (4.51)$$

К одному из типичных примеров дифракции Френеля относят явления на границе света и тени в пределах конечных расстояний от препятствия (4.51).

Поскольку дифракцию Френеля определяют небольшие участки волновой поверхности, прилегающей к краям препятствий вблизи оси z, то при решении дифракционной задачи на большом расстоянии от препятствия вблизи оси z ($R \approx z \gg \lambda$), когда

$$x^2 + y^2 \ll z^2, \quad (4.52)$$

можно произвести дальнейшее упрощение интегрального представления волнового поля в виде (4.46). Например, при рассмотрении дифракции от отверстия в непрозрачном экране на расстояниях, бóльших линейных размеров отверстия с поверхностью S, когда элемент $dS = dx'dy'$ пробегает область интегрирования, радикал в знаменателе (4.46) изменяется незначительно; его можно заменить на z и вынести из-под знака интегралов. В то же время множитель kR в общем случае изменяется на много длин волн и поэтому экспонента быстро осциллирует, причем согласно (4.52)

$$R = \sqrt{(x-x')^2 + (y-y')^2 + z^2} \approx z + \frac{(x-x')^2 + (y-y')^2}{2z}. \quad (4.53)$$

Следовательно, напряженность дифрагированного поля

$$\dot{E}(x, y, z) = \frac{1}{j\lambda z} e^{jkz} \iint_S \dot{E}(x', y', 0)\, e^{jk[(x-x')^2+(y-y')^2]/(2z)}\, dx'\, dy' \quad (4.54)$$

и импульсная характеристика свободного пространства

$$h(x, y, z) \approx \frac{1}{j\lambda z} e^{jkz} e^{jk(x^2+y^2)/(2z)} \quad (4.55)$$

выражены через функцию Френеля

$$z(t) = \omega_0 \sqrt{2}\, e^{j\omega_0^2 t^2} \quad (4.56)$$

и *преобразование Френеля*

$$\Phi_n(t) = \omega_0 \sqrt{2} \int_{-\infty}^{\infty} f(\tau) e^{j\omega_0^2 (t-\tau)}\, d\tau. \quad (4.57)$$

Преобразование Френеля играет в оптике и голографии такую же роль, как и преобразование Фурье.

Очевидно, что в свободном пространстве бесконечно малой глубины осуществляется преобразование Дирака

$$\lim_{z\to 0} h(x, y, z) = \delta(x, y, z), \tag{4.58}$$

т. е. «мелкое» свободное пространство не вызывает дифракционного размытия сигнала.

Волновые явления на больших расстояниях позади экрана называют *дифракцией Фраунгофера.* В этом случае решение дифракционных задач упрощается при выполнении неравенства, обратного по знаку условию применимости геометрической оптики (4.33),

$$z\chi^2/(2k) = z(u^2 + v^2)/(2k) \gg \pi. \tag{4.59}$$

При выполнении условия (4.59) упрощение расчета достигается не вследствие упрощения частотной характеристики (4.30), а благодаря возможности использования *метода стационарной фазы* для приближенной оценки дифракционного интеграла (4.20). Этот метод утверждает, что

$$I = \int_{-\infty}^{\infty} F(x) e^{j\varphi(x)} dx \approx F(x_0) e^{j\varphi(x_0)} \int_{-\infty}^{\infty} e^{j0{,}5(x-x_0)^2 \varphi''(x_0)} dx =$$

$$= F(x_0) \sqrt{\frac{2\pi}{\mp\varphi''(x_0)}}\, e^{j[\varphi(x_0) \mp \pi/4]}, \tag{4.60}$$

если x_0 — корень уравнения $\varphi'(x) = 0$, $(x - x_0')^3 \varphi'''(x_0)/6 \ll \pi$ и $F(x)$ медленно изменяется на вещественном интервале $\Delta x = \sqrt{2\pi/\mp\varphi''(x_0)}$. В интеграле (см. формулу (4.25)) $F(x) = \dot{\mathcal{E}}(u)$, $\varphi(x) = z\sqrt{k^2 - u^2} + ux$, $u_0 = xk/\sqrt{z^2 + k^2} = xk/R$, поэтому

$$\dot{E}(x, z) \approx \frac{1 + j}{\sqrt{2\lambda R}} e^{jkr}\, \dot{\mathcal{E}}\left(\frac{kx}{R}\right). \tag{4.61}$$

Особенностью дифракции Фраунгофера является то, что распределение напряженности поля на далеких расстояниях от препятствия $\dot{E}(x, z)$ пропорционально спектру исходного поля $\dot{\mathcal{E}}(u)$ на препятствии. В частности,

$$\frac{\dot{E}(x, z)}{\dot{E}(0, z)} = \frac{\dot{\mathcal{E}}(kx/R)}{\dot{\mathcal{E}}(0)}. \tag{4.62}$$

При этом оказывается, что радиус-вектор R точки, в которой находится дифрагированное поле (4.61), наклонен к первоначальному направ-

лению распространения волн $\mathbf{e}_z$ под тем же углом $\theta\,(\sin\theta = xR)$, единственная пространственная частота которого $u = k\sin\gamma = kx/R$ определяет плотность спектра ($\dot{\mathcal{E}}(u)$) в рассматриваемой точке ($\theta = \gamma$). Другими словами, дифрагированное поле в любой точке характеризуется только одним лучом, который идет от центра препятствия ($x = z = 0$) в эту точку. Частотные компоненты оказываются разделенными (в пространстве).

При дифракции Френеля частотные компоненты u, v взаимно перемешиваются. Поле в каждой точке определяется совместным действием семейства гармонических компонент.

Рассмотрим несколько примеров дифракции Фраунгофера.

§ 4.9. Дифракция на отверстиях

К типичным задачам, приближенно решаемым спектральным методом, относится дифракция волны на отверстии в плоском непрозрачном экране. Пусть на экран с отверстием падает однородная плоская волна амплитудой E_0 нормально к поверхности экрана. Необходимо определить поле за экраном. На теневой стороне экрана ($z = 0$) поле в соответствии с правилом Кирхгофа полагаем равным нулю, а на отверстии — равным полю падающей волны. Если найти начальную спектральную функцию $\dot{\mathcal{E}}(u, v)$, то тем самым будет определено поле за экраном в дальней зоне (4.61).

Дифракция на бесконечной щели. Рассмотрим щель шириной $2a = D$, расположенную в плоскости $z = 0$ параллельно оси y. В этом случае

$$E(x,0) = \begin{cases} E_0 & \text{при } |x| \leqslant a, \\ 0 & \text{при } |x| > a \end{cases}$$

и спектральная функция

$$\dot{\mathcal{E}}(u) = \int_{-\infty}^{\infty} E(x,0)\,\mathrm{e}^{-\mathrm{j}ux}\,\mathrm{d}x = \int_{-a}^{a} E_0\,\mathrm{e}^{-\mathrm{j}ux}\,\mathrm{d}x =$$

$$= 2aE_0\frac{\sin au}{au} = \dot{\mathcal{E}}(0)\frac{\sin ka\theta}{ka\theta} = \dot{\mathcal{E}}(\theta). \qquad (4.63)$$

Если параллельно плоскости щели расположить экран на достаточно большом расстоянии L, то в соответствии с (4.61) и (4.63) можно наблюдать (в оптическом диапазоне) чередование светлых и темных полос. Центральная полоса самая яркая, боковые имеют убывающую яркость.

Ширину A центральной полосы можно определить из условия $ka\theta_0 = \pi$, т. е. $A = 2\theta_0 L = 2\pi L/(ka) = \lambda L/D$.

Дифракция на круглом отверстии. Рассмотрим отверстие радиусом $a = D/2$, расположенное в плоскости $z = 0$. В этом случае

$$\dot{E}(x, y, 0) = \begin{cases} E_0 & \text{при } |r| \leqslant a, \\ 0 & \text{при } |r| > a, \end{cases}$$

где $\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y$. Если ввести в рассмотрение поперечный волновой вектор

$$\chi = u\mathbf{e}_x + v\,\mathbf{e}_y \tag{4.64}$$

и учесть, что $ux + vy = \chi\mathbf{r} = \chi r\cos\varphi$ и $\mathrm{d}S = \mathrm{d}y\,\mathrm{d}x = r\,\mathrm{d}\varphi\,\mathrm{d}r$ (φ — угол между векторами χ и $\mathbf{r}$), то спектральная функция

$$\dot{\mathscr{E}}(u, v) = \int_{-\infty}^{+\infty}\int \dot{E}(x, y, 0)\mathrm{e}^{-j(ux+vy)}\,\mathrm{d}x\,\mathrm{d}y = \int_0^a\int_0^{2\pi} E_0\,\mathrm{e}^{-j\chi r\cos\varphi}\,r\,\mathrm{d}r\,\mathrm{d}\varphi =$$

$$= 2\pi E_0\int_0^r rJ_0(\chi r)\,\mathrm{d}r = 2\pi\frac{a}{\chi}E_0 J_1(\chi a) = \dot{\mathscr{E}}(\theta),$$

так как $\chi = k\sin\theta \approx k\theta$. Полученное выражение определяет диаграмму направленности

$$F(\theta) = \frac{|F(\theta)|}{E_{\max}} = \frac{\left|\dot{\mathscr{E}}(\theta)\right|}{\left|\dot{\mathscr{E}}(0)\right|} = \frac{|J_1(ka\theta)|}{ka\theta/2}. \tag{4.65}$$

Следовательно, на удаленном на расстоянии L экране дифракционная картина имеет вид чередующихся ярких и темных колец. Радиус центрального, наиболее яркого пятна можно получить, приравнивая аргумент функции Бесселя первого порядка значению ее первого корня ($ka\theta_0 = 3{,}84$):

$$r_0 = \theta_0 L = 1{,}22\lambda L/D. \tag{4.66}$$

Рассмотренные примеры не только знакомят с расчетом поля излучения из отверстий в непрозрачных экранах. С их помощью можно также рассчитать поля поверхностных антенн с однородным синфазным полем в раскрывах рассмотренной формы. Например, решение задачи об излучении из круглого отверстия описывает также поле излучения рупорной, параболической и линзовой антенн. У реальных антенн поле неоднородно по амплитуде и фазе, что приводит к расширению главного лепестка диаграммы направленности по сравнению с исследованным случаем. Вместе с тем формула (4.65) дает оценку предельной направленности поверхностных антенн с круглым раскрывом.

Дифракция на синусоидальной решетке. Предположим, что решетка расположена параллельно оси y в плоскости $z = 0$. В этой плоскости за счет периодического изменения пропускания экрана напряженность

$$\dot{E}(x,0) = \frac{E_0}{2}\left(1 + m\cos\frac{2\pi x}{\Lambda}\right) = \frac{E_0}{2}\left(1 + \frac{m}{2}e^{j2\pi x/\Lambda} + \frac{m}{2}e^{-j2\pi x/\Lambda}\right), \quad (4.67)$$

где $m \leqslant 1$ — глубина модуляции; Λ — период решетки.

Спектральная функция

$$\dot{\mathscr{E}}(u) = \int_{-\infty}^{\infty}\dot{E}(x,0)e^{-jux}\,dx = \frac{E_0}{2}\int_{-\infty}^{\infty}\left[e^{-jux} + \frac{m}{2}e^{j(A-u)x} + \frac{m}{2}e^{-j(A+u)x}\right]dx =$$

$$= \pi E_0\left[\delta(-u) + \frac{m}{2}\delta(A-u) + \frac{m}{2}\delta(-A-u)\right] =$$

$$= \pi E_0\left[\delta(u) + \frac{m}{2}\delta(u-A) + \frac{m}{2}\delta(u+A)\right]$$

показывает, что дифрагированный свет состоит из трех лучей, соответствующих трем пространственным частотам: 0, A, $-A$. Первый центральный луч распространяется, не отклоняясь от направления исходной плоской волны ($\theta = 0$), а два боковых симметрично отклонены от центрального на углы

$$\theta = \pm\arcsin(A/k) = \pm\arcsin(\lambda/\Lambda) \approx \pm\lambda/\Lambda, \quad (4.68)$$

если $\Lambda \gg \lambda$.

Рассмотренная решетка образована экраном, который осуществляет *амплитудную модуляцию* падающей однородной плоской волны. Подобные решетки используют для снятия амплитудной и частотной характеристик оптических анализаторов спектра.

Перейдем к рассмотрению решетки, образующейся в результате *фазовой модуляции* той же плоской волны. Дифракция на фазовой решетке лежит в основе действия ультразвукового дефлектора.

Среда (жидкость или кристалл), по которой распространяется акустическая волна, представляет собой периодическую последовательность областей сжатия и разрежения. В результате *упругооптического эффекта* или, как его еще называют, явления фотоупругости, пространственные изменения плотности среды в стоячей волне приводят к пространственному изменению ее показателя преломления:

$$n = n_0 + \Delta n\cos(Kx - \Omega t), \quad (4.69)$$

где n_0 — показатель преломления невозмущенной среды; Δn — амплитуда изменений показателя преломления; $K = 2\pi/\Lambda$; $\Omega = 2\pi f_{зв}$.

В случае как стоячей, так и проходящей звуковой волны получающаяся решетка имеет период, равный длине ультразвуковой волны $\Lambda = v_{зв}/f_{зв}$. Для большинства прозрачных твердых тел (стекло, кварц) скорость ультразвуковых волн составляет 5000...6000 м/с, а для жидкостей — 1000...1500 м/с. Меньшим поглощением обладают ксилол и вода. Для них $\Lambda = 10^{-3}$ см = 10 мкм при $f_{зв} = 10^8$ Гц.

Скорость распространения света примерно на шесть порядков больше скорости акустической волны. Поэтому за время прохождения света поперек акустической волны ее можно считать неподвижной, и дифракционную картину можно рассчитывать при произвольной мгновенной интенсивности акустического поля.

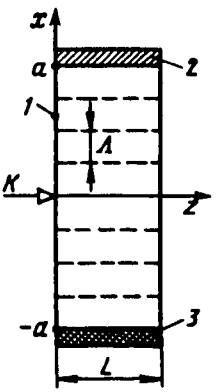


Рис. 4.4. Акустооптическая ячейка

Основным элементом дефлектора является акустооптическая ячейка *1* из фотоупругого материала (рис. 4.4). Ультразвуковая волна возбуждается вдоль оси *x* специальным излучателем *3* (кварц или турмалин). На противоположном конце ячейки закрепляют согласующуюся среду *2*, которая предназначена для гашения отраженных акустических волн. Первичная световая волна распространяется вдоль оси *z*, т. е. перпендикулярно направлению ультразвуковых волн. Если ширина акустического столба *L* в направлении распространения света достаточно мала ($L < \Lambda$), то можно в первом приближении считать, что световые лучи проходят через акустический столб прямолинейно. В этом случае изменения показателя преломления (4.69) влияют только на фазу и не меняют амплитуду света, прошедшего через столб:

$$\dot{E}(x,0) = E_0 \, e^{j(2\pi L/\lambda)[n_0 + \Delta n \cos(\pi x - \Omega t)]} \operatorname{rect}(x/a),$$

где $\operatorname{rect}\left(\frac{x}{a}\right) = \begin{cases} 1 & \text{при } |x| \leqslant a; \\ 0 & \text{при } |x| > a. \end{cases}$

Спектральная функция

$$\dot{\mathscr{E}}(u) = \int_{-\infty}^{\infty} \dot{E}(x,0) e^{-jux} dx = 2aE_0 \, e^{j2\pi n_0 L/\lambda} \sum_{m=-\infty}^{\infty} e^{jm(\Omega t - \pi/2)} \times$$

$$\times J_m\left(2\pi\Delta n \frac{L}{\lambda}\right) \frac{\sin[(2\pi a/\lambda)(\sin\theta - m\lambda/\Lambda)]}{(2\pi a/\lambda)(\sin\theta - m\lambda/\Lambda)} \qquad (4.70)$$

определяет направление, соотношение интенсивностей и частоты дифракционных максимумов.

Как следует из (4.70), для *m*-го максимума $\theta_m \doteq \arcsin(m\lambda/\Lambda)$. Угол отклонения первого бокового максимума $\theta_1 = \arcsin(\lambda/\Lambda) \approx \lambda/\Lambda = \lambda f_{зв}/v_{зв}$.

Следовательно, изменяя частоту ультразвукового генератора, можно изменять направление светового луча: $\Delta\theta_1 = \lambda \Delta f_{зв} / v_{зв}$.

Напряженность поля в m-м дифракционном максимуме пропорциональна $J_m(\Delta\varphi)$, где $\Delta\varphi = 2\pi\Delta nL/\lambda = k\Delta nL$ — амплитудное значение разности фаз, вносимой при прохождении оптического луча через фазовую решетку.

Экспоненциальный множитель в (4.70) показывает, что частота m-го максимума равна $\omega + m\Omega$, т. е. несколько сдвинута относительно частоты света в результате эффекта Доплера (напомним, что $f_{зв} \leqslant 10^8$ Гц, а $f \leqslant 10^{12}$ Гц). Постоянство интенсивностей дифракционных максимумов и сдвиг частоты связаны с тем, что бегущая ультразвуковая волна создает движущуюся решетку с постоянным соотношением максимумов и минимумов.

Для стоячей ультразвуковой волны спектральная функция отличается от (4.70) только тем, что в нее вместо $m(\Omega t - \pi/2)$ входит $2\pi n L/\lambda - \pi/\lambda$ и вместо $\Delta\varphi$ — $\Delta\varphi \cos\Omega t$. Следовательно, здесь происходит амплитудная модуляция дифрагмированного света во времени, а частота остается постоянной. Это объясняется тем, что образуемая стоячей волной решетка неподвижна в пространстве, но глубина фазовой модуляции меняется во времени (с частотой акустической волны).

В рассмотренном ультразвуковом дефлекторе энергия света, приходящаяся даже на первый дифракционный максимум, относительно невелика из-за распределения энергии по многим максимумам. Расчеты показывают, что энергию дифрагированного света удается сосредоточить практически только в одном максимуме первого порядка при увеличении толщины ультразвуковой ячейки $L \gg \Lambda^2/(\pi\lambda)$ и наклоне первичного светового луча к оси z под углом

$$\varphi = \arcsin[\lambda/(2\Lambda)] \approx \lambda/(2\Lambda). \qquad (4.71)$$

Явления, происходящие при этих условиях, называют *дифракцией Брэгга.*

Дифракцию на объемной фазовой решетке можно упрощенно представить как совокупность элементарных отражений световой волны от границ раздела между слоями (толщиной $\Lambda/2$) среды с разными показателями преломления ($n + \Delta n$ и $n - \Delta n$). При подобном описании амплитуда продифрагировавшей волны представляет собой сумму амплитуд таких элементарных волн. Поэтому максимальное значение дифракции будет достигнуто лишь при синфазном суммировании этих волн. Условие Брэгга (4.71) и представляет аналитическое выражение условия фазового синхронизма, учитывающего равенство углов падения и отражения. Ус-

ловие (4.71) является скалярной формой векторного соотношения $\mathbf{k}_2 = \mathbf{k}_1 + \mathbf{k}$, отражающего закон дискретного изменения импульса фотона падающей волны с волновым вектором $\mathbf{k}_1$. *Вектор решетки* $\mathbf{k}$ в таком процессе формирования вторичной волны (с волновым вектором $\mathbf{k}_2$) характеризует минимальный квант количества движения, на который может измениться импульс среды с записанной в ней дифракционной решеткой.

§ 4.10. Дифракция в фокусе линзы

Допустим, что рассматриваемый предмет можно представить в виде синфазной излучающей поверхности П в плоскости xy с известным распределением $f(x, y)$ скалярной компоненты вектора $\mathbf{E}$ поля излучения. Тогда элемент поверхности $\mathrm{d}x\,\mathrm{d}y$ в соответствии с принципом Гюйгенса испускает элементарную волну напряженностью $f(x, y)\,\mathrm{d}x\,\mathrm{d}y$ во всех направлениях.

Для волнового вектора $\mathbf{k}$, определяемого направляющими косинусами α, β, γ (рис. 4.5), фаза волны, выходящей из точки M, соответствующей радиус-вектору $\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y$ в плоскости $z = 0$, опережает на $k\zeta = \mathbf{kr} = k(\alpha x + \beta y)$ фазу элементарной волны, испускаемой с тем же волновым вектором из центра предмета, находящегося в начале координат. Результирующая волна в направлении, определяемом направляющими косинусами α, β, γ на основе принципа суперпозиции равняется сумме всех элементарных волн $\dot{f}(x, y)\mathrm{e}^{\mathrm{j}k(\alpha x+\beta y)}\,\mathrm{d}x\,\mathrm{d}y$.

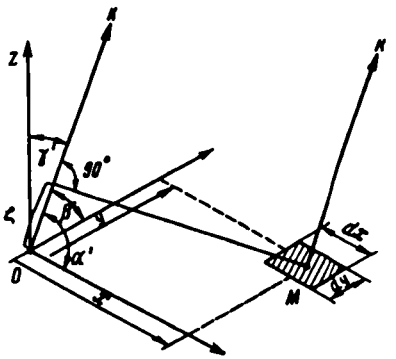


Рис. 4.5. Определение разности путей между параллельными элементарными лучами, испускаемыми синфазной поверхностью предмета

Изображение предмета обычно получают с помощью линзы. Элементарные волны, посылаемые или пропускаемые удаленным предметом в направлении α, β, γ собираются совершенной фокусирующей системой в одну точку P *фокальной плоскости* (рис. 4.6, *а*). В этой точке, если не учитывать конечных размеров линзы, напряженность электрического поля

$$\dot{E}_P = \iint_{\Pi} \dot{f}(x, y)\mathrm{e}^{\mathrm{j}k(\alpha x+\beta y)}\,\mathrm{d}x\,\mathrm{d}y. \qquad (4.72)$$

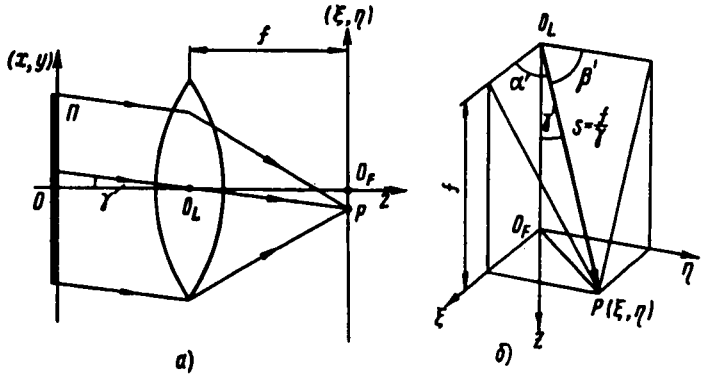


Рис. 4.6. Траектории лучей в фокусирующей системе

Таким образом, при известном распределении $\dot{f}(x, y)$ амплитуды электромагнитного поля предмета $П$, можно рассчитать дифракционную картину $\dot{E}(\xi, \eta)$, образующую изображение. При этом каждая точка в фокальной плоскости формируется за счет информации, полученной от всех точек предмета. Однозначное соотношение между распределениями поля в предмете и света в дифракционной картине имеет вид преобразования Фурье

$$\dot{E}(\xi, \eta) = \int_{-\infty}^{+\infty}\int \dot{f}(x, y)\, e^{j(ux+vy)}\, dx\, dy = \mathcal{F}[\dot{f}(x, y)] = F(u, v).$$

Распределение поля в фокальной плоскости обычно задается координатами ξ и η (см. рис. 4.6, *а*). Начало этой системы координат располагается на оси линзы, совпадающей с осью z координатной системы предмета, а оси ξ и η параллельны соответственно осям x и y. Тогда координаты точки $P(\xi, \eta)$, в которую фокусируются все элементарные лучи, испускаемые предметом в направлении $\alpha, \beta, \gamma \approx 1$, определяются из соотношений (рис. 4.6, *б*) $\xi = f\alpha/\gamma \approx f\alpha$, $\eta = f\beta/\gamma \approx f\beta$ или $\alpha = \cos\alpha' \approx \xi/f$, $\beta = \cos\beta' \approx \eta/f$.

Таким образом, дифракционную картину (4.72) в фокальной плоскости описывает уравнение

$$\dot{E}(\xi, \eta) = \iint_{П} \dot{f}(x, y)\, e^{j2\pi(\xi x+\eta y)/(\lambda f)}\, dx\, dy = F(u, v), \qquad (4.73)$$

которое характеризует распределение напряженности поля, формирующего изображение предмета в фокальной плоскости. Каждой точке с ко-

ординатами ξ и η соответствуют в преобразовании Фурье пространственные частоты $u = \xi/(f\lambda) = \alpha/\lambda$ и $v = \eta/(f\lambda) = \beta/\lambda$.

Из уравнения (4.73) следует, в частности, что волновая природа света не позволяет сфокусировать световой луч в идеальную точку даже в случае идеальной линзы и строго параллельного когерентного светового потока. Например, при освещении цилиндрической линзы, имеющей бесконечную протяженность в направлении оси $y(\eta)$, световым пучком в виде однородной $[f(x, y) = 1]$ плоской электромагнитной волны шириной D, направленной вдоль оптической оси z линзы ($\gamma = 1$), уравнение для напряженности электрического поля (4.73) становится одномерным :

$$E(\xi) = \int_{-D/2}^{D/2} e^{j2\pi\xi x/(\lambda f)}\,dx = D\frac{\sin\dfrac{\pi D\xi}{\lambda f}}{\dfrac{\pi D\xi}{\lambda f}}. \qquad (4.74)$$

Все приемники (фотоэлектрические, фотографические и т.д.) регистрируют только интенсивность

$$I(\xi) \doteq \dot{E}\dot{E}^{*} = D^2\frac{\sin^2\dfrac{\pi D\xi}{\lambda f}}{\left(\dfrac{\pi D\xi}{\lambda f}\right)^2},$$

которую в оптике принято называть *дифракционной картиной*, соответствующей апертуре D.

Нетрудно видеть, что дифракционная картина цилиндрической линзы (4.74) имеет тот же характер, что и дифракционная картина бесконечной щели (4.63) в дальней зоне с той же апертурой. Полуширина центральной, наиболее яркой, полосы определяется координатой первого минимума дифракционной картины $\xi_0 = \lambda f/D$. Для круглой апертуры того же размера аналогичный расчет радиуса центрального пятна дифракционной картины приводит к значению

$$r_0 = 1{,}22\lambda f/D, \qquad (4.75)$$

называемому иногда *радиусом диска Эйри*.

Если в начале координат ($x = y = 0$) предметной плоскости ($z = 0$) находится единичный точечный источник, описываемый δ-функцией, то согласно (4.73)

$$\dot{E}(\xi, \eta) = \int_{-\infty}^{+\infty}\int \delta(x, y)e^{-j(ux+vy)}\,dx\,dy = [e^{-j(ux+vy)}]_{\substack{x=0\\y=0}} = 1, \qquad (4.76)$$

т. е. линза в плоскости $\xi\eta$ создает равномерное поле однородной плоской волны с постоянной фазой. Если точечный источник смещен в предметной плоскости относительно оптической оси линзы и находится в точке ($x = a,\ y = b$), то

$$\dot{E}(\xi, \eta) = \int_{-\infty}^{+\infty}\int \delta(x-a, y-b)\, e^{-j(ux+vy)}\, dx\, dy = e^{-j(ua+vb)} = e^{-jk(\xi a+b\eta)/f}, \quad (4.77)$$

т. е. плоскость $\xi\eta$ пересекает также плоская волна с одинаковой амплитудой $|\dot{E}| = 1$, наклоненная к оси ξ под углом α ($\operatorname{tg}\alpha = a/f$) и к оси η — под углом β ($\operatorname{tg}\beta = b/f$).

Дифракционную картину единичного точечного источника можно рассматривать как импульсный отклик оптической системы.

Оптические системы, если их использовать в качестве элементов преобразования сигналов в виде электромагнитных волн, обладают двумерной гибкостью, отсутствующей у электрических систем. Это не исключает введения и использования времени в качестве дополнительной третьей координаты.

Уравнение (4.73), описывая дифракционную картину в фокальной плоскости линзы, лишь косвенно характеризует распределение электромагнитного поля в плоскости изображения. Связь объекта (плоскость xy) и его изображения имеет более сложный характер. Она выражается через распределение по *плоскости изображения* $x'y'$ комплексных амплитуд поля $\dot{d}(x', y')$, создаваемого единичным точечным источником, находящимся в начале координат ($x = 0, y = 0$).

В идеальной фокусирующей системе распределение поля определяется только дифракцией, а в реальной системе оно учитывает все виды искажений и, в частности, аберрацию и рассеяние в плоскости приемника. Если предположить в первом приближении, что искажения не меняются при перемещении источника в плоскости объекта, то изображение единичного источника в точке с координатами x, y может быть представлено функцией $\dot{d}(x' - gx, y' - gy) = \dot{d}(x' - x_0, y' - y_0)$, где g — линейное (поперечное) увеличение прибора; $x_0 = gx$; $y_0 = gy$. Если распределение комплексных амплитуд на объекте описывает некоторая функция $\dot{\Omega}(x, y)$, то функция $\dot{\Omega}(x_0, y_0)$ характеризует распределение амплитуд поля на том изображении, которое должно быть получено по законам геометрической оптики. Действительная амплитуда поля на плоскости изображения в точке с координатами x', y' может быть найдена суммированием тех составляющих, которые обусловле-

ны распределением амплитуд поля излучения от всех точек объекта с учетом их удельного веса:

$$E(x', y') = \iint \dot{d}(x' - x_0, y' - y_0)\dot{\Omega}(x_0, y_0)\,dx_0\,dy_0 = \dot{d} \circledast \dot{\Omega}.$$

Если элементы объекта излучают абсолютно некогерентные колебания, то вместо суммирования комплексных амплитуд происходит суммирование освещенностей. Тогда распределение освещенности изображения

$$i(x', y') = \iint s(x' - x_0, y' - y_0)o(x_0, y_0)\,dx_0\,dy_0 = s \circledast o.$$

где $o = \dot{\Omega}\dot{\Omega}$ — распределение яркости на объекте; $s = \dot{d}\dot{d}$ — функция разброса, дающая распределение освещенности от точечного объекта.

Согласно теореме свертки $I(u', v') = O(u', v')S(u', v')$ — преобразование Фурье изображения равно произведению преобразований Фурье объекта и изображения изолированной точки.

Таким образом, формирование изображения в оптическом устройстве можно также рассматривать как процесс прохождения оптического сигнала через фильтр пространственных частот. Каждой составляющей $O(u', v')$ соответствует свой коэффициент передачи $S(u', v')$. Его можно назвать и функцией фильтрования, поскольку она определяет передачу прибором каждой частотной составляющей.

§ 4.11. Оптические системы обработки информации

Оптические методы обработки информации можно разделить на три группы. К первой группе относят методы, основанные на теории пространственной фильтрации оптического изображения. В оптических методах второй группы для создания устройств ввода, логического преобразования, запоминания и вывода информации используют электрооптические явления. Эту группу методов и устройств применяют в оптоэлектронике. Третья группа охватывает методы полностью оптической обработки, при которой световой поток является не только носителем информации, но и управляющим сигналом.

В настоящей главе изложены методы первой группы. Способность оптических систем выполнять двумерное преобразование Фурье позволила применять к этим системам общие принципы теории колебаний, методы теории электрических цепей и связи, при помощи которых были решены основные вопросы, связанные со структурой оптического изображения, сформированного с помощью когерентного освещения. Такой подход значительно упростил методы анализа и синтеза оптических сис-

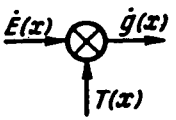


Рис. 4.7. Эквивалентная схема плоского транспаранта

тем, поскольку дал возможность рассматривать их как линейные многополюсники с определенными коэффициентами передачи.

Основными элементами когерентной оптической системы являются модулирующая среда, собирающая линза, свободное пространство и приемник света.

В качестве модулирующей среды обычно используют плоский транспарант, т. е. тонкий диапозитив с пропусканием

$$\underline{T}(x, y) = T(x, y)\mathrm{e}^{j\Phi(x, y)}.$$

Если транспарант освещает волна $\dot{E}(x, y) = E(x, y)\mathrm{e}^{j\theta(x, y)}$, то на его выходе возникает волна

$$\dot{g}(x, y, 0) = T(x, y)E(x, y)\mathrm{e}^{j[\Phi(x, y)+\theta(x, y)]} = g(x, y, 0)\mathrm{e}^{j\varphi(x, y, 0)}. \quad (4.78)$$

На рис.4.7 показана эквивалентная схема транспаранта, на которой знак × означает операцию умножения.

Тонкую линзу можно рассматривать как плоский фазовый транспарант, не обладающий ослаблением, но создающий некоторую задержку $t(x, y)$ и соответствующее запаздывание фазы. Предположим, что линза образована двумя пересекающимися цилиндрическими поверхностями (рис. 4.8, *а*).

На расстоянии x от оси линзы путь луча через транспарант складывается из участков AB, BC и CD. При подсчете этих расстояний необходимо исходить из того, что элементы оптических систем, используемых в качестве аналоговых вычислительных машин, находятся на сравнительно близких расстояниях, соответствующих дифракции Френеля ($x^2 \ll z^2$). Поэтому

$$CD = r_1 - z = \sqrt{z^2 + x^2} - z \approx x^2/(2z) \approx x^2/(2r_1),$$

$$AB = r_2 - z \approx x^2/(2r_2),$$

$$BC = e - AB - CD \approx e - \frac{x^2}{2}\left(\frac{1}{r_1} + \frac{1}{r_2}\right).$$

Из последнего выражения следуют соотношения между полушириной линзы a и ее толщиной e:

$$e = \frac{a^2}{2}\left(\frac{1}{r_1} + \frac{1}{r_2}\right)$$

и между расстоянием от оси x и запаздыванием t:

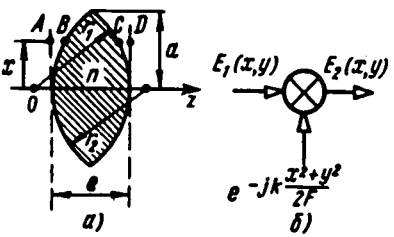


Рис. 4.8. Линза (*а*) и ее эквивалентная схема (*б*)

$$t = \frac{BC}{v} + \frac{AB+CD}{c} = \frac{e}{v} - \frac{x^2}{2v}\left(\frac{1}{r_1}+\frac{1}{r_2}\right) + \frac{x^2}{2c}\left(\frac{1}{r_1}+\frac{1}{r_2}\right) = \tau_0 - \frac{x^2}{2cF},$$

где фокусное расстояние F определяется выражением

$$\frac{1}{F} = (n-1)\left(\frac{1}{r_1}+\frac{1}{r_2}\right) = \mu, \qquad (4.79)$$

а постоянная составляющая $\tau_0 = e/v$. В общем случае она зависит от начала отсчета времени. При соответствующем выборе начала отсчета значение τ_0 можно опустить. Аналогичным образом для сферических линз можно получить

$$t(x, y) = \tau_0 - \frac{x^2 + y^2}{2cF}.$$

Следовательно, в случае гармонической волны сферическая линза может быть представлена функцией пропускания

$$T_л(x, y) = e^{jkt(x,y)} = e^{-jk(x^2+y^2)/(2F)} \qquad (4.80)$$

и эквивалентной схемой рис. 4.8, *б*. Цилиндрическая линза имеет функцию пропускания $e^{-jkx^2/(2F)}$ или $e^{-jky^2/(2F)}$ в зависимости от ее ориентации.

Свободное пространство, заключенное между двумя плоскостями $z = 0$ и $z = D$ может быть описано импульсной характеристикой (4.55)

$$h(x, y) = \frac{C_1}{D} e^{jk(x^2+y^2)/(2D)} \qquad (4.81)$$

или частотной характеристикой (4.50)

$$\zeta(u, v) = C_2 e^{-j(u^2+v^2)/(4\alpha)} = C_2 e^{-j(u^2+v^2)D/(2k)}, \qquad (4.82)$$

где $\alpha = k/(2D) = \pi/(\lambda D)$, C_1, C_2 — постоянные числа, не зависящие ни от $r = \sqrt{x^2 + y^2}$, ни от $\chi = \sqrt{u^2 + v^2}$ и обычно опускаемые при рассмотрении распределения сигналов (изображений) в поперечных плоскостях (x, y).

На эквивалентной схеме свободное пространство изображают прямоугольником (рис. 4.9, *а*), в котором записано френелевское ядро $\exp\dfrac{jk(x^2 + y^2)}{2D}$ и который означает операцию свертки.

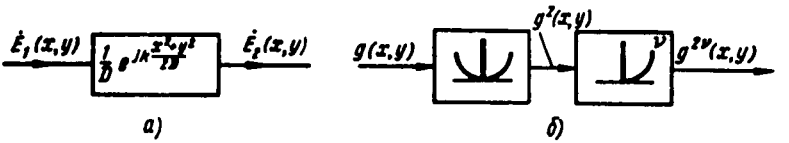


Рис. 4.9. Эквивалентные схемы свободного пространства (*а*) и пространственного приемника (*б*)

Пространственный приемник регистрирует выходной сигнал оптической системы $E_2(x, y)$ при помощи фотопленки, электрооптического преобразователя, фотоумножителя, фотодиодной мозаики (матрицы) или фотоэлемента. Подобные устройства реагируют на интенсивность светового потока. В пределах диапазона линейной передачи интенсивности светового потока приемник можно рассматривать как квадратичный детектор с выходным сигналом $E_2^2(x, y)$. Однако этот (динамический) диапазон сравнительно невелик (около 25 дБ). Вне этого диапазона кривая передачи интенсивности обычно нелинейна. Поэтому в эквивалентную схему приемника (рис. 4.9, *б*) вводят последовательно с квадратичным детектором нелинейный четырехполюсник, который дает на выходе сигнал, пропорциональный входному сигналу, возведенному в некоторую степень ν, равную коэффициенту контрастности, т. е. $\sim|E_2(x, y)|^{2\nu}$.

Оптическую систему в целом можно рассматривать как фильтр пространственных частот. При этом аналогично временны́м линейным фильтрам в электрических цепях действие оптического фильтра можно представить в виде линейного оператора $\mathscr{L}$, воздействующего на входное пространственное распределение светового поля $E_{\text{вх}}(x, y)$ так, что на выходе системы $E_{\text{вых}}(x, y) = \mathscr{L}[E_{\text{вх}}(x, y)]$.

Для оптического фильтра, как и для временно́го, линейным оператором можно пользоваться в том случае, когда оптическая система подчиняется *принципу наложения* (суперпозиции). Однако суперпозиция имеет различный смысл для систем с когерентным и некогерентным освещением. В первом случае принцип линейности относится к амплитудам светового потока, а во втором — к его интенсивностям. Очевидно, что если пассивная оптическая система линейна, то она обладает и свойством *устойчивости*, т. е. всякий сигнал с конечной энергией создает на выходе сигнал также с конечной энергией.

Условие инвариантности, или *транспозиции*, могло бы выполняться в оптической системе в том случае, когда распределение освещенности в изображении точки не меняется, а система перемещается в плоскости объекта. В реальных системах аберрационные коэффициенты зависят от угловых координат. Однако условием инвариантности в оптике можно пользоваться в пределах области, где выполняются условия стигматического изображения, когда поперечное увеличение не зависит от размеров объектов (теорема Лагранжа—Гельмгольца) и передаточная функция (импульсная или частотная характеристика) постоянна. В пределах каждой из таких *изопланарных зон*, на которые можно разбить плоскости изображения и объекта, $\mathscr{L}[E_{\text{вх}}(x \pm x_0, y \pm y_0)] = E_{\text{вых}}(x \pm x_0, y \pm y_0)$.

Условие физической осуществимости в электрических системах реализуется вследствие того, что независимой переменной является время, и поэтому сигнала на выходе не будет до тех пор, пока не появится сигнал на входе. Поскольку в оптических системах функцию независимой переменной выполняют координаты точки на плоскости, то условие физической осуществимости не играет той роли, какую оно выполняет во временно́й области. Это позволяет синтезировать такие характеристики фильтров, которые нельзя или трудно осуществить в электрических системах.

В заключение отметим, что коэффициент пропускания сферической линзы (4.80) и отклик свободного пространства (4.81) имеют близкую структуру и их можно объединить, взяв за основу в качестве оператора единую *обобщенную функцию Френеля*

$$\Phi(\mathbf{r}, f) = \mathrm{e}^{jkfr^2/2}. \tag{4.83}$$

Тогда импульсный отклик свободного пространства

$$h(\mathbf{r}) = d\Phi(\mathbf{r}, d), \tag{4.84}$$

а коэффициент пропускания линзы

$$T_л(\mathbf{r}) = \overset{*}{\Phi}(\mathbf{r}, f), \tag{4.85}$$

причем *диоптрийные параметры* $f = F^{-1}$, $d = D^{-1}$. Обобщенная функция (4.83) обладает свойствами, аналогичными свойствам преобразований Фурье и Френеля. Преобразование Фурье обобщенной функции Френеля

$$F(u) = \int_{-\infty}^{\infty} \mathrm{e}^{jkfx^2/2}\mathrm{e}^{-jux}\,\mathrm{d}x = \frac{\sqrt{j\pi}}{\sqrt{kf/2}}\mathrm{e}^{-ju^2/(2kf)} = \tag{4.86}$$

$$= \sqrt{j2\pi kf_0}\,\mathrm{e}^{-jkf_0u^2/2} = \sqrt{j2\pi kf_0}\,\overset{*}{\Phi}(u, f_0),$$

где $f_0 = 1/(k^2 f)$.

Для свободного пространства бесконечно малой глубины отклик (4.58) равен δ-функции

$$\lim_{d\to\infty} \mathrm{d}\Phi(\mathbf{r}, d) = \delta(\mathbf{r}), \tag{4.87}$$

т. е. здесь не происходит никакого дифракционного размытия оптического сигнала.

Перейдем к рассмотрению некоторых основных методов и типичных систем оптической обработки информации.

§ 4.12. Многоканальный анализ спектра

Простейшей является оптическая система, содержащая в себе модулирующую среду, первое свободное пространство, собирающую сферическую линзу и второе свободное пространство (рис. 4.10, *а*). Следуя эквивалентной схеме этого оптического каскада (рис. 4.10, *б*), можно сразу записать общее выражение для определения на его выходе, т. е. в плоскости P_3, комплексной амплитуды напряженности поля:

$$\dot{E}(\mathbf{r}) = \underbrace{\int_{P_2} d_2\Phi(\mathbf{x}_1 - \mathbf{r}, d_2)}_{\text{свертка во втором свободном пространстве}} \underbrace{\dot{\Phi}(x_1, f)}_{\text{линза}}\, \mathrm{d}x_1 \underbrace{\int_{P_1} E_0 f(\mathbf{x})}_{\text{модуляция}} \underbrace{d_1\Phi(\mathbf{x} - \mathbf{x}_1, d_1)}_{\text{свертка в первом свободном пространстве}} \mathrm{d}\mathbf{x}$$

где $\mathbf{r} = \xi\mathbf{e}_x + \eta\mathbf{e}_y$; $\mathbf{x} = x\mathbf{e}_x + y\mathbf{e}_y$ и $\mathbf{x}_1 = x_1\mathbf{e}_x + y_1\mathbf{e}_y$.

С учетом соотношения $\Phi(\mathbf{x} - \mathbf{x}_1, f) = \Phi(\mathbf{x}, f)\Phi(\mathbf{x}_1, f)\mathrm{e}^{-jkf\mathbf{x}\mathbf{x}_1}$

$$\dot{E}(\mathbf{r}) = d_1 d_2 E_0 \int_{P_2}\int_{P_1} \underline{f}(\mathbf{x})\Phi(\mathbf{x}, d_1)\Phi(\mathbf{x}_1, d_1)\mathrm{e}^{-jkd_1\mathbf{x}\mathbf{x}_1}\dot{\Phi}(\mathbf{x}_1, f)\times$$

$$\times\Phi(\mathbf{x}_1, d_2)\Phi(\mathbf{r}, d_2)\mathrm{e}^{-jkd_2\mathbf{x}_1\mathbf{r}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x}_1 = d_1 d_2 E_0\Phi(\mathbf{r}, d_2)\int_{P_2}\int_{P_1}\underline{f}(\mathbf{x})\Phi(\mathbf{x}, d_1)\times$$

$$\times\Phi(\mathbf{x}_1, d_1 - f + d_2)\mathrm{e}^{-jk(d_1\mathbf{x}_1\mathbf{x} + d_2\mathbf{x}_1\mathbf{r})}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x}_1.$$

С учетом свойства Фурье-образа (4.86)

$$\int\Phi(x, f_1)\mathrm{e}^{-jkf_2ux}\,\mathrm{d}x = \frac{c}{f_1}\dot{\Phi}(u, f_2^2/f_1)$$

уравнение оптического каскада в общем случае приобретает окончательный вид

$$G(r) = \frac{\dot{E}(r)}{E_0} = \frac{d_1 d_2}{d_1 - f + d_2}\Phi(\mathbf{r}, d_2)\times$$

$$\times\int_{P_1}\underline{f}(\mathbf{x})\Phi(\mathbf{x}, d_1)\dot{\Phi}\left(\mathbf{x} + \frac{d_2}{d_1}\mathbf{r}, \frac{d_1^2}{d_1 - f + d_2}\right)\mathrm{d}\mathbf{x}. \qquad (4.88)$$

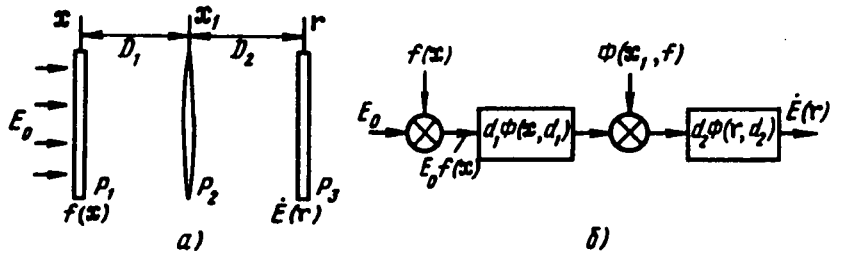


Рис. 4.10. Простейший оптический каскад (*а*) и его эквивалентная схема (*б*)

Рассмотрим два характерных примера. В первом из них оптический каскад удовлетворяет условию $d_1 + d_2 - f = 0$, т. е. $d_1 + d_2 = f$ или

$$\frac{1}{D_1} + \frac{1}{D_2} = \frac{1}{F}. \tag{4.89}$$

При этом, как следует из (4.87)

$$\frac{d_1^2}{d_1 + d_2 - f} \dot{\Phi}\left(\mathbf{x} + \frac{d_2}{d_1}\mathbf{r}, \frac{d_1^2}{d_1 + d_2 - f}\right) = \delta\left(\mathbf{x} + \frac{d_2}{d_1}\mathbf{r}\right)$$

и на выходе каскада оптический сигнал

$$G(\mathbf{r}) = \frac{d_2}{d_1}\Phi(\mathbf{r}, d_2)\int \Phi(\mathbf{x}, d_1)\underline{f}(\mathbf{x})\delta\left(\mathbf{x} + \frac{d_2}{d_1}\mathbf{r}\right)\mathrm{d}\mathbf{x} =$$

$$= \frac{d_2}{d_1}\Phi(\mathbf{r}, d_2)\underline{f}\left(-\frac{d_2}{d_1}\mathbf{r}\right)\Phi\left(-\frac{d_2}{d_1}\mathbf{r}, d_1\right) = \underline{f}\left(-\frac{d_3}{d_2}\mathbf{r}\right)\Phi\left(\mathbf{r}, d_2 + \frac{d_2^2}{d_1}\right)C\left(\frac{d_2}{d_1}\right), \tag{4.90}$$

где C — комплексное число, зависящее от отношения d_2/d_1.

Из выражения (4.90) следует, что оптический сигнал $G(\mathbf{r})$ равен исходному сигналу $f(\mathbf{x})$, но перевернутому и изменившему масштаб.

Таким образом, когда размеры оптического каскада удовлетворяют уравнению линзы (4.89), он выполняет функции *изображающей системы*, и условие (4.89) является условием образования изображения.

Обобщенная функция Френеля в (4.90) описывает эквивалентный точечный освещающий источник, расположенный на расстоянии $D = (d_2 + d_2^2/d_1)^{-1}$ слева от транспаранта.

Во втором примере выясним условия, при которых оптический каскад может осуществлять *преобразование Фурье*. Для этого следует преобразовать подынтегральное выражение (4.88) следующим образом:

$$\Phi(\mathbf{x}, d_1)\dot{\Phi}\left(\mathbf{x} + \frac{d_2}{d_1}\mathbf{r}, \frac{d_1^2}{d_1 - f + d_2}\right) = \Phi(\mathbf{x}, d_1)\dot{\Phi}\left(\mathbf{x}, \frac{d_1^2}{d_1 - f + d_2}\right) \times$$

$$\times \dot{\Phi}\left(\frac{d_2}{d_1}\mathbf{r}, \frac{d_1^2}{d_1 - f + d_2}\right)\exp\left[-jk\mathbf{x}\mathbf{r}\frac{d_2}{d_1}\frac{d_1^2}{d_1 - f + d_2}\right] =$$

$$= \Phi\left(\mathbf{x}, d_1 - \frac{d_1^2}{d_1 - f + d_2}\right)\dot{\Phi}\left(\mathbf{r}, \frac{d_2^2}{d_1^2}\frac{d_1^2}{d_1 - f + d_2}\right)\exp\left[-jk\mathbf{x}\mathbf{r}\frac{d_1 d_2}{d_1 - f + d_2}\right].$$

После этого из уравнения оптического каскада

$$G(\mathbf{r}) \doteq \dot{\Phi}\left(\mathbf{r}, d_2 - \frac{d_2^2}{d_1 - f + d_2}\right) \int_{P_1} \underline{f}(\mathbf{x}) \Phi\left(\mathbf{x}, d_1 - \frac{d_1^2}{d_1 - f + d_2}\right) \times$$

$$\times \exp\left[-jk\mathbf{x}\mathbf{r}\frac{d_1 d_2}{d_1 - f + d_2}\right] d\mathbf{x} = \int_{P_1} \underline{f}(\mathbf{x}) e^{-jk\mathbf{x}\mathbf{r}/f}\, d\mathbf{x} = F(\chi), \qquad (4.91)$$

если $d_1 - \dfrac{d_1^2}{d_1 - f + d_2} = 0$; $d_2 - \dfrac{d_2^2}{d_1 - f + d_2} = 0$, т. е. $d_1 = d_2 = f$ или $D_1 = D_2 = F$.

Таким образом, если поместить исходный сигнал $f(x)$ в переднюю фокальную плоскость линзы, то выходной сигнал $G(\mathbf{r})$ в задней фокальной плоскости будет фурье-образом входного сигнала.

Способность линзы осуществлять преобразование Фурье приходящего электромагнитного сигнала используют в оптических анализаторах спектра. При этом в качестве модулирующей среды наряду с диапозитивами и транспарантами применяют акустооптическую ячейку, в которой пьезоэлектрический преобразователь возбуждается исследуемым сигналом $u(t)$, в результате чего возникают ультразвуковые волны той же частоты и формы, что и у входного сигнала. Под воздействием ультразвуковой волны происходит модуляция коэффициента преломления в направлении (x) распространения этой волны, причем изменение показателя преломления Δn пропорционально напряжению сигнала на входе. Следовательно, световой поток, выходящий из модулятора, будет образован пространственно модулированной по фазе $\varphi(x)$ волной, отображающей форму входного сигнала $u(t)$. Поскольку на практике почти всегда амплитуда модуляции фазы $\Delta\varphi_m \leqslant 0{,}3$, то в первом дифракционном максимуме выходного сигнала (4.70) функция Бесселя $J_1(\Delta\varphi_m) \approx \Delta\varphi_m/2$, т. е. выходной сигнал можно рассматривать как спектр входного сигнала (даже в системах с большим динамическим диапазоном).

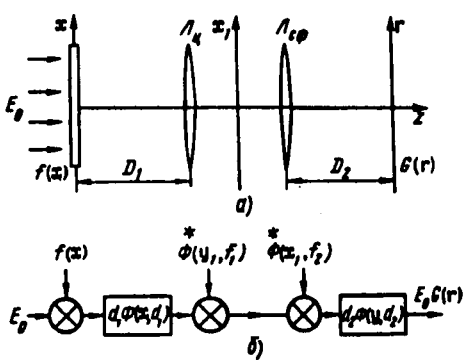


Рис. 4.11. Многоканальный анализатор спектра ($а$) и его эквивалентная схема ($б$)

Двухмерная гибкость оптических систем дает возможность простым способом выполнять *одновременно* спектральный анализ большого количества входных каналов. В этом заключается одно из основных преимуществ оптического анализатора спектра.

На рис. 4.11 в качестве примера показана схема многоканального анализатора спектра, составленного из цилиндрической и сферической линз и выполняющего преобразование Фурье каждого из каналов в направлении оси x, а также позволяющего получать их изображения, сдвинутые вдоль оси y.

Сигнал на выходе такого каскада

$$G(\mathbf{r}) = \Phi(\mathbf{r}, d_2) \int_{P_1} \int_{P_2} \underline{f}(\mathbf{x}) \Phi(\mathbf{x}, d_1) \Phi(y_1, d_1 - f_1 + d_2 - f_2) \Phi(x_1, d_1 - f_2 + d_2) \times$$

$$\times e^{-jkd_1 x_1 x} e^{-jkd_2 x_1 r} d\mathbf{x}_1 d\mathbf{x} = \Phi(\mathbf{r}, d_2) \int \underline{f}(\mathbf{x}) \Phi(\mathbf{x}, d_1) \dot{\Phi}\left(x + \frac{d_2}{d_1}\xi, \frac{d_1^2}{d_1 - f_2 + d_2} \right) \times$$

$$\times \dot{\Phi}\left(y + \frac{d_2}{d_1}\eta, \frac{d_1^2}{d_1 - f_2 + d_2 - f_1} \right) d\mathbf{x} = \Phi(r, d_2) \Phi\left(\frac{d_2}{d_1}\eta, d_1 \right) \int \underline{f}\left(x, -\frac{d_2}{d_1}\eta \right) \times$$

$$\times \Phi(\mathbf{x}, d_1) \dot{\Phi}\left(x + \frac{d_2}{d_2}\xi, \frac{d_1^2}{d_1 - f_2 + d_2} \right) dx = \Phi(r, d_2) \Phi\left(\eta, d_2 + d_1 \frac{d_2^2}{d_1^2} \right) \times$$

$$\times \int \underline{f}\left(x, -\frac{d_2}{d_1}\eta \right) \Phi\left(x, d_1 - \frac{d_1^2}{d_1 + d_2 - f} \right) \exp\left(-jk \frac{d_1 d_2}{d_1 + d_2 - f_2} x\xi \right) dx =$$

$$= \Phi(\xi, d_2) \Phi\left(\eta, f_2 + \frac{f_2^2}{f_1} \right) \int \underline{f}\left(x, -\frac{d_2}{d_1}\eta \right) e^{-jkf_2 x\xi} dx.$$

Предполагается, что в системе, дающей изображение вдоль оси y,

$$\frac{d_1^2}{d_1 - f_2 + d_2 - f_1} \dot{\Phi}\left(y + \frac{d_2}{d_1}\eta, \frac{d_1^2}{d_1 - f_2 + d_2 - f_1} \right) = \delta\left(y + \frac{d_3}{d_2}\eta \right),$$

а в системе, осуществляющей преобразование Фурье в направлении оси x,

$$\Phi\left(x, d_1 - \frac{d_1^2}{d_1 + d_2 - f_2} \right) = 1.$$

Второе предположение выполняется при условии $d_2 = f_2$, а первое — при условии $d_1 + d_2 = f_1 + f_2$. Отсюда следует, что $d_1 = f_1$ и

$$\frac{1}{D_2} + \frac{1}{D_3} = \frac{F_1 + F_2}{F_1 F_2} = \frac{1}{F},$$

т. е. две линзы $Л_ц$ и $Л_{сф}$ действуют в этом случае как одна с фокусным расстоянием F.

В одном из подобных анализаторов были использованы ультразвуковые модуляторы в 24 независимых каналах. Прямоугольная ванна модулятора с дистиллированной водой выполнена из алюминия с тефлоновым покрытием и окнами, закрытыми стеклами с хорошими оптическими

свойствами. Каждый ультразвуковой преобразователь изготовлялся из пьезоэлектрического кристалла в виде кварцевой пластинки с резонансной частотой 20 МГц. Поглощение ультразвуковых колебаний происходит в тефлоне. Уменьшение интерференции между каналами осуществлялось специальной диафрагмой разделения каналов. В системе съема выходных данных был применен механический метод смещения измерительной диафрагмы (ширина — 1 мкм, длина — 1 мм), непосредственно за которой установлен фотоумножитель. К винту перемещения щели был прикреплен вал потенциометра. При этом снимаемое с потенциометра напряжение было пропорционально положению щели (ξ). В других системах отсчета результаты фиксировались на фотопленку с последующим фотометрированием.

§ 4.13. Корреляционная фильтрация

Оптическое устройство, состоящее из линзы и транспаранта с нанесенными на него исходными данными, выполняет функцию двухмерного фильтра. Коэффициент передачи зависит от размеров элементов оптической системы. Полосу пропускания такого фильтра существенно ограничивают поперечные размеры линзы (*апертура* $2a$) и аберрация (искажения).Если апертура линзы достаточно велика, а аберрации пренебрежимо малы, то для малых апертурных углов, стягиваемых рамкой транспаранта в изопланарной области, оптическая система ведет себя как линейный фильтр. В такой системе оптические сигналы подчиняются принципу суперпозиции.

Оптические линейные системы, выполняющие преобразование Фурье, открывают возможность осуществления анализа и различных фильтраций пространственного спектра. Например, непрозрачная маска в плоскости Фурье на вполне определенной пространственной частоте удаляет соответствующую гармонику, содержащуюся в исходном сигнале. Наоборот, узкая щель или отверстие в этом месте выделяет периодический сигнал данной частоты на фоне посторонних гармоник. Аналогичную, но несравнимо более сложную задачу решают с помощью пространственной фильтрации двухмерных изображений. Оптическая фильтрация в сочетании с голографией позволяет не только регистрировать, но и преобразовывать любой комплекснозначный стационарный оптический сигнал.

Любая задача фильтрации сводится к обнаружению полезного сигнала известной формы на фоне аддитивных шумов. К этой проблеме относится и распознавание образов, например поиск слова «Иванов» на одной из страниц книги. Из теории фильтров известно, что если результирующий сигнал на входе $\underline{f}(t) = \underline{m}(t) + \underline{n}(t)$ состоит из полезного сигна-

ла $\underline{m}(t)$ и случайного шума $\underline{n}(t)$, то отношение мощности сигнала к мощности шума на выходе фильтра будет максимально, если частотная характеристика фильтра

$$G(\omega) \doteq \frac{\dot{M}(\omega)}{S_n(\omega)}, \tag{4.92}$$

где $\dot{M}(\omega)$ — комплексно сопряженный фурье-образ полезного сигнала; $S_n(\omega)$ — энергетический спектр шума на входе фильтра.

Фильтр, обладающий характеристикой (4.92), можно назвать *абсолютно оптимальным*. Если, кроме того, шум $\underline{n}(t)$ описывает свойства белого шума и спектр $S_n(\omega) = S_0$ не зависит от частоты, то оптимальное значение частотной характеристики

$$G(\omega) \doteq \dot{M}(\omega). \tag{4.93}$$

Линейный фильтр с такой характеристикой называют *согласованным*.

Согласованные фильтры позволяют осуществить дальнейшее эффективное подавление шумов и помех. Из ряда методов подавления можно упомянуть использование систем с интегральными преобразованиями типа преобразований Фурье и Френеля. Такие системы дают возможность обратного преобразования с восстановлением исходного полезного сигнала или изображения, «очищенного» от помех. Для распознавания образов используется корреляционный метод подавления шумов.

На выходе коррелятор выдает не сам полезный сигнал, а функцию автокорреляции для сигнала и кросс-корреляции сигнала и помехи. Для этого на вход согласованного фильтра подают не сам сигнал $f(t)$, а его фурье-образ $F(\omega)$ и производят обратное преобразование Фурье сигнала на выходе фильтра $F(\omega)\dot{M}(\omega)$.

Корреляционная фильтрация лежит в основе оптических методов распознавания образов. Далее рассмотрена ее принципиальная схема на примере поиска слова «Иванов».

Голограмма согласованного фильтра в первых экспериментах такого рода изготовлялась по схеме, изображенной на рис. 4.12. В предметной плоскости xy располагался диапозитив со словом «Иванов». Функция пропускания диапозитива $\underline{m}(x, y)$ определяла распределение в плоскости xy комплексных амплитуд прошедшей волны при освещении диапозитива однородной плоской волной. Между фотопластинкой (плоскость ξ, η) и диапозитивом находилась линза $Л$ с фокусным расстоянием f. Распределение комплексных амплитуд волны, прошедшей диапозитив, в плоскости изображения определяется преобразованием Фурье-функции $\underline{m}(x, y)$

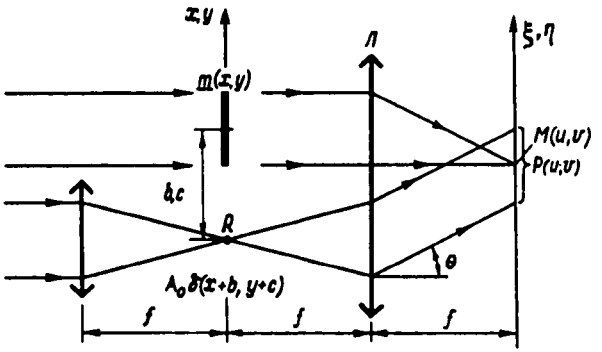


Рис. 4.12. Схема изготовления голографического фильтра

$$\dot{M}(u, v) = M\,e^{j\varphi} = \mathcal{F}[\underline{m}(x, y)],$$

где, как и раньше, $u = \xi/(\lambda f)$; $v = \eta/(\lambda f)$.

Опорный точечный источник R с амплитудой A_0 в плоскости предмета можно описать с помощью δ-функции: $A_0\delta(x+b, y+c)$, где b, c — сдвиг R относительно начала координат x, y. В фокальной плоскости линзы $Л$ в результате преобразования Фурье опорный сигнал представляет собой плоскую волну с комплексной амплитудой

$$\dot{P}(u, v) = A_0\,e^{-j\varphi_0} = \mathcal{F}[A_0\delta(x+b, y+c)],$$

где $\varphi_0 = bu + cv$.

Комплексная амплитуда результирующего электрического поля в плоскости фотопластинки

$$E(u, v) = M(u, v) + \dot{P}(u, v).$$

Следовательно, регистрируемая в фильтре интенсивность

$$I(u, v) = E(u, v)\dot{E}^{*}(u, v) = (M\,e^{j\varphi} + A_0\,e^{-j\varphi_0})(M\,e^{-j\varphi} + A_0\,e^{j\varphi_0}) =$$
$$= M^2 + A_0^2 + M\,e^{-j\varphi}\,A_0\,e^{-j\varphi_0} + M\,e^{j\varphi}\,A_0\,e^{j\varphi_0}.$$

Для простоты расчетов предположим, что комплексное амплитудное пропускание T фильтра пропорционально распределению интенсивности $I(u, v)$, зарегистрированной в голограмме фильтра, причем коэффициент пропорциональности равен единице. Это предположение верно в том случае, когда интенсивность опорного пуска A_0^2 примерно в 10 раз боль-

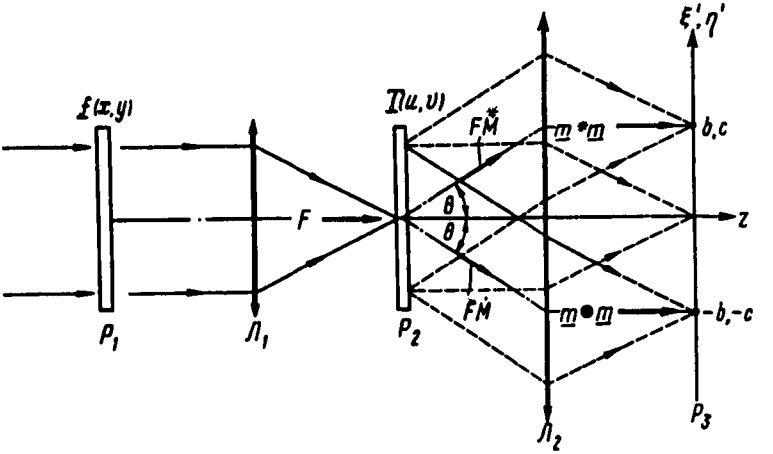


Рис. 4.13. Схема корреляционной фильтрации

ше величины M^2, а также когда экспозицию и проявление фотопленки производят при контрасте $\nu = -2$. Для других значений ν комплексную амплитуду на выходе фильтра умножают на $\nu/2$, хотя комплексная амплитуда волны, вышедшей из фильтра, практически пропорциональна $I(u, v)$.

Поиск слова «Иванов» на одной из страниц книги производят по схеме, представленной на рис. 4.13. Диапозитив страницы располагают в плоскости P_1 и освещают плоской волной. Предполагают, что диапозитив воспроизводится линзами $Л_1$ и $Л_2$ без изменения масштаба. В плоскости P_2 располагают голограмму согласованного фильтра слова «Иванов». Комплексная амплитуда поля на входе фильтра определяется фурье-образом сигнала

$$F(u, v) = \mathcal{F}[\underline{f}(x, y)] = M(u, v) + \dot{N}(u, v),$$

где $\underline{f}(x, y)$ — распределение комплексных амплитуд на выходе диапозитива P_1; $\dot{N} = \mathcal{F}[\underline{n}]$ — фурье-образ шума ($\underline{n}$), т. е. сигнала от всех остальных слов.

Комплексная амплитуда поля на выходе фильтра

$$F(u, v)T(u, v) = (M^2 + A_0^2)F(u, v) + F(u, v)\dot{M}(u, v)A_0\, e^{-j\varphi_0} + \\ + F(u, v)M(u, v)A_0\, e^{j\varphi_0}. \qquad (4.94)$$

Световое поле после фильтра образует три пучка. Первый член уравнения (4.94) определяет основной пучок, направленный вдоль оптической оси z системы со структурой, близкой к Фурье-образу сигнала $F(u, v) = \mathcal{F}[\underline{f}(x, y)]$, если $M^2 \ll A_0^2$. Второй пучок (второй член в уравнении (4.94)) идет в том же направлении, что и опорный пучок P, используемый в процессе получения фильтра (см. рис. 4.12). Третий пучок, соответствующий последнему члену уравнения (4.94), отклоняется относительно оси в противоположную сторону.

Отфильтрованное поле подвергается преобразованию Фурье с помощью линзы $Л_2$. В процессе преобразования возникают функции, называемые сверткой и корреляцией. По определению, функция *кросс-корреляции*

$$\varphi_{12}(\tau) = \int_{-\infty}^{\infty} \dot{f}_1(t)\dot{f}_2^*(\tau + t)\,\mathrm{d}t = \dot{f}_1 * \dot{f}_2.$$

Кросс-корреляцию можно представить как предел суммы произведений предмета $\dot{f}_1(x)$ на предмет $\dot{f}_2^*(x)$ при всевозможных сдвигах их друг относительно друга около точки τ. *Автокорреляцию* можно рассматривать как частный случай кросс-корреляции при одинаковых значениях функций обоих предметов.

Преобразование Фурье первого слагаемого уравнения (4.94)

$$\mathcal{F}[(M^2 + A_0^2)F(u, v)] = \underline{f}(x, y) * \dot{\mathcal{F}}^{-1}(M^2 + A_0^2)^*$$

приводит к выражению, определяющему компоненту, сосредоточенную вблизи точки $\xi' = 0$, $\eta' = 0$.

Сигнал, передаваемый верхним боковым пучком перед линзой, состоял из двух слагаемых:

$$\dot{F}\dot{M}^* = (\dot{M} + \dot{N})\dot{M}^* = \dot{M}\dot{M}^* + \dot{N}\dot{M}^*.$$

---

* $\int_{-\infty}^{\infty} M^2(u)A_0\,\mathrm{e}^{j2\pi u'u}\,\mathrm{d}u = \int_{-\infty}^{\infty} M^2(u)\,\mathrm{e}^{j2\pi u'u}\left[\int_{-\infty}^{\infty} \underline{f}(x)\,\mathrm{e}^{j2\pi ux}\,\mathrm{d}x\right]\mathrm{d}u =$

$= \int_{-\infty}^{\infty} \underline{f}(x)\left[\int_{-\infty}^{\infty} M^2(u)\,\mathrm{e}^{j2\pi(u'+x)u}\,\mathrm{d}u\right]\mathrm{d}x = \int_{-\infty}^{\infty} \underline{f}(x)\left[\int_{-\infty}^{\infty} M^2(u)\,\mathrm{e}^{j2\pi(u'+x)u}\,\mathrm{d}u\right]^*\mathrm{d}x =$

$= \int_{-\infty}^{\infty} \underline{f}(x)[\dot{\mathcal{F}}^{-1}(M^2)]_{u'+x}\,\mathrm{d}x = \int_{-\infty}^{\infty} \underline{f}(u'+x)[\dot{\mathcal{F}}^{-1}(M^2)]_{u'}\,\mathrm{d}x = \underline{f} * \dot{\mathcal{F}}^{-1}(M^2).$

После фурье-преобразования линзой $Л_2$ функция $\dot{M}\dot{M}$ превращается в сигнал автокорреляции искомого образа $\underline{m}(x)$:

$$\mathcal{F}(\dot{M}\dot{M}) = \underline{m} * \underline{m}^{*}.$$

Этот сигнал появится в центре верхнего изображения. Сигнал $\dot{N}\dot{M}$ будет распространяться в направлении $\dot{M}\dot{M}$, но из-за случайной природы шума значение кросс-корреляции его фурье-образа $\mathcal{F}(\dot{N}\dot{M}) = \underline{n} * \underline{m}$ будет достаточно мало́.

Рассмотренная система обеспечивает максимальное подавление шума, поскольку оптический фильтр эквивалентен комплексно сопряженному фильтру. Этот фильтр в обоих случаях пропускает только те частоты, которые содержатся в полезном сигнале. В рассмотренной системе роль сигнала с положительной амплитудой играет функция автокорреляции $\underline{m} * \underline{m}$ в виде светового пятна на плоскости P_3 (см. рис. 4.13), в которое линза $Л_2$ преобразует волну, формирующуюся в плоскости P_2. Интенсивность корреляционного максимума может быть использована как мера сходства между входным сигналом и эталоном (фильтром).

До сих пор предполагалось, что положение центра фильтруемого сигнала совпадает с положением центра сигнала, по которому изготовлялся фильтр ($x = y = 0$). Только в этом случае сигнал автокорреляции $\underline{m} * \underline{m}$ появится в центре верхнего бокового изображения ($\xi' = b$, $\eta' = c$). Если же анализируемый сигнал смещен относительно опорного сигнала на расстояние $\mathbf{x}_1$, когда $\underline{m}_1 = \underline{m}(\mathbf{x} + \mathbf{x}_1)$, то

$$\dot{M}_1(u, v) = \int_{-\infty}^{+\infty}\int \underline{m}(x + x_1, y + y_1)\, e^{j2\pi(ux+vy)}\, dx\, dy =$$

$$= \int_{-\infty}^{+\infty}\int \underline{m}(x + x_1, y + y_1)\, e^{j2\pi[u(x+x_1)+v(y+y_1)]} e^{-j2\pi(ux_1+vy_1)} d(x + x_1) d(y + y_1) =$$

$$= \dot{M}(u, v)\, e^{-j2\pi(ux_1+vy_1)}.$$

---

* $\int_{-\infty}^{\infty} \dot{M}(u)\dot{M}(u) e^{j2\pi u'u} du = \int_{-\infty}^{\infty} \dot{M}(u) e^{j2\pi u'u} \left[\int_{-\infty}^{\infty} \underline{m}(x) e^{j2\pi ux} dx\right] du =$

$= \int_{-\infty}^{\infty} \underline{m}(x) \left[\int_{-\infty}^{\infty} \dot{M}(u) e^{j2\pi(u'+x)u} du\right] dx = \int_{-\infty}^{\infty} \underline{m}(x) \left[\int_{-\infty}^{\infty} \dot{M}(u) e^{-j2\pi(u'+x)u} du\right]^{*} dx =$

$= \int_{-\infty}^{\infty} \underline{m}(x) \dot{m}(x + u') dx = \underline{m} * \underline{m}.$

Таким образом, спектр $\dot{M}_1$ отличается от спектра М только множителем $e^{-j2\pi(ux+vy_1)}$, который приводит к дополнительному наклону бокового пучка. Операция фильтрации позволяет получить

$$\dot{M}_1\dot{M}\,e^{-j\varphi_0} = \dot{M}_1\dot{M}\,e^{-j2\pi(ub+vc)} = \dot{M}_1\dot{M}\,e^{-j2\pi[u(b+x_1)+v(c+y_1)]}.$$

Поэтому если сигнал $\dot{M}\dot{M}$ создавал волну, направленную под углами, определяемыми фазой $\varphi_0 = 2\pi(ub + vc)$, то волна сигнала $\dot{M}_1\dot{M}$ распространяется под углом, который соответствует фазе $2\pi[u(b+x_1)+v(c+y_1)]$. Изменение угла в области фурье-образа приводит к изменению координаты выходного сигнала в плоскости P_3 без каких-либо искажений. Следовательно, при наличии в электромагнитном поле $\underline{f}(x, y)$ в плоскости P_1 нескольких сигналов $\underline{m}(x + x_i, y + y_i)$, расположенных в различных местах (x_i, y_i), в плоскости выходного сигнала P_3 появляются яркие пятна в тех местах $(b + x_i, c + y_i)$, где находились исходные сигналы $\underline{m}_i$. Однако это будут не изображения предметов (слова «Иванов»), а их автокорреляции $\underline{m} * \underline{m}$. Пространственное разделение отфильтрованных изображений получается благодаря достаточно большому сдвигу опорного пучка $R(b > x_i, c > y_i)$ относительно фильтра $\underline{m}(x, y)$ в процессе получения его голограммы (см. рис. 4.12).

Сигнал $\dot{M}\dot{M}$ после фильтра P_2 отклонится от оптической оси на тот же угол, что и сигнал $\dot{M}\dot{M}$, но в противоположную сторону, определяя в выходной плоскости P_3 сигнал свертки $\mathcal{F}(\dot{M}\dot{M}) = \underline{m} \circledast \underline{m}$.

Для сигнала шума типа $\dot{N}\dot{M}$ свертка, как и в предыдущем случае, мала из-за рассеяния по всему изображению.

Подобным же образом осуществляется распознавание букв и геометрических фигур, а также определение их координат. Рассмотренный метод распознавания образов (оптическая фильтрация с помощью голограммы) может быть использован для автоматизации ввода в машину графических заданий, в частности для автоматического чтения и автоматизации библиотечного дела.

Голографические фильтры входят составной частью в корреляционные координаторы автоматических систем навигации и управления движением летательных объектов (самолетов, космических кораблей и т. д.). В некоторых случаях необходимо использовать в качестве ориентира всю местность по курсу полета, т. е. местность с практически неограниченным количеством участков земной или лунной поверхности в качестве сложных ориентиров. Конструкция одного из первых голографических

координаторов для автономной навигации космического аппарата по наземным ориентирам изображена на рис. 4.14. Он состоит из телеобъектива *1*, формирующего текущее изображение местности на фототермопластическом материале или фотопленке *12*, газового лазера *6*, магазина *10* корреляционных фильтров *9*, телевизионной камеры *7*, считывающего устройства *4* на видиконе и блока управления *5*. Для уменьшения габаритных размеров в оптической схеме, содержащей коллиматор *3*, преобразующую *2* и отображающую *8* линзы и призму, используется система неподвижных зеркал *11* и подвижное зеркало *13*. Ход световых лучей при формировании изображения местности (ориентира) на регистрирующей среде показан пунктирными линиями. Ход световых лучей при опознавании и определении координат ориентира показан сплошными линиями (подвижное зеркало при этом опущено). Координаты ориентира определяют положение яркой точки (максимума корреляционного поля) на мишени видикона. Опознавание нескольких ориентиров одновременно позволяет найти не только координаты объекта, но и углы его ориентации.

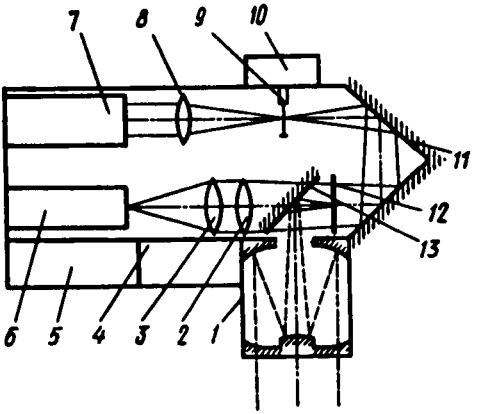


Рис. 4.14. Голографический координатор

Было установлено, что опознавание даже отдельных малоконтрастных изображений участков пустыни происходило с отношением сигнал/помеха в выходной плоскости коррелятора не ниже 100. Оказалось, что опознавание сложного наземного ориентира возможно и в том случае, когда облака закрывают до 90% его поверхности. Результаты опознавания ориентиров на поверхности Луны практически не зависели от угла возвышения Солнца (отношение сигнал/помеха было не ниже 30). Опознавание участков звездного неба было проведено при отношении сигнал/помеха около 400, точность измерения местоположения ориентира на снимке имела величину порядка 10^{-4} линейного размера поля зрения и была ограничена пределами разрешающей способности видикона. На высоте полета 180 км ошибка измерения координат ориентиров на местности была не ниже 15 м*.

---

* *Василенко Г. И.* Голографическое опознавание образов. —М.: Советское радио, 1977.

Комплексно сопряженный фильтр (коррелятор) — основной узел оптической аналоговой вычислительной машины. Операции в такой машине, как и во всякой когерентной оптической системе при интерференционной форме записи информации, выполняют по программе вычислительных машин одновременного действия. Информацию в оптический коррелятор обычно вводят в виде фотографий на диапозитивах, текста на микрофильмах или с помощью управляемых транспарантов. Если исключить время на образование формы записи информации, производящейся вне оптической машины, то можно ввести понятие быстродействия блока оптического коррелятора. Например, умножение вектора, имеющего 10^4 компонент, на квадратную матрицу с $10^4 \times 10^4$ элементами отнимает около 1 мс, т. е. происходит со скоростью 10^{11} оп/с. Это примерно в 10^5 раз выше скорости обычных электронных вычислительных машин.

Основной проблемой при проектировании оптического процессора является проблема расчета и изготовления пространственно-частотного фильтра. Фильтры могут быть сформированы не только как фурье-голограммы, но и синтезированы с помощью ЭВМ методами цифровой голографии. Запись фильтра в виде фурье-голограммы не является универсальным методом, так как используется только для записи действительной неотрицательной двухмерной функции, представляемой путем амплитудной модуляции транспаранта. Синтез фильтров методами цифровой голографии позволяет изготовлять комплексные фильтры общего типа с произвольными передаточными функциями. Попытки применения ЭВМ для автоматического опознавания образов часто оказываются недостаточно эффективными. Это связано с тем, что большинство ЭВМ являются машинами последовательного действия и не рассчитаны на параллельный анализ больших массивов данных при опознавании сложных изображений. В таких ЭВМ поиск объектов производится в ходе последовательного перебора всех элементов разложения изображения, причем в каждый момент времени выполняется только одна или несколько вычислительных операций, а анализ изображения осуществляется по малым его участкам. Время обработки изображения с числом элементов разложения, соответствующим стандарту вещательного телевидения, даже на ЭВМ с высокой производительностью может исчисляться десятками часов. Электронные вычислительные комплексы параллельного действия, использующие матричные процессоры и многоканальные аналоговые вычислители, оказываются сложными, тяжелыми и дорогими.

Голографические вычислительные устройства, приспособленные естественным образом к параллельному анализу больших информационных массивов в виде изображений, оказываются проще и экономичнее. Они позволяют организовать процесс автоматического анализа по программе одновременного действия, при которой происходит не только опознава-

ние и локализация образа, но и вызов анализируемым образом отклика в виде другого образа, используя фантомное изображение или ассоциативную голографическую память. Их быстродействие намного превышает быстродействие электронных корреляторов как дискретных, так и аналоговых.

Голографический координатор для навигации космических аппаратов (см. рис. 4.14) относится к одному из гибридных оптико-цифровых устройств, сочетающих в себе преимущества как оптических процессоров, так и ЭВМ. Их перспективность следует из того, что в большинстве случаев вычисление корреляционных функций и спектров сигналов не является самоцелью. Результаты таких вычислений предназначены для дальнейшей оперативной обработки: принятия определенных решений, управления работой оптических процессоров и т. п. Эти функции успешно выполняет ЭВМ. На оптическое устройство возлагается быстродействующая обработка массивов информации, связанная со спектральным анализом, вычислением корреляционных функций и т. д. Для вывода информации наряду с видиконами и матрицами фотодиодов в гибридных оптико-цифровых устройствах используются многоэлементные фотоприемники с зарядовой связью (ПЗС), входящие в состав единой интегральной структуры, реализующей регистры сдвига и другие элементы для последующей обработки сигналов.

В рассмотренном координаторе наглядно соединяются две характерные особенности современной оптики: 1) проникновение в оптику идей информатики, когда линейная операция над оптическим изображением рассматривается методами пространственно-частотного фурье-анализа и теории систем с конечной полосой пропускания; 2) введение в оптические системы ЭВМ и специализированных процессоров, которые обеспечивают возможность адаптации, осуществляют количественную обработку информации, содержащуюся в оптическом сигнале.

Наиболее эффективным путем миниатюризации оптических процессоров является переход к замкнутому циклу, при котором световой пучок в свободное пространство не выводится, обработка сигналов выполняется на *оптических волноводах* и функциональных узлах.

Одномерное преобразование Фурье может быть реализовано в пленочном волноводе с меняющимся в поперечном сечении по параболическому закону показателем преломления и обладающим линзоподобными свойствами. Двухмерное преобразование Фурье реализует оптическое волокно с квадратичной зависимостью диэлектрической проницаемости. Обеспечение соответствующей интерференции волн в оптических волноводах приводит к схеме твердотельного микроволноводного коррелятора, аналогичной классической схеме согласованной фильтрации (см. рис. 4.13). Обработка сигнала без вывода световых пучков в свободные пространства обеспечивает кроме миниатюрности также нечувствительность к внешним помехам (фоновым засветам, вибрациям, пыли) и удобство эксплуатации.

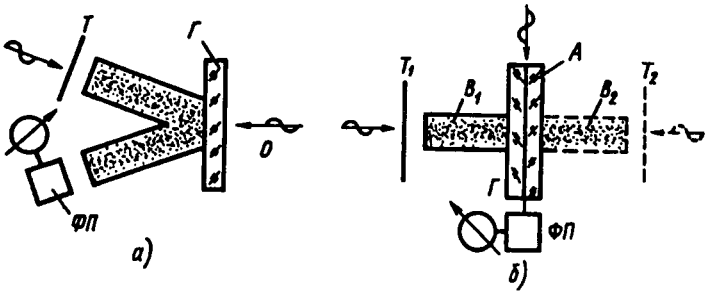


Рис. 4.15. Схемы малогабаритных оптических процессоров

В процессоре (рис. 4.15, *а*) фурье-спектр транспаранта *Т* формируется в плоскости общего торца волноводного разветвителя с нанесенной на него толстослойной фоторегистрирующей средой *Г*, используемой для формирования голограммы встречным пучком *О*. Дифракционная эффективность получаемого таким образом фильтра повышается, а частотно-селективные свойства полученной голограммы позволяют снизить требования к когерентности лазера при распознавании.

В оптическом процессоре (рис. 4.15, *б*) для получения голографического фильтра эталонного сигнала используется опорный сигнал, поступающий по тонкопленочному волноводу в направлении стрелки *А* под углом 90° к эталонному сигналу через транспарант T_1. Фоторегистрирующий слой *Г* прилегает к поверхности тонкопленочного световода. На этапе распознавания во входную плоскость процессора проступает двухмерный анализируемый сигнал от транспаранта T_2. В плоскости корреляционного фильтра *Г* световод B_2 с параболическим профилем формирует такое распределение поля, соответствующее фурье-спектру сигнала, которое, проходя через фильтр, возбуждает в тонкопленочном волноводе поверхностную волну. Тонкопленочная линза на планарном волноводе фокусирует восстановленную поверхностную волну и направляет ее во входной зрачок фотоприемника ФП. Фотоприемник по интенсивности сигнала корреляции определяет меру оптического сходства распознаваемого и эталонного сигналов. Оперативность процессора повышается возможностью независимого перебора эталонных и распознаваемых транспарантов при последовательном их анализе и использовании первого фурье преобразующего устройства B_1 при записи эталонного фильтра.

Одним из основных функциональных узлов *интегрально-оптических корреляторов*, спектроанализаторов и других интегрально-оптических устройств обработки информации являются фокусирующие элементы. К простейшим из них относится планарная линза, образуемая локальным изменением толщины или показателя преломления волноводного

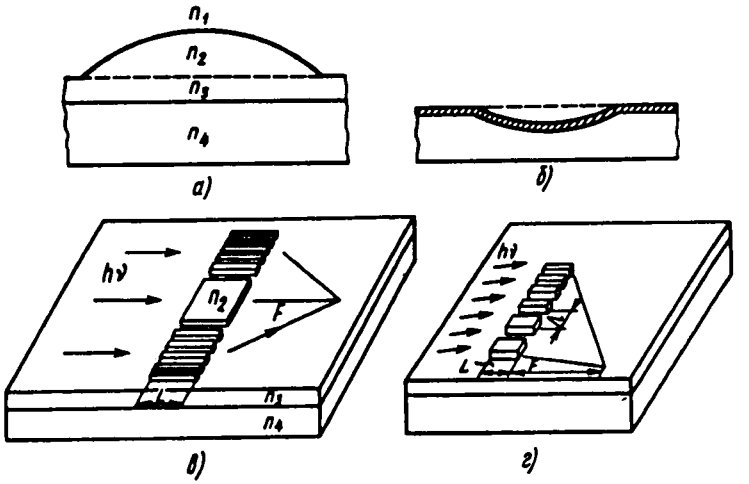


Рис. 4.16. Волноводные линзы:
а — Люнеберга; *б* — геодезическая; *в*, *г* — с тонкой и толстой решетками

слоя. Ее разрешающая способность ограничивается аберрациями, причем аберрация комма оказывается принципиально неустранимой.

Поиски новых принципов построения волноводных линзовых устройств привели к созданию двухмерных интегрально-оптических систем с удовлетворительными характеристиками в виде линз Люнеберга, геодезических линз и линз типа дифракционной решетки.

Тонкопленочная *линза Люнеберга* (рис. 4.16, *а*) располагается на подложке с показателем преломления n_4, на которую последовательно нанесены два волноводных слоя с показателями преломления n_2 и n_3. Фокусирующая и разрешающая способность линзы тем выше, чем больше разность показателей преломления волноводов и подложки. Верхний волноводный слой обладает круговой симметрией. В линзе необходима коррекция сферической аберрации.

В *геодезической линзе* волновод в подложке создается после изготовления в ней углубления или возвышения сферической или асферической формы (рис. 4.16, *б*). Линзу можно применять в многомодовых устройствах, так как постоянство параметров волноводного слоя исключает хроматическую аберрацию. Для снижения других аберраций применяют различные компенсационные методы, в частности, в области углубления формируют пленочную линзу с обратным знаком аберрации.

Дифракционные линзы технологически проще геодезических и линз Люнеберга. В волноводной линзе Френеля необходимый фазовый сдвиг получают за счет изменения эффективного показателя преломления

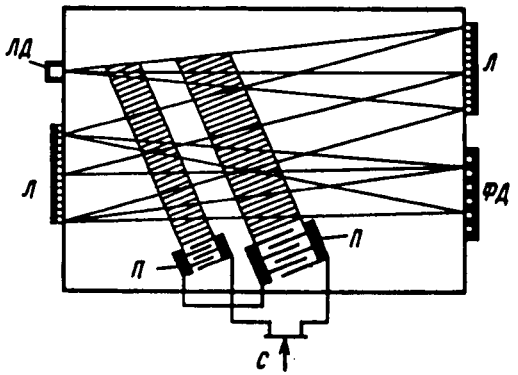


Рис. 4.17. Интегрально-оптический спектроанализатор:
ЛД — лазерный диод; *Л* — брегговская линза; *ФД* — матрица фотодекторов; *П* — преобразователи

волновода или нанесением на волновод металлического или диэлектрического покрытия. Поглощение или фазовый сдвиг в нечетных зонах Френеля осуществляется нанесением или вытравливанием соответствующих областей волновода (рис. 4.16, *в*). В градиентных линзах параметры волноводного слоя меняются плавно по определенному закону. Это позволяет увеличить их дифракционную эффективность в 2,5 раза (по сравнению с эффективностью ступенчатых линз) и довести ее до 90%. Однако при работе с наклонными пучками происходит значительное ухудшение характеристик линзы. Брегговские линзы (рис. 4.16, *г*) называют также линзами с толстой решеткой, в отличие от френелевских, являющихся линзами с тонкой решеткой (их ширина L в направлении распространения волны меньше).

Схема широкополосного акусто-оптического *анализатора спектра* сигнала *С* приведена на рис. 4.17. В качестве фокусирующих элементов в нем используются брегговские линзы с переменным периодом, работающие на отражение. Размеры подложки удалось уменьшить по площади примерно в 10 раз за счет неосевой фокусировки линз. Применение двух встречно-штыревых акустооптических преобразователей с разными центральными частотами (500 и 1000 МГц) позволило довести полосу частот спектроанализатора до 1 ГГц (330...1333 МГц).

Оптическое *запоминающее устройство*, использующее волноводные голограммы, выгодно отличается от их аналога, построенного на объемных компонентах. Голограмма площадью 1 мм^2 может быть сформирована в тонком слое светочувствительной среды *1*, нанесенной на тонкопленочном волноводе *6* (рис. 4.18). Опорный луч *4* вводится через призму *7* и подложку *5*. Предметная волна *2* формируется транспарантом *3*. Если в качестве фоточувствительной среды используется As_2S_3, то для записи голограммы удобно применить аргоновый лазер (λ = 514,5 нм), а для восстановления — гелий-неоновый лазер (λ = 632,8 нм). Восстановление волной типа Н позволяет достигнуть 92% дифракционной эффективности. На пластине размером 5×5 см можно уместить около 1000 голограмм, каждая из которых на 1 мм^2 мо-

жет содержать около 10^4 бит информации. Коммутацию голограмм можно проводить полосковыми связанными волноводами или управляемыми брэгговскими решетками.

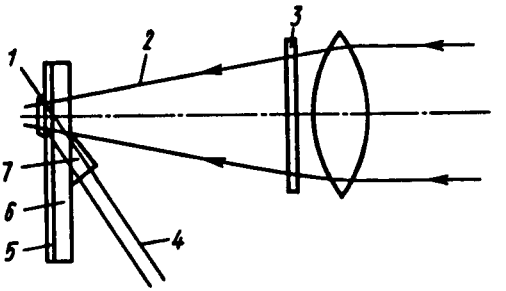


Рис. 4.18. Схема записи волноводной голограммы

Цифровой коррелятор с пространственным интегрированием изображен на рис. 4.19. Он сформирован на пластине *1* ниобата лития и рассчитан на обработку 32-разрядных кодов при быстродействии 32 МБит/с. Фильтрующая функция (эталонный код) внешними электронными устройствами регистрируется на интегрально-оптическом пространственном модуляторе *10*. При этом в результате электрооптического изменения коэффициента преломления волноводного слоя образуется стационарная (неподвижная) дифракционная решетка, которая дает возможность модулировать однородный входной световой пучок *4* параллельно по нескольким информационным каналам (количество каналов равно количеству секций модулятора). Изменяющийся во времени входной сигнал на клеммах *5* формирует при помощи электроакустического преобразователя *6* движущуюся дифракционную решетку, осуществляющую ввод данных поперек светового пучка. Единица кодируется высокочастотной поверхностной акустической волной, а «минус единица» — низкочастотной. Умножение выполняется в результате последовательного прохождения через неподвижную и движущуюся дифракционные решетки; интегрирование обеспечивает линза *8*. Световой пучок, соответствующий совпадающим разрядам входной посылки, попадает на один из фотодетекторов *7*, а несовпадающие — на другой. Полная биполярная функция корреляции формируется после вычитания сигналов фотодетекторов. Призменные элементы связи *3* и *9* расположены на контактных площадках *2*. Несмотря на относительно низкую скорость ввода и вывода данных (32 МБит/с),

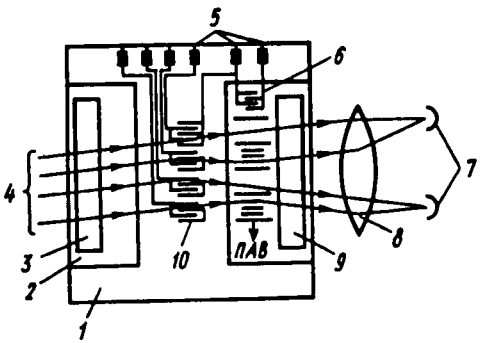


Рис. 4.19. Цифровой оптический коррелятор

внутренняя скорость вычислений эквивалентна $4 \cdot 10^9$ оп/с. Время, требуемое на выполнение операций умножения и суммирования, определяется временем прохождения светового пучка через коррелятор и составляет примерно 10^{-12} с.

Применение интегрально-оптической конструкции корреляторов и поверхностных акустических волн обеспечивает компактность, надежность и стойкость к механическим воздействиям, снижает мощность входного электрического сигнала до десятков милливатт.

§ 4.14. Математические операции, осуществляемые оптическими методами

Функции устройств, предназначенных для обработки информации, сводятся в общем случае к решению различных систем уравнений, описывающих анализируемые процессы. Решение этих уравнений предусматривает выполнение простейших операций (алгебраических или интегродифференциальных). Как было показано, простейшие когерентные оптические системы способны выполнять многие из подобных операций.

Комплексно сопряженный фильтр, изображенный на рис. 4.13, представляет собой когерентное оптическое устройство, с помощью которого осуществляются операции *свертки* и *корреляции*. При рассмотрении элементов оптической ячейки было показано, что в зависимости от расположения выходного устройства линза может осуществлять *преобразования Фурье* и *Френеля*. С помощью оптических систем можно выполнять и другие математические операции, в частности умножение, интегрирование и дифференцирование функций.

В основе *операции умножения* двух (и более) комплексных функций лежит свойство плоского транспаранта с пропусканием $\underline{T} = \underline{f}(x, y)$ при освещении волной $E_1 = \varphi(x, y)$ формировать волну (4.78) $E_2 = g(x, y) = \underline{f}(x, y)\varphi(x, y)$, описываемую произведением функций пропускания и падающей волны. При задании распределения $\varphi(x, y)$ с помощью другого транспаранта в плоскости P_1 (см. рис. 4.10, *а*), его следует, как и транспарант $\underline{f}(x, y)$ в плоскости P_3, располагать от линзы на расстоянии $D_1 = D_2 = 2F$ в соответствии с уравнением линзы (4.89). В схеме рис. 4.13 в плоскости P_2 осуществляется умножение (4.94) фурье-образов $F(u, v)M(u, v) = \dot{\Phi}_1(u, v)$ и $F(u, v)\dot{M}(u, v) = \dot{\Phi}_2(u, v)$ функций $\underline{f}(x, y)$ и $\underline{m}(x, y)$.

Операцию интегрирования некоторой функции $\underline{f}(x, y)$, описывающей пространственное распределение света, $G(\mathbf{r}) = \int_{P_1} \underline{f}(\mathbf{x})\,\mathrm{d}\mathbf{x}$ легко осуществить с помощью элементарной оптической системы (см. рис. 4.10, *а*), в которой $D_1 = D_2 = F$ и $\mathbf{r} = 0$, поскольку при этом экспоненциальный множитель преобразования Фурье (4.91) превращается в единицу. Практически реализовать второе условие позволяет диафрагма с точечным отверстием, либо оптический приемник малого размера, располагаемый точно на оптической оси ($x = y = 0$) в плоскости P_3.

Из теоремы о спектре производной сигнала

$$F_n(\chi) = \int_{-\infty}^{\infty} \frac{\mathrm{d}^n f}{\mathrm{d}\mathbf{x}^n} \mathrm{e}^{-j\chi\mathbf{x}}\,\mathrm{d}\mathbf{x} = (j\chi)^n F(\chi)$$

следует, что *операция дифференцирования* производится в области пространственных частот, т. е. в области преобразования Фурье. Например, чтобы найти первую производную функции, необходимо ее спектр сначала умножить на χ, а затем произвести обратное преобразование Фурье. Это означает, что в плоскости пространственных частот необходимо поставить транспарант с нулевой прозрачностью в центре и постепенно (линейно) увеличивающейся прозрачностью к краям (по мере возрастания пространственной частоты $\chi = \sqrt{u^2 + v^2}$), а также полуволновую пластинку, осуществляющую сдвиг фазы на 180° (перемену знака) для отрицательных пространственных частот. Для получения второй производной транспарант должен иметь прозрачность, изменяющуюся по закону $\chi^2 = u^2 + v^2$.

В отличие от ЭВМ в оптических системах затруднено выполнение некоторых элементарных операций. Например, для *вычитания* из одной комплексной функции $\underline{f}_1(x)$ другой $\underline{f}_2(x)$ необходимо выполнить следующие операции:

1) образовать суммы каждого из сигналов с δ-функцией

$$\underline{\varphi}_1(\mathbf{x}) = a\delta(\mathbf{x} - \mathbf{x}_0) + \underline{f}_1(\mathbf{x}),$$

$$\underline{\varphi}_2(\mathbf{x}) = a\delta(\mathbf{x} - \mathbf{x}_0) + \underline{f}_2(\mathbf{x});$$

2) получить сначала их фурье-образы $\Phi_i(\chi) = \mathcal{F}(\underline{\varphi}_i) = a\mathrm{e}^{-j\chi\mathbf{x}_0} + F_i(\chi)$, а затем энергетические спектры $S_i(\chi) = \Phi_i^2 = \Phi_i\dot{\Phi}_i$;

3) суммировать полученные спектральные плотности

$$S_1(\chi)+S_2(\chi)=2a^2+F_1^2(\chi)+F_2^2(\chi)+$$
$$+[F_1(\chi)-F_2(\chi)]e^{-j\chi x_0}+[\overset{*}{F}_1(\chi)-\overset{*}{F}_2(\chi)]e^{j\chi x_0};$$

4) выполнить обратное преобразование Фурье:

$$\mathcal{F}^{-1}(S_1+S_2)=2a^2\delta(\mathbf{x})+\underline{f}_1*\underline{f}_1+\underline{f}_2*\underline{f}_2+a[\underline{f}_1(\mathbf{x})-\underline{f}_2(\mathbf{x})]*\delta(\mathbf{x}-\mathbf{x}_0)+$$
$$+a\delta(\mathbf{x}-\mathbf{x}_0)*[\overset{*}{f}_1(\mathbf{x})-\overset{*}{f}_2(\mathbf{x})]=2a^2\delta(\mathbf{x})+\underline{f}_1*\underline{f}_1+\underline{f}_2*\underline{f}_2+$$
$$+a[\underline{f}_1(\mathbf{x}_0-\mathbf{x})-\underline{f}_2(\mathbf{x}_0-\mathbf{x})]+a[\underline{f}_1(\mathbf{x}+\mathbf{x}_0)-\underline{f}_2(\mathbf{x}+\mathbf{x}_0)].$$

В полученном выражении первый член есть δ-функция в точке $x = 0$, второй и третий члены — автокорреляционные функции, симметричные относительно точки $x = 0$, а четвертый и пятый члены — искомые разности функций $\underline{f}_1(x)-\underline{f}_2(x)$, разнесенные относительно точки $x = 0$ на расстояние $\pm x_0$.

Цилиндрическая линза может выполнять ряд операций над числовыми массивами (например логическое суммирование элементов строк или столбцов). Цилиндрический линзовый растр может осуществлять мультипликации числового столбца или строки, а также операцию транспонирования (замену строк столбцами). В последнем случае ось цилиндрических линз располагается вдоль главных диагоналей числовых ячеек. Волоконно-оптические жгуты при соответствующем расположении оптических волокон без труда реализуют операции совмещения, разветвления и сдвига числовой картины, а также ее транспонирование.

Таким образом, выполняя операции умножения и преобразования Фурье, можно осуществить широкий спектр математических операций по обработке информации. Основным преимуществом современных оптических вычислительных устройств является возможность практически мгновенного выполнения математических операций над комплексными функциями двух переменных по программе одновременного действия, что позволяет создать высокопроизводительные системы обработки *массивов* информации.

Помимо преобразований Фурье в оптических устройствах могут быть реализованы и другие линейные интегральные преобразования. В общем случае *линейное интегральное преобразование* позволяет представить преобразованную функцию (трансформанту) в выходной плоскости (u, v) в виде

$$\varphi(u,v)=\iint_{\Omega} f(x,y)K(x,y,u,v)\,\mathrm{d}x\,\mathrm{d}y,$$

где $f(x, y)$ — исходная функция (оригинал), представляющая собой распределение комплексной амплитуды светового поля во входной плоско-

сти xy оптического процессора; Ω — область интегрирования, определяемая апертурой входного зрачка оптического процессора; $K(x, y, u, v)$ — ядро преобразования, т. е. импульсная реакция оптического процессора.

В оптических системах, реализующих преобразование Фурье, ядро $K = \exp[-j(xu + yv)]$. Помимо этого существует ряд интегральных преобразований, представляющих собой разложение исходной функции в ряд по системе ортогональных функций $K_{ij}(x, y)$ двух переменных. Совокупность коэффициентов разложения

$$a_{ij} = \iint_{\Omega} f(x, y) K_{ij}(x, y)\, \mathrm{d}x\, \mathrm{d}y$$

называют *обобщенным спектром исходной функции* $f(x, y)$. Подобные системы образуют, например, *функции Уолша.*

Одномерную функцию Уолша в зависимости от координаты x в общем случае определяют на бесконечном интервале $(-\infty < x < \infty)$ по рекуррентной формуле

$$\mathrm{Wal}(m, x) = \mathrm{Wal}(2p + q, x) =$$

$$= (-1)^{\{p/2+q\}}[\mathrm{Wal}(p, 2x + 1/2) + (-1)^{p+q}\, \mathrm{Wal}(p, 2x - 1/2)],$$

где $q = 0$ или 1; $p = 0, 1, 2, \ldots$; $\{p/2 + q\}$ — наибольшее целое, не превосходящее $(p/2 + q)$, причем

$$\mathrm{Wal}(0, x) = \begin{cases} 1; & -1/2 < x < 1/2, \\ 0; & x < -1/2, x > 1/2. \end{cases}$$

Таким образом, функции Уолша принимают только два значения: +1 и –1. Они легко реализуются в оптике при помощи фазовых пространственно-временны́х модуляторов.

Двухмерные функции Уолша являются произведением соответствующих одномерных функций

$$\mathrm{Wal}(m, n, x, y) = \mathrm{Wal}(m, x)\, \mathrm{Wal}(n, y),$$

где m, n — номера функций, равные количеству смен знака на единичном интервале $(0 < x, y < 1)$. На этом интервале они образуют полную ортонормированную систему функций.

Одно из применений интегральных преобразований в оптике связано с тем, что они позволяют осуществить наиболее эффективные методы *кодирования* изображений. Эти методы в значительной степени определяют эффективность работы оптических устройств обработки, хранения, передачи и отображения информации. Кодирование позволяет решить следующие задачи: 1) повышение пропускной способности каналов передачи и производительности устройств обработки в результате уменьшения избыточности информации; 2) увеличение емкости голограмм на

один участок регистрирующей среды, а также помехозащищенности устройств обработки и хранения информации путем применения специальных кодов для выявления и устранения ошибок и сбоев; 3) формирование и выделение информативных признаков в устройствах распознавания изображений; 4) формирование ключевых слов (признаков) в устройствах ассоциативного поиска информации.

Существенно повысить эффективность схем кодирования и преобразования оптических сигналов можно за счет применения в них оптических пространственных модуляторов или транспарантов, управляемых электрическими или оптическими сигналами. Именно они, обеспечивая управление импульсными реакциями оптических систем, позволяют реализовать функциональные возможности параллельной обработки больших массивов информации в реальном масштабе времени. Управляемые транспаранты, по аналогии с устройствами, управляющими импульсными реакциями электронных и электрических систем, можно назвать также перестраиваемыми или *динамическими фильтрами.*

Одним из примеров системы, в которой используются оптические динамические фильтры, является ассоциативное запоминающее устройство. Эта система, как и другие когерентные оптические процессы, включает в себя в качестве необходимого элемента оптические квантовые генераторы и рассматривается в последней главе книги.

§ 4.15. Автоматизация проектирования оптических устройств

Особенность оптических устройств как объектов автоматизации проектирования состоит в необходимости рассмотрения и описания сложных дифракционных явлений практически на всех стадиях создания устройства — от формирования замысла конструктора до назначения допусков на параметры. Разработка физических и математических моделей, входящих в математическое обеспечение САПР, оказывается возможной на основе результатов *теории* и *методов резонансной угловой фильтрации.* Эти результаты обеспечивают: единство физических и математических принципов, применяемых при разработке моделей; универсальность модели, заключающуюся в возможности описания различных классов устройств; блочность моделей, обеспечивающую компоновку более сложных моделей и алгоритмов из простых; адаптируемость моделей, связанную с возможностью использования их с учетом различных степеней приближения описания устройств на всех этапах проектирования.

Проектируемое оптическое устройство рассматривается здесь как резонансный частотно-угловой фильтр, в котором определяющими являются физические явления, связанные с преобразованием частотно-углового спектра волновых полей. В результате осуществляется преобразование их пространственно-временно́й формы.

Простейшим базовым элементом или базовым элементом нулевого уровня (БЭО) предлагается считать границу раздела двух диэлектриков. В качестве базо-

вого элемента оптических устройств или базового элемента первого уровня (БЭ1) принят диэлектрический слой, заключенный между двумя полубесконечными средами. При этом базовый элемент первого уровня получают каскадным соединением двух базовых элементов нулевого уровня: $БЭ1_{13} = БЭ0_{12} + БЭ0_{23}$.

В устройствах формирования используют спектральные и импульсные математические модели базовых элементов. В первых из них рассматриваются коэффициенты передачи по проходящей (S_{21}) и отраженной (S_{11}) волнам для установления связи между спектрами этих волн. При этом учитываются также спектры свободных колебаний системы и соседних участков. Импульсные пространственно-временны́е модели строят с помощью обратных преобразований Фурье и Лапласа от S_{11} и S_{21}.

Математические модели сложных структур, соответствующие различным реальным устройствам, получают из моделей базовых элементов путем одно-, двух- и трехмерного наращивания с учетом обратной связи между элементами.

Автоматизированное проектирование включает в себя следующие этапы: анализ структурной схемы устройства и выделение базовых элементов; построение физической модели устройства из базовых элементов различных уровней; создание математической модели по заданному изображению физической модели; формирование структурной схемы анализа (синтеза) устройства с использованием полученной математической модели; описание структурной схемы в терминах выбранного языка программирования; проведение проектных расчетов и принятие решений по реализации устройства [72].

Раздел второй

ЭЛЕКТРОМАГНИТНЫЕ ЯВЛЕНИЯ В АНИЗОТРОПНОЙ ЛИНЕЙНОЙ СРЕДЕ

Глава 5
Электромагнитное поле в немагнитных диэлектриках

Анизотропные вещества обладают различными свойствами в разных направлениях. В электрически анизотропной среде вектор поляризации в общем случае не параллелен вектору напряженности электрического поля. В магнитно-анизотропной среде вектор намагниченности в общем случае не параллелен вектору напряженности магнитного поля. В среде с анизотропными потерями вектор плотности тока не параллелен вектору напряженности электрического поля. Природа перечисленных видов анизотропии различна. Однако общим является то, что соответствующие «материальные» уравнения, связывающие векторы, входящие в уравнения Максвелла, приобретают тензорный характер, определяемый тензорным характером соответствующих проницаемостей и проводимостей. В данной главе рассмотрены свойства диэлектриков, в которых магнитной анизотропией можно пренебречь.

§ 5.1. Тензор диэлектрической проницаемости

Электрическое поле в анизотропном материале встречается в слоистой изоляции, грунте, веществах с волокнистым строением, кристаллах и т. д. В общем случае зависимость смещения и поляризации от напряженности поля в анизотропных диэлектриках имеет сложный характер:

$$\begin{aligned} D_1 &= \varepsilon_{11}\varepsilon_0 E_1 + \varepsilon_{12}\varepsilon_0 E_2 + \varepsilon_{13}\varepsilon_0 E_3; \\ D_2 &= \varepsilon_{21}\varepsilon_0 E_1 + \varepsilon_{22}\varepsilon_0 E_2 + \varepsilon_{23}\varepsilon_0 E_3; \\ D_3 &= \varepsilon_{31}\varepsilon_0 E_1 + \varepsilon_{32}\varepsilon_0 E_2 + \varepsilon_{33}\varepsilon_0 E_3. \end{aligned} \qquad (5.1)$$

В прямоугольной системе координат индексы 1, 2, 3 относятся к координатным осям x, y, z.

Линейные преобразования, определяемые системой линейных уравнений (5.1), могут быть записаны в виде

$$\mathbf{D} = \varepsilon_r \varepsilon_0 \mathbf{E}, \qquad (5.2)$$

если ввести в рассмотрение матрицу преобразования

$$\varepsilon_r = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{bmatrix}, \tag{5.3}$$

называемую *тензором диэлектрической проницаемости.*

Тензор второго ранга (5.3) обозначают иногда ε_{ik}, а вместо (5.2) записывают

$$D_{ik} = \varepsilon_{ik}\varepsilon_0 E_k. \tag{5.4}$$

Когда магнитное поле отсутствует, тензор ε_{ik} симметричен: $\varepsilon_{ik} = \varepsilon_{ki}$. Поэтому максимально возможное количество независимых коэффициентов преобразования для тензора диэлектрической проницаемости может быть снижено в общем случае с девяти до шести.

В листах слюды, асбесте, кристаллических веществах, древесноволокнистых плитах оси анизотропии прямолинейны, в изоляции кабелей, бакелитовых цилиндрах — криволинейны. В большинстве случаев оси анизотропии перпендикулярны друг другу (*ортогональная анизотропия*).

Если оси координат выбрать таким образом, чтобы они совпадали с осями анизотропии (главными диэлектрическими осями), то тензор приобретает диагональную форму

$$\varepsilon_r = \begin{bmatrix} \varepsilon_x & 0 & 0 \\ 0 & \varepsilon_y & 0 \\ 0 & 0 & \varepsilon_z \end{bmatrix}, \tag{5.5}$$

где ε_x, ε_y, ε_z — *главные диэлектрические проницаемости.*

Наглядно представить зависимость диэлектрической проницаемости $\varepsilon_r = D/(\varepsilon_0 E)$ от направления можно, если построить геометрическое место точек конца нормированного вектора смещения $\mathbf{D}/(\varepsilon_0 E)$, определяемого направляющими косинусами $\cos\alpha$, $\cos\beta$ и $\cos\gamma$ в декартовой системе координат, совпадающей с осями анизотропии. В этом случае

$$E^2 = E_x^2 + E_y^2 + E_z^2 = \left(\frac{D_x}{\varepsilon_x\varepsilon_0}\right)^2 + \left(\frac{D_y}{\varepsilon_y\varepsilon_0}\right)^2 + \left(\frac{D_z}{\varepsilon_z\varepsilon_0}\right)^2 =$$

$$= \frac{D^2}{\varepsilon_0^2}\left(\frac{\cos^2\alpha}{\varepsilon_x^2} + \frac{\cos^2\alpha}{\varepsilon_y^2} + \frac{\cos^2\alpha}{\varepsilon_z^2}\right);$$

$$\frac{x_\varepsilon^2}{\varepsilon_x^2} + \frac{y_\varepsilon^2}{\varepsilon_y^2} + \frac{z_\varepsilon^2}{\varepsilon_z^2} = 1, \tag{5.6}$$

где $x_\varepsilon = \varepsilon_r \cos\alpha$; $y_\varepsilon = \varepsilon_r \cos\beta$; $z_\varepsilon = \varepsilon_r \cos\gamma$ — координаты поверхности, которая носит название *эллипсоида диэлектрической проницаемости*. Полуоси этого эллипсоида совпадают с главными диэлектрическими осями.

§ 5.2. Электромагнитная волна в анизотропной среде

Для немагнитной ($\mu_r = 1$) и прозрачной (без потерь) среды поле монохроматической волны описывается уравнениями Максвелла

$$\mathrm{rot}\,\mathbf{H} = j\omega\mathbf{D}; \qquad \mathrm{rot}\,\mathbf{E} = -j\omega\mu_0\mathbf{H}. \tag{5.7}$$

В плоской волне, распространяющейся в среде без потерь, волновое число k вещественно. Поэтому уравнения (1.5а), можно переписать в виде

$$\mu_0\omega\mathbf{H} = \mathbf{k}\times\mathbf{E}, \qquad \omega\mathbf{D} = -\mathbf{k}\times\mathbf{H}, \tag{5.8}$$

определяющем взаимное расположение векторов поля (рис. 5.1). Для такого расположения характерны: взаимная перпендикулярность векторов **k**, **D** и **H** (**B**); копланарность векторов **D**, **E**, **П** и **k**; поперечность векторов **H** и **D** (но не **E**) по отношению к направлению волнового вектора **k**; несовпадение направления потока мощности $\mathbf{П} = \mathbf{E}\times\mathbf{H}$ (луча) с направлением волнового вектора. Вектор **П** составляет с вектором **k** угол δ, равный углу между **E** и **D**.

Показатель преломления, или *вектор рефракции*

$$\mathbf{n} = \frac{c_0}{\omega}\mathbf{k} \tag{5.9}$$

в анизотропных средах связан с законами преломления иначе, чем в изотропных средах. Кроме того, модуль вектора **n** в анизотропной среде зависит от его направления. Согласно уравнениям (5.8) и (5.9) $c_0\mu_0\mathbf{H} = \mathbf{n}\times\mathbf{E}$, $c_0\mathbf{D} = -\mathbf{n}\times\mathbf{H}$ и

$$\mathbf{D}/\varepsilon_0 = \mathbf{n}\times\mathbf{E}\times\mathbf{n} = n^2\mathbf{E} - \mathbf{n}(\mathbf{nE}). \tag{5.10}$$

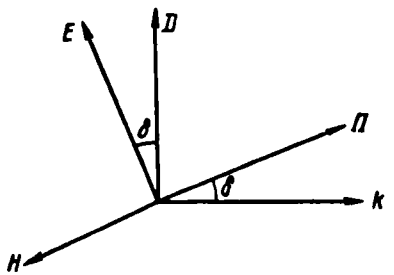


Рис. 5.1. Векторы поля электромагнитной волны в анизотропном диэлектрике

Приравнивая компоненты векторов, стоящих в правых частях выражений (5.4) и (5.10), можно получить три однородных линейных уравнения для трех составляющих вектора **E**. Например, для i-й составляющей $n^2E_i - n_in_kE_k = \varepsilon_{ik}E_k$ или

$$(n^2\delta_{ik} - n_i n_k - \varepsilon_{ik})E_k = 0.$$

Из условия совместности этих уравнений определитель

$$\left|n^2\delta_{ik} - n_i n_k - \varepsilon_{ik}\right| = 0. \qquad (5.11а)$$

Если совместить декартовы координаты с *главными диэлектрическими осями* (главные оси тензора ε_{ik}) и обозначить главные значения тензора ε_x, ε_y и ε_z, то последнее соотношение приводит к *уравнению Френеля*

$$n^2(\varepsilon_x n_x^2 + \varepsilon_y n_y^2 + \varepsilon_z n_z^2) - [n_x^2\varepsilon_x(\varepsilon_y + \varepsilon_z) +$$
$$+ n_y^2\varepsilon_y(\varepsilon_x + \varepsilon_z) + n_z^2\varepsilon_z(\varepsilon_x + \varepsilon_y)] + \varepsilon_x\varepsilon_y\varepsilon_z = 0. \qquad (5.11б)$$

Это уравнение определяет в неявном виде закон дисперсии, т. е. зависимость волнового вектора от частоты.

При заданном направлении волнового вектора выражение (5.11б) представляет собой биквадратное уравнение с вещественными коэффициентами. Таким образом, каждому направлению **n** соответствуют в общем случае два различных абсолютных значения волнового вектора. Можно убедиться, что эти две волны с одинаковым направлением волнового вектора оказываются линейно поляризованными в двух взаимно перпендикулярных плоскостях. Следовательно, условия распространения волн в анизотропных средах существенно отличны от изотропных. Напомним, что в общем случае при распространении в изотропной среде эллиптически поляризованной волны двум направлениям поляризации соответствует один и тот же волновой вектор.

Уравнение Френеля (5.11б) определяет в координатах n_x, n_y и n_z некоторую поверхность, называемую *поверхностью волновых векторов.*

Более наглядно связь между направлением поляризации волны и показателем преломления кристалла передает его *оптическая индикатриса.* Если совместить координатные оси с осями анизотропии, то из выражения для плотности энергии электрического поля

$$2w_e = ED = \frac{D_x^2}{\varepsilon_x\varepsilon_0} + \frac{D_y^2}{\varepsilon_y\varepsilon_0} + \frac{D_z^2}{\varepsilon_z\varepsilon_0}$$

уравнение индикатрисы можно записать в виде

$$\frac{x_n^2}{\varepsilon_x} + \frac{y_n^2}{\varepsilon_y} + \frac{z_n^2}{\varepsilon_z} = 1, \qquad (5.12)$$

где $x_n = D_x/(2w_e\sqrt{\varepsilon_0})$, $y_n = D_y/(2w_e\sqrt{\varepsilon_0})$, $z_n = D_z/(2w_e\sqrt{\varepsilon_0})$ — нормированные проекции вектора смещения. Уравнение (5.12) описывает эллипсоид, который иногда называют *эллипсоидом волновых нормалей* или *эллипсоидом индексов.*

Для нахождения коэффициента преломления с помощью оптической индикатрисы необходимо через начало координат провести плоскость, перпендикулярную направлению распространения волны, и плоскость поляризации, проходящую через векторы **D** и **n**. Отрезок линии пересечения этих двух плоскостей между началом координат и поверхностью эллипсоида равен показателю преломления n для данной поляризации и данного направления волны.

Не приводя громоздкого обоснования этого результата в общем виде, нетрудно убедиться в его справедливости для волн, распространяющихся вдоль координатных осей. Для этого достаточно обратить внимание на то, что вектор смещения всегда перпендикулярен направлению волнового вектора, и что полуоси эллипсоида (5.12) $\sqrt{\varepsilon_x}$, $\sqrt{\varepsilon_y}$, $\sqrt{\varepsilon_z}$ совпадают со значениями показателя преломления волн, поляризованных вдоль соответствующих осей анизотропии x, y, z.

Иногда уравнение индикатрисы (5.12) записывают в виде

$$a_0^2 x_n^2 + b_0^2 y_n^2 + c_0^2 z_n^2 = 1, \qquad (5.13)$$

где $a_0 = \varepsilon_x^{-1/2}$; $b_0 = \varepsilon_y^{-1/2}$; $c_0 = \varepsilon_z^{-1/2}$ — нормированные значения главных фазовых скоростей распространения волн с одноименными поляризациями.

В прямоугольной системе координат, совпадающей с кристаллографическими осями, уравнение индикатрисы (5.13) в общем случае усложняется:

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{23}yz + 2a_{31}zx + 2a_{12}xy = 1, \qquad (5.14)$$

где a_{11}, a_{22}, ..., a_{12} — поляризационные константы, зависящие от взаимного расположения главных осей индикатрисы и кристаллографических осей.

Всякий неравноосный эллипсоид имеет два и только два круговых сечения, проходящих через центр. Нормали к круговым сечениям индикатрисы называют *оптическими осями*. Вдоль этих осей скорость распространения (и показатель преломления) имеет одно значение, не зависящее от поляризации волны. В остальных направлениях распространения упомянутые сечения имеют эллиптическую форму, и скорость волны зависит от поляризации. Разность $n_{max} - n_{min} = B$ называют *двупреломлением кристалла*. Кристаллы с неравноосными индикатрисой, поверхностью волновых векторов и эллипсоидом диэлектрической проницаемости называют оптически двухосными. К ним относят кристаллы с ромбической, моноклинной и триклинной системами.

Кристаллы, в которых можно выбрать два (и более) кристаллографически эквивалентных направления, лежащих в одной плоскости, называют *оптически одноосными*. К ним относят кристаллы тригональной, тетрагональной и гексагональной систем.

Кристаллы кубической системы обладают тремя кристаллографически эквивалентными взаимно ортогональными направлениями, совпадающими с главными диэлектрическими осями. Индикатриса и эллипсоиды имеют шаровую форму. Такие кристаллы изотропны и оптически эквивалентны аморфному телу.

§ 5.3. Обыкновенные и необыкновенные волны

Остановимся подробнее на оптических свойствах одноосных кристаллов. С оптической осью кристалла совпадает одна из главных осей тензора ε_{ik}. С ее направлением обычно совмещается ось z, и соответствующее главное значение ε_{ik} обозначают $\varepsilon_{\parallel}$. Главные значения $\varepsilon_{\perp}$ диэлектрического тензора в плоскости, перпендикулярной оптической оси, совпадают, а направления соответствующих главных осей x и y произвольны.

Если в (5.11б) подставить $\varepsilon_x = \varepsilon_y = \varepsilon_{\perp}$ и $\varepsilon_z = \varepsilon_{\parallel}$, то уравнение Френеля примет вид произведения двух квадратичных множителей

$$(n^2 - \varepsilon_{\perp}^2)[(n_x^2 + n_y^2)\varepsilon_{\perp} + n_z^2\varepsilon_{\parallel} - \varepsilon_{\perp}\varepsilon_{\parallel}] = 0,$$

которое распадается на два уравнения второго порядка:

$$n^2 = \varepsilon_{\perp}; \tag{5.15а}$$

$$\frac{n_x^2 + n_y^2}{\varepsilon_{\parallel}} + \frac{n_z^2}{\varepsilon_{\perp}} = 1. \tag{5.15б}$$

Уравнениям (5.15а) и (5.15б) отвечают две раздельные поверхности волновых векторов — сфера и эллипсоид. Они показывают, что в одноосном кристалле могут распространяться волны двух типов.

По отношению к волнам одного типа — *обыкновенным* — кристалл ведет себя как изотропное тело с показателем преломления $n = \sqrt{\varepsilon_{\perp}}$. Абсолютное значение волнового вектора в этом случае не зависит от направления и равно $\omega n / c_0$.

В волнах второго типа (*необыкновенных*) модуль волнового вектора зависит от угла θ его наклона к оптической оси, так как, согласно (5.15б)

$$\frac{\sin^2\theta}{\varepsilon_{\parallel}} + \frac{\cos^2\theta}{\varepsilon_{\perp}} = \frac{1}{n^2}. \tag{5.16}$$

В отличие от обыкновенной волны направления волнового вектора и вектора групповой скорости $\mathbf{v}_{гр} = \partial\omega/\partial\mathbf{k}$ в необыкновенной волне не совпадают.

§ 5.4. Двойное лучепреломление

На поверхности анизотропной среды законы отражения и преломления (1.45), определяющие направления распространения отраженных и преломленных волн, остаются прежними. Волновые векторы и векторы рефракции (5.9) как отраженных, так и преломленных волн лежат в плоскости падения. Имеет прежний вид и соотношение (1.47) между углами падения и преломления волн.

Принципиальная особенность отражения и преломления волн в анизотропной среде связана с тем, что граничным условиям удовлетворяют в общем случае две пары отраженных и преломленных волн, соответствующие двум решениям уравнения (5.11б) в каждой из сред. В частном случае, когда однородная волна падает из изотропной среды на поверхность одноосного кристалла, образуются две преломленные волны: обыкновенная и необыкновенная. Поскольку показатели преломления у них различны, то в соответствии с (1.47) и направления этих волн будут различными.

Определение направления необыкновенной волны усложняется тем, что показатель преломления (5.16) зависит от этого же направления. Для геометрического способа решения этой задачи необходимо направления координат поверхности волновых векторов привести в соответствие с расположением поверхности анизотропной среды, из начала координат провести вектор рефракции падающей волны, и через конец этого вектора опустить перпендикуляр (направление $\mathbf{e}_q$) к границе раздела. Точки пересечения этого перпендикуляра с двумя поверхностями волновых векторов (или векторов рефракции) определяют направление и длину искомых двух векторов рефракции преломленных волн, как это следует из (1.45) и (5.9).

Таким образом, в общем случае, каждая подающая волна в анизотропной среде вызывает две преломленные волны с разными направлениями и скоростями распространения. Это явление получило название *двойного лучепреломления.*

Наблюдаемое направление распространения лучей задается не волновым, а лучевым вектором $\mathbf{s}$; его направление совпадает с групповой скоростью $\partial\omega/\partial\mathbf{k}$ и вектором Пойнтинга, а значение определяется равенством $\mathbf{n}\cdot\mathbf{s}=1$. В отличие от обыкновенной волны аналогичной обычным преломленным волнам в изотропных средах с лучевым вектором, совпадающим по направлению с волновым вектором, направление лучевого вектора необыкновенной волны не совпадает в общем случае ни с волновым вектором, ни с плоскостью падения.

Двойное лучепреломление впервые наблюдалось в исландском шпате, представляющем собой кристаллическую разновидность углекислого

кальция $CaCO_3$. Для электромагнитных колебаний с длиной волны 569,3 нм в кальците $n_{\parallel} = 1{,}468$ и $n_{\perp} = 1{,}658$.

Двойное лучепреломление можно обнаружить и при нормальном падении электромагнитной волны на поверхность кристалла, когда угол падения равен нулю. При этом один из двух лучей служит продолжением первичного пучка, а у другого луча угол преломления больше нуля. Это отступление от законов преломления в изотропных средах и послужило причиной того, что второй пучок назвали необыкновенным лучом, а первый — обыкновенным.

Обыкновенный и необыкновенный лучи на выходе из кристалла линейно поляризованы в двух взаимно перпендикулярных плоскостях. Различие направлений поляризации лучей в одноосном кристалле лежит в основе работы многих поляризаторов. Так называют комбинации простых кристаллов, служащие для выделения из луча света с произвольной поляризацией линейно поляризованного луча.

§ 5.5. Квадратичный электрооптический эффект

Электромагнитная анизотропия может появиться в изотропном теле (с диэлектрическим тензором $\varepsilon_{ik} = \varepsilon_r \delta_{ik}$) под влиянием механических воздействий, электрического и магнитного полей как результат изменения тензора диэлектрической проницаемости. Независимо от степени изменения диэлектрической проницаемости (нередко весьма слабого) это приводит к качественному изменению электромагнитных, в частности оптических, свойств вещества.

Изменение оптических свойств вещества под действием электрического поля носит название *электрооптического эффекта*. Под действием электрического поля изотропная среда может стать анизотропной, а одноосный кристалл — двухосным. В общем случае под действием электрического поля происходят деформация и поворот оптической индикатрисы: изменяются как размеры ее полуосей, так и их направление. После этого главные оси индикатрисы x', y', z' уже не будут совпадать с кристаллографическими осями x, y, z, а поляризационные константы a_{ik} в уравнении (5.14) будут зависеть от приложенного к кристаллу электрического поля.

Если обозначить через a, b, c нормированные значения главных скоростей волн новой индикатрисы, то новые значения поляризационных констант a_{ik} в кристаллографической системе координат можно найти по правилам преобразования координат:

$$
\begin{aligned}
a_{11} &= a^2\alpha_1^2 + b^2\beta_1^2 + c^2\gamma_1^2;\\
a_{22} &= a^2\alpha_2^2 + b^2\beta_2^2 + c^2\gamma_2^2;\\
a_{33} &= a^2\alpha_3^2 + b^2\beta_3^2 + c^2\gamma_3^2;\\
a_{12} &= a^2\alpha_1\alpha_2 + b^2\beta_1\beta_2 + c^2\gamma_1\gamma_2;\\
a_{23} &= a^2\alpha_2\alpha_3 + b^2\beta_2\beta_3 + c^2\gamma_2\gamma_3;\\
a_{31} &= a^2\alpha_3\alpha_1 + b^2\beta_3\beta_1 + c^2\gamma_3\gamma_1,
\end{aligned}
\qquad (5.17)
$$

где α, β, γ — косинусы углов, образованных главными осями новой оптической индикатрисы (x', y', z') с кристаллографическими осями (x, y, z). Например, $\alpha_1 = \cos(\hat{\mathbf{e}_{x'}\mathbf{e}_x})$, $\alpha_2 = \cos(\hat{\mathbf{e}_{x'}\mathbf{e}_y})$ и $\beta_1 = \cos(\hat{\mathbf{e}_{y'}\mathbf{e}_x})$.

Количественную оценку электрооптического эффекта получают из соотношений, выражающих зависимость поляризационных констант от напряженности внешнего электрического поля E_0. Электрооптические эффекты могут быть линейные или квадратичные в зависимости от того, какой степени напряженности E_0 пропорциональны изменения оптических параметров. Примером квадратичного эффекта может служить *эффект Керра.*

Он проявляется в том, что под действием электрического поля в некоторых жидкостях возникает двойное лучепреломление. При этом изотропное тело в электрическом поле приобретает оптические свойства одноосного кристалла.

Главное значение диэлектрического тензора в направлении постоянного поля E_0 $\varepsilon_{\parallel} = \varepsilon_{\perp} + f(E_0)$. Положение двух других ортогональных главных осей произвольно, соответствующие им два главных значения одинаковы и равны $\varepsilon_{\perp}$.

Если $n_{\parallel}$ — показатель преломления света с длиной волны λ при расположении электрического вектора волны параллельно вектору E_0, а $n_{\perp}$ — то же, при ортогональном расположении электрических векторов волны и постоянного поля, то, как было обнаружено, сдвиг фаз между этими компонентами волны, выраженный в долях длины волны,

$$
\Gamma = \frac{n_{\parallel} - n_{\perp}}{\lambda} l = KE_0^2 l, \qquad (5.18)
$$

где K — постоянная Керра.

Сдвиг фаз между необыкновенным и обыкновенным лучами приводит к тому, что после выхода из ячейки Керра плоскость поляризации результирующей волны оказывается повернутой на тем больший угол, чем больше разность коэффициентов преломления $n_{\parallel}$ и $n_{\perp}$ и чем больше путь l луча в ячейке.

Совокупность расположенных вдоль луча поляризатора ячейки Керра и анализатора называют *электрооптическим затвором* или *затвором Керра*. Если плоскости пропускания поляризатора и анализатора повернуты друг относительно друга на 90°, то затвор прозрачен только при приложении к нему напряжения, называемого полуволновым. При этом напряжении угол поворота плоскости поляризации ячейкой составляет 90°. Если к электродам ячейки Керра приложено переменное напряжение, то яркость выходящего луча будет меняться, и все устройство может быть названо *электрооптическим модулятором*.

Модуляционную характеристику целесообразно найти сначала для общего случая, когда плоскость поляризации света на входе ячейки Керра составляет произвольный угол α с горизонтальной осью y, и световой луч распространяется также горизонтально вдоль оси x, т. е. перед входом в ячейку ($x = 0$) напряженность электрического поля световой волны $E(0,t) = E_m \sin \omega t$. Под воздействием внешнего поля E_0, направленного вертикально (вдоль оси z), ячейка приобретает свойства одноосного кристалла. При этом вертикальная компонента световой волны $E_e(0,t) = E_m \sin\alpha \sin \omega t$ проходит ячейку как необыкновенная волна, а горизонтальная компонента $E_o(0,t) = E_m \cos\alpha \sin \omega t$ — как обыкновенная. Если длина ячейки Керра l, то на выходе из нее $E_e(l,t) = E_m \sin\alpha \sin(\omega t - \varphi_e)$; $E_o(l,t) = E_m \cos\alpha \sin(\omega t - \varphi_o)$, где φ_e, φ_o — сдвиг фаз необыкновенной и обыкновенной волн относительно фазы волны на входе в ячейку.

Если ось наибольшего пропускания анализатора повернута на 90° относительно оси входного поляризатора, то анализатор пропустит только те компоненты обыкновенной и необыкновенной волн, которые параллельны оси наибольшего пропускания анализатора:

$$E_e(L,t) = a \sin(\omega t - \varphi'_e);$$
$$E_o(L,t) = a \sin(\omega t - \varphi'_o),$$

где $a = E_m \sin\alpha \cos\alpha$; $\varphi'_e = \varphi_e + \varphi_a$; $\varphi'_o = \varphi_o + \varphi_a$, φ_a — сдвиг фазы на участке L — l. На выходе из модулятора результирующая световая волна имеет напряженность

$$E(L,t) = E_e(L,t) - E_o(L,t) = a[\sin(\omega t - \varphi'_e) - \sin(\omega t - \varphi'_o)] =$$
$$= a(\sin\omega t \cos\varphi'_e - \cos\omega t \sin\varphi'_e - \sin\omega t \cos\varphi'_o + \cos\omega t \sin\varphi'_o) = A \sin(\omega t - \varphi),$$

причем $A \sin\varphi = a(\cos\varphi'_e - \cos\varphi'_o)$; $A\cos\varphi = a(\sin\varphi'_e - \sin\varphi'_o)$.

Тогда $A^2/a^2 = 2 - 2\cos(\varphi'_e - \varphi'_o) = 2(1 - \cos\varphi) = 4\sin^2\varphi/2$, где $\varphi = \varphi'_e - \varphi'_o = \varphi_e - \varphi_o = 2\pi(n_e - n_o)l/\lambda = 2\pi\Gamma$.

Отношение интенсивностей световых лучей на выходе из модулятора (I) и на входе в ячейку Керра (I_0) $I/I_0 = A^2/E_m^2 = \sin^2 2\alpha \sin^2(\varphi/2)$.

Таким образом, потери в модуляторе будут наименьшими при $\alpha = 45°$. В этом случае модуляционная характеристика имеет вид

$$I = I_0 \sin^2 \pi\Gamma \tag{5.19}$$

или, если учесть (5.18),

$$I = I_0 \sin^2\left(\pi K \frac{l}{d^2} U^2\right), \tag{5.20}$$

где $U = E_0 d$ — напряжение на электродах ячейки; d — расстояние между ними.

Напряжение, необходимое для перехода от затемненного состояния ($I = 0$) к просветленному ($I = I_0$), получило название *полуволнового*, так как оно соответствует разности фаз обыкновенного и необыкновенного лучей $\varphi = \varphi_e - \varphi_0 = 180°$ и $\Gamma = 0{,}5$.

Наибольшее применение в ячейках Керра получил нитробензол $C_6H_5NO_2$, для которого постоянная Керра составляет 400 см/В^2, относительная диэлектрическая проницаемость 36, пробивная напряженность 150 кВ/см и время молекулярной релаксации 10^{-9} с.

§ 5.6. Линейный электрооптический эффект

В кристаллах при сравнительно малых интенсивностях света квадратичные эффекты пренебрежимо малы. Электрооптический эффект в кристаллах иногда называют *эффектом Поккельса*. Это линейный электрооптический эффект, так как для него изменения поляризационных констант (5.17)

$$\begin{aligned}
a_{11} - a_0^2 &= r_{11}E_x + r_{12}E_y + r_{13}E_z;\\
a_{22} - b_0^2 &= r_{21}E_x + r_{22}E_y + r_{23}E_z;\\
a_{33} - c_0^2 &= r_{31}E_x + r_{32}E_y + r_{33}E_z;\\
a_{23} &= r_{41}E_x + r_{42}E_y + r_{43}E_z;\\
a_{31} &= r_{51}E_x + r_{52}E_y + r_{53}E_z;\\
a_{32} &= r_{61}E_x + r_{62}E_y + r_{63}E_z,
\end{aligned} \tag{5.21}$$

где E_x, E_y, E_z — проекции вектора напряженности внешнего электрического поля на кристаллографические оси; r_{ik} — электрооптические коэффициенты.

Количественной оценкой электрооптического эффекта в кристаллах служит матрица

$$\mathbf{r} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ r_{41} & r_{42} & r_{43} \\ r_{51} & r_{52} & r_{53} \\ r_{61} & r_{62} & r_{63} \end{bmatrix}, \tag{5.22}$$

в которую входят в общем случае 18 независимых коэффициентов.

Линейный электрооптический эффект возникает у кристаллов тех классов, в которых наблюдается пьезоэлектричество. К их числу относятся кристаллы ниобата $LiNbO_3$ и танталата $LiTaO_3$ лития, а также дигидрофосфата калия KH_2PO_4 (KDP). Применяют иногда и дигидрофосфат аммония $NH_4H_2PO_4$ (ADP). Кристаллы KDP и ADP прозрачны и имеют $r_{63} \approx (8{,}5...10{,}5)\cdot 10^{-10}$ см/В. По годности к обработке, хрупкости, гигроскопичности и растворимости они сравнимы с кристаллами поваренной соли.

Симметрия кристаллов существенно ограничивает количество независимых коэффициентов матрицы (5.22),отличных от нуля. Например, в кристаллах дигидрофосфатов их остается только два:

$$\mathbf{r} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ r_{41} & 0 & 0 \\ 0 & r_{41} & 0 \\ 0 & 0 & r_{63} \end{bmatrix}.$$

Это упрощает уравнения (5.21) для поляризационных констант:

$$\begin{aligned} a_{11} &= a_0^2; \quad a_{23} = r_{41}E_x; \\ a_{22} &= b_0^2; \quad a_{31} = r_{41}E_y; \\ a_{33} &= c_0^2; \quad a_{12} = r_{63}E_z. \end{aligned} \tag{5.23}$$

Если же учесть, что кристаллы дигидрофосфатов в отсутствие электрического поля являются одноосными ($a_0^2 = b_0^2 = o^2$, $c_0^2 = e^2$), то

$$a_{11} = a_{22} = o^2, \quad a_{33} = e^2. \tag{5.24}$$

Окончательное уравнение оптической индикатрисы для кристаллов со структурой дигидрофосфатов может быть получено из уравнения (5.14) с учетом обозначений в равенствах (5.23) и (5.24):

$$o^2(x^2+y^2)+e^2z^2+2r_{41}(E_xzy+E_yzx)+2r_{63}E_zxy=1. \qquad (5.25)$$

Наибольшее распространение на практике получили *основные* кристаллические образцы, у которых поверхность пластины срезается перпендикулярно оптической оси (в отсутствие электрического поля). Электроды обычно прикладывают к этим поверхностям пластины, а ось z совмещают с оптической осью. Тогда вектор напряженности электрического поля имеет только одну компоненту ($E_z=E_0$; $E_x=E_y=0$), и уравнение индикатрисы (5.25) приобретает вид

$$o^2(x^2+y^2)+e^2z^2+2r_{63}E_zxy=1. \qquad (5.26)$$

Из него следует, что ось z' новой оптической индикатрисы совпадает с осью z индикатрисы в отсутствие электрического поля. В этом случае $\gamma_3=1$, $\gamma_1=\gamma_2=\beta_3=\alpha_3=0$.

Чтобы найти остальные косинусы, уравнение (5.26) необходимо привести к координатным осям, совпадающим с осями новой оптической индикатрисы

$$x'=\alpha_1x+\alpha_2y,\quad y'=\beta_1x+\beta_2y=-\alpha_2x+\alpha_1y_1,$$

так как поворот координат x' и y' происходит в плоскости xy ($\alpha_1=\beta_2$, $\alpha_2=-\beta_1$). Уравнение

$$[o^2(\alpha_1^2+\alpha_2^2)+2\alpha_1\alpha_2r_{63}E_z]x'^2+[o^2(\alpha_1^2+\alpha_2^2)-2\alpha_1\alpha_2r_{63}E_z]y'^2+$$
$$+e^2z^2(\alpha_1^2+\alpha_2^2)+2r_{63}E_z(\alpha_1^2-\alpha_2^2)x'y'=(\alpha_1^2+\alpha_2^2)^2$$

должно иметь канонический вид, в котором отсутствуют члены, содержащие произведения координат, т. е. $2r_{63}E_z(\alpha_1^2-\alpha_2^2)x'y'=0$ и $\alpha_1=\alpha_2=\pm\sqrt{2}/2$. Таким образом, независимо от напряженности $E_0>0$ оси x' и y' новой оптической индикатрисы оказываются повернутыми на 45° относительно кристаллографических осей x и y.

Обратные показатели преломления новой индикатрисы

$$a^2=o^2(\alpha_1^2+\alpha_2^2)+2\alpha_1\alpha_2r_{63}E_z=o^2+r_{63}E_0;$$
$$b^2=o^2(\alpha_1^2+\alpha_2^2)-2\alpha_1\alpha_2r_{63}E_z=o^2-r_{63}E_0; \qquad (5.27)$$
$$c^2=e^2.$$

Таким образом, во внешнем электрическом поле оптическая индикатриса приобретает форму неравноосного эллипсоида ($a\neq b\neq c$). Это

означает, что одним из следствий электрооптического эффекта является превращение одноосного кристалла в двухосный.

Новые показатели преломления можно определить, исходя из значений (5.27):

$$n_{x'} = 1/a = (o^2 + r_{63}E_0)^{-1/2} = n_o(1 + n_o^2 r_{63}E_0)^{-1/2} \approx$$
$$\approx n_o - 0{,}5n_o^3 r_{63}E_0 = n_o - \Delta n;$$
$$n_{y'} = 1/b = (o^2 - r_{63}E_0)^{-1/2} = n_o + \frac{1}{2}n_o^3 r_{63}E_0 = n_o + \Delta n; \qquad (5.28)$$
$$n_{z'} = 1/c = 1/e = n_e,$$

если учесть, что $n_o^2 r_{63}E_0 \ll 1$.

Двупреломление для волн, распространяющихся вдоль главных осей новой индикатрисы,

$$B_{x'} = n_{y'} - n_{z'} = n_o - n_e + 0{,}5n_o^3 r_{63}E_0 = B_0 + \Delta n;$$
$$B_{y'} = n_{x'} - n_{z'} = n_o - n_e - 0{,}5n_o^3 r_{63}E_0 = B_0 - \Delta n;$$
$$B_{z'} = n_{y'} - n_{x'} = n_o^3 r_{63}E_0 = 2\Delta n.$$

Следовательно, изменение двупреломления кристаллов под действием электрического поля ($\Delta n = 0{,}5n_o^3 r_{63}E_0$) будет иметь максимальное значение ($2\Delta n$) для луча, распространяющегося вдоль оси z. В других направлениях электрооптический эффект усложняется естественным двупреломлением, достигающим максимального значения $B_0 = n_o - n_e$ в направлениях осей x' и y'.

Анализ электрооптического эффекта под действием внешнего поля, приложенного вдоль осей x и y, показывает, что в этих случаях эффект менее выражен, чем при электрическом поле, направленном вдоль естественной оптической оси z.

Кристаллы ниобата и танталата лития также являются одноосными ($a_0 = b_0$). При температурах ниже точки Кюри (T_{Nb} = 1470 К и T_{Ta} = 890 К) у них отличны от нуля следующие компоненты тензора (5.22): $r_{13} = r_{23}$, r_{33}, $r_{42} = r_{51}$, $r_{22} = -r_{12} = -r_{61}$. Для этого класса кристаллов уравнение оптической индикатрисы в общем случае имеет вид

$$(a_0^2 - r_{22}E_y + r_{13}E_z)x^2 + (a_0^2 + r_{22}E_y + r_{13}E_z)y^2 + (c_0^2 + r_{33}E_z)z^2 +$$
$$+2r_{42}E_y yz + 2r_{42}E_x xz - 2r_{42}E_x xy = 1.$$

Если луч и внешнее поле направлены вдоль оптической оси ($E_0 = E_z$), то уравнение индикатрисы упрощается:

$$(a_0^2 + r_{13}E_0)(x^2 + y^2) + (c_0^2 + r_{33}E_0)z^2 = 1$$

или

$$(x^2 + y^2)/(n_o + \Delta n_o)^2 + z^2/(n_e + \Delta n_e)^2 = 1,$$

причем $\Delta n_o = -n_o^3 r_{13} E_0/2$ и $\Delta n_e = -n_e^3 r_{33} E_0/2$.

При распространении света вдоль оптической оси относительная фазовая задержка $\Gamma_1 = n_o^3 r_{13} E_0 l/(2\lambda)$. При распространении света перпендикулярно оптической оси с поляризацией, перпендикулярной оси, фазовая задержка оказывается такой же, как и в предыдущем продольном электрооптическом эффекте. При поперечном эффекте с поляризацией в направлении управляющего поля $\Gamma_2 = n_e^3 r_{33} F_0 l/(2\lambda)$. Для ниобата лития $r_{33} = 30{,}8$ пм/В, $r_{13} = 8{,}6$ пм/В, $n_o = 2{,}29$, $n_e = 2{,}20$. Коэффициент r_{33} для ниобата и танталата лития в 3—4 раза больше, чем коэффициент r_{63} для KDP и ADP. Поэтому ниобат и танталат лития оказываются более эффективными (полуволновые напряжения у них соответственно равны 4,45 и 3,92 кВ). Кроме того, ниобат и танталат лития, как и другие сегнетоэлектрические перовскиты, являются достаточно механически прочными и негигроскопическими. В кристалле этой группы $Sr_{0,75}Ba_{0,25}Nb_2O_3$ полуволновое напряжение равно 37 В, однако температура Кюри значительно ниже (333 К).

§ 5.7. Электрооптические явления в жидких кристаллах

Состояние вещества, внутренняя структура которого имеет черты, характерные как для жидкости, так и для твердого тела, называют *мезаморфным*. Вещество в этом состоянии называют *жидким кристаллом* или *анизотропной жидкостью*, поскольку оно обладает оптической анизотропией. В основном это органические вещества с молекулами вытянутой формы.

Преимущественное направление ориентации длинных осей молекул характеризуют единичным вектором, называемым *директором*.

Жидкие кристаллы можно разделить на три группы, отличающиеся своими физическими и прежде всего оптическими свойствами: нематические, смектические, холестерические.

Нематические жидкие кристаллы, или *нематики*, обладают нитевидной структурой (нема — по-гречески «нить»): длинные оси молекул лежат вдоль линий, параллельных определенному направлению (рис. 5.2, *а*), а их центры размещены хаотично и отдельных слоев не образуют.

Смектические жидкие кристаллы, или *смектики*, характеризуются наличием слоев одинаковой толщины, близкой к длине молекул. Эти

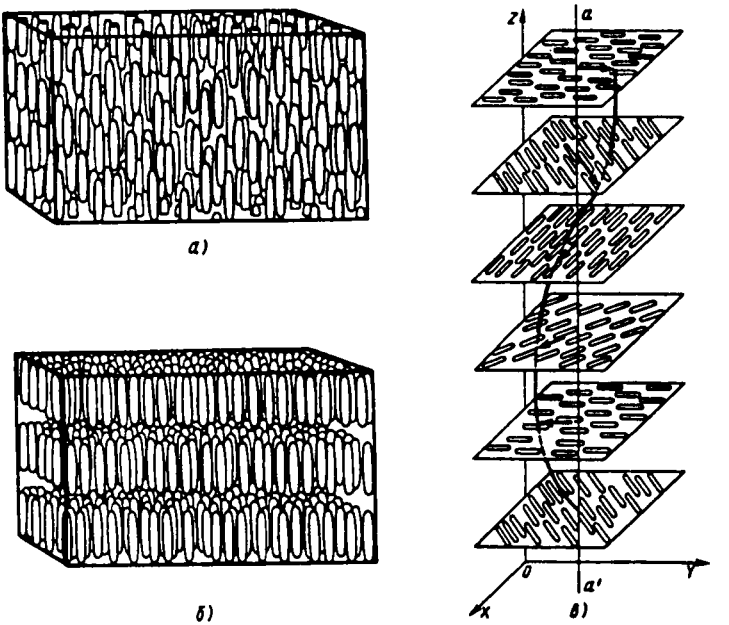


Рис. 5.2. Ориентация молекул в нематических (*а*), смектических (*б*) и холестерических (*в*) жидких кристаллах

слои лежат один над другим на одинаковом расстоянии (рис. 5.2, *б*) и легко перемещаются параллельно друг другу. Примером смектического двойного слоя является пленка мыльного пузыря (смегма — по-гречески «мыло»).

Холестерические жидкие кристаллы, или *холестерики*, образуются главным образом производными холестерина. В этих жидких кристаллах длинные оси молекул параллельны плоскостям мономолекулярных слоев, причем направление ориентации длинных осей молекул в каждом последующем слое отклонено приблизительно на 15′ по сравнению с предыдущим слоем. В результате суммирования отклонений по всем слоям образуется типичная для холестерика спиральная молекулярная структура (рис. 5.2, *в*).

В оптических ячейках жидкокристаллический слой располагается между прозрачными электродами, метод обработки которых существенно влияет на начальную ориентацию молекул. Ионным травлением и вакуумным напылением тонких пленок можно добиться *гомогенной* (плоской или планарной) или *гомеотропной* (нормальной) структуры в зависимости от того, как располагаются молекулы: параллельно или нормально к поверхности электродов.

Ориентация молекул жидких кристаллов вызывает анизотропию электрических и магнитных свойств и, как следствие, двупреломление и вращение плоскости поляризации. В холестериках угол вращения плоскости поляризации света достигает 18°на 1 мкм пути.

Ориентирующее действие электрического поля на молекулы жидких кристаллов связано с их электрическим моментом. Поэтому во внешнем электрическом поле каждая молекула будет стремиться к такому положению, чтобы вектор электрического момента стал параллельным вектору напряженности поля. Энергия взаимодействия электрического момента молекулы изотропной жидкости с внешним полем в несколько десятков тысяч раз меньше энергии теплового движения молекул. В этом случае упорядоченное ориентирование в электрическом поле происходит только частично и заметно в том случае, когда собственные электрические моменты молекул и напряженность поля велики, как это имеет место при эффекте Керра в нитробензоле.

В жидких кристаллах ориентация молекул силами межмолекулярного взаимодействия приводит к тому, что электрическое поле воздействует на ориентированные группы, состоящие из сотен тысяч молекул. Вследствие этого энергия взаимодействия с внешним электрическим полем возрастает в несколько тысяч раз и превышает энергию теплового движения. Энергия деформации структуры жидких кристаллов мала и по сравнению с энергией естественной анизотропии твердых кристаллов. Поэтому для изменения молекулярной структуры жидких кристаллов достаточны сравнительно слабые электрические поля, а также незначительные температурные колебания и механические воздействия. К заметному изменению ориентации молекул и направления оптической оси приводит приложение к тонкому жидкокристаллическому слою даже небольшого напряжения (несколько вольт). Пороговое напряжение материала определяется упругими силами, возвращающими молекулы к первоначальному положению.

Ориентация молекул в электрическом поле зависит от знака диэлектрической анизотропии кристалла $\Delta\varepsilon = \varepsilon_{\parallel} - \varepsilon_{\perp}$, где $\varepsilon_{\parallel}$ — диэлектрическая проницаемость при расположении оси молекулы вдоль направления напряженности электрического поля; $\varepsilon_{\perp}$ — то же, при перпендикулярной ориентации поля. Если направление собственного электрического момента перпендикулярно длинной оси молекулы, то диэлектрическая анизотропия отрицательна. В электрическом поле молекулы ориентируются перпендикулярно полю. Если молекулы обладают собственным электрическим моментом в направлении длинной оси, то анизотропия положительна, и молекулы стремятся расположиться вдоль направления напряженности электрического поля.

Ориентирующее действие электрического поля проявляется как без участия электропроводности (полевые эффекты), так и с ее участием (токовые эффекты).

Небольшие внешние электрические поля приводят к значительным изменениям ориентации директора, а, следовательно, к изменению оптических свойств жидких кристаллов и электрооптическим эффектам, не имеющим аналогов в твердых и жидких телах.

Рассмотрим электрооптические свойства нематиков для случая отрицательной анизотропии ($\Delta\varepsilon = \varepsilon_{\parallel} - \varepsilon_{\perp} < 0$) с гомеотропной структурой. В небольших (но больше пороговых) полях директор переориентируется и станет перпендикулярным внешнему полю, т. е. гомеотропная структура переходит в планарную с соответствующей переориентацией оптической индикатрисы.

Электрооптические эффекты, обусловленные такой переориентацией, называют *ориентационными*. Они бывают нескольких видов.

В рассмотренном примере переориентация молекул возникает в результате продольного изгиба. Связанный с этим ориентационный эффект называют *B-эффектом* (bend — продольный изгиб). Если переориентация возникает в результате поперечного изгиба, то соответствующий ориентационный эффект называют *S-эффектом* (splay — поперечный изгиб). Если переориентация сопровождается превращением закрученной структуры с осью спирали вдоль электрического поля в гомеотропную, то возникающий ориентационный эффект называют *T-эффектом* (twist — кручение).

Интересен *S*-эффект, возникающий при растворении в нематике («хозяин») плеохроичного красителя («гость»), молекулы которого обладают дихроизмом поглощения. При ориентации молекул красителя одновременно с переориентацией нематика электрическим полем спектр поглощения образца (цвет) изменяется. Это явление называют *эффектом гость — хозяин*.

Неоднородное распределение директора вызывает *флексоэлектрический эффект*. К одному из его проявлений относится формирование периодической структуры в виде полосовых доменов. Пороговое напряжение и период доменной структуры определяются анизотропией диэлектрической проницаемости, коэффициентами упругости и разностью флексомодулей, характеризующих взаимодействие вектора поляризации с внешним электрическим полем. Обратно пропорциональная зависимость периода структуры от напряжения на жидкокристаллическом слое позволяет управлять шагом формируемой дифракционной решетки.

В сильных электрических полях в нематиках с анизотропной электропроводностью возникает гидродинамическая неустойчивость, проявляющаяся в возникновении вихревых нитей тока, поворачивающих моле-

кулы нематика. Вихревые трубки воздействуют на свет как фазовая решетка, дающая в некоторых случаях дифракционную картину, образованную системой собирающих и рассеивающих линз. Подобные электрооптические эффекты называют *электрогидродинамическими*.

При еще больших электрических полях возникает интенсивное турбулентное движение нематика и сильное рассеяние света, характерное для *динамического рассеяния*.

Электрооптические эффекты в холестериках можно разбить на три группы.

Первая группа эффектов обусловлена раскручиванием спирали холестерика, она аналогична ориентационным эффектам в нематиках. Если электрическое поле перпендикулярно оси спирали холестерика с $\Delta\varepsilon > 0$, то раскручивание спирали приводит к *красному смещению* — происходит смещение длины волны селективного отражения в сторону больших длин волн. При этом зависимость шага спирали от напряженности поля совершенно такая же, как и длины волны селективного отражения. Если электрическое поле приложено вдоль оси спирали, то деформация спирали приводит к *синему смещению*, при котором происходит уменьшение шага спирали, выход молекулы из холестерической плоскости и смещение длины селективного отражения в сторону меньших длин волн.

Вторую группу электрооптических эффектов составляют переходы, связанные с изменением текстуры в электрическом поле. Например, в холестерике с $\Delta\varepsilon > 0$ в поле, перпендикулярном опорным поверхностям, конфокальная текстура может перейти в текстуру «отпечатков пальцев». Конфокальная текстура получила название от конической поверхности, основанием которой является эллипс, а вершина лежит на гиперболе, проходящей через один из фокусов эллипса, плоскость которого перпендикулярна плоскости гиперболы. Текстура отпечатка пальцев возникает у конфокальной текстуры с косым наклоном спиралей относительно опорных стекол у холестерика с достаточно большим шагом спирали. При данном переходе текстур образец изменяет свое состояние от сильно рассеивающего до прозрачного. Это изменение сопровождается значительным полевым и временным гистерезисом в поле, перпендикулярном опорным поверхностям, планарная структура может претерпеть пространственно периодическую деформацию, а при дальнейшем увеличении поля — перейти в текстуру «отпечатков пальцев».

Для холестериков с $\Delta\varepsilon < 0$ переход конфокальной структуры в планарную проявляется в переходе образца из рассеивающего состояния в прозрачное.

К третьей группе электрооптических эффектов относится *динамическое рассеяние с памятью*, наблюдаемое в смесях холестериков и нематиков с $\Delta\varepsilon < 0$ и $\Delta\sigma > 0$. Если такой холестерик первоначально имел

планарную структуру, то после приложения напряжения выше порогового возникает турбулентное движение жидкости, приводящее к сильному рассеянию света, причем рассеяние сохраняется и после снятия электрического напряжения, когда устанавливается конфокальная текстура. Время памяти может быть доведено до нескольких дней. Для стирания используется переменное напряжение, возвращающее текстуру в первоначальное состояние.

Электрооптические эффекты в смектиках наблюдаются двух видов.

К первому относят *переход* в смектике с $\Delta\varepsilon > 0$ в электрическом поле, перпендикулярном опорным поверхностям, *планарной текстуры в гомеотропную*. В процессе перехода возникает промежуточная текстура с пристеночными дислокациями и дисклинизациями, приводящими к сильному рассеянию света. Если при достижении этой структуры отключить напряжение, она останется стабильной, а конечная гомеотропная текстура — метастабильной (после выключения напряжения она самопроизвольно релаксирует в промежуточную рассеивающую текстуру).

Ко второму виду можно отнести обратный эффект, связанный с *переходом гомеотропной текстуры* смектика с $\Delta\varepsilon < 0$ *в планарную*.

Рассмотренные электрооптические эффекты наблюдаются у смектиков с небольшой электрической проводимостью (меньше 10^{-11} См·см$^{-1}$). При большой электрической проводимости ($10^{-9}...10^{-10}$ См·см$^{-1}$) смектика с $\Delta\varepsilon > 0$ процесс переключения сопровождается электрогидродинамической неустойчивостью. После отключения напряжения образуется рассеивающая текстура, которую можно вернуть в первоначальную гомеотропную переменным напряжением. В смектиках с $\Delta\varepsilon < 0$ и планарной начальной ориентацией электрооптический эффект также связан с электрогидродинамической неустойчивостью и приводит к уменьшению пропускания света. При планарной начальной структуре электрогидродинамическая неустойчивость приводит к динамическому рассеянию.

Жидкие кристаллы успешно выдерживают конкуренцию с другими материалами (например, электрооптическими керамическими и магнитооптическими пленками) для пространственно-временны́х модуляторов света благодаря высокой чувствительности к управляющим сигналам, хорошим оптическим и пороговым свойствам, технологичности изготовления многоэлементных устройств большого формата. Срок службы жидких кристаллов доведен до $5 \cdot 10^4$ ч. В тонких слоях время переключения может быть снижено до 10 мкс.

§ 5.8. Электрооптические модуляторы света

Электрооптические эффекты применяют в модуляторах света различного типа. Изменения коэффициентов преломления и скорости распространения электромагнитной волны под воздействием внешнего напряжения позволяют легко осуществить фазовую модуляцию света. Для амплитудной модуляции света используют различие фазовых скоростей световых лучей, поляризованных вдоль осей x' и y'.

Один из вариантов *модулятора с продольным полем* показан на рис. 5.3. Основным его элементом является электрооптический кристалл с электродами *4*, которые имеют отверстия для пропускания света. Кристалл расположен между двумя скрещенными поляризаторами *1* и *3*. Луч *2* через отверстия в электродах *4* проходит через кристалл *5* вдоль кристаллографической оси *z*.

Если напряжение на электродах отсутствует, то луч после поляризатора *1* в кристалле не изменяет своей поляризации, и поляризатор *3* не пропускает световой поток.

Во внешнем электрическом поле оси оптической индикатрисы x' и y' расположены под углом 45° относительно кристаллографических осей x и y. При показанном штриховкой расположении поляризаторов световой луч на входе кристалла окажется поляризованным в направлении кристаллографической оси y. Линейно поляризованная волна A может быть разложена на две компоненты $A_{x'} = A_{y'} = A/\sqrt{2}$, поляризованные вдоль осей индикатрисы x' и y'. Показатели преломления (5.28) для этих компонент различны, поэтому различны и фазовые скорости. Когда на выходе кристаллов между ними образуется сдвиг фаз 90° результирующая волна будет иметь круговую поляризацию и частично пройдет через поляризатор *3*. Когда же сдвиг фаз составит 180° (при полуволновом напряжении), суммарная волна на выходе кристалла снова станет линейно поляризованной с направлением поляризации, перпендикулярным исходному (на входе кристалла). В этом случае свет полностью пройдет через поляризатор.

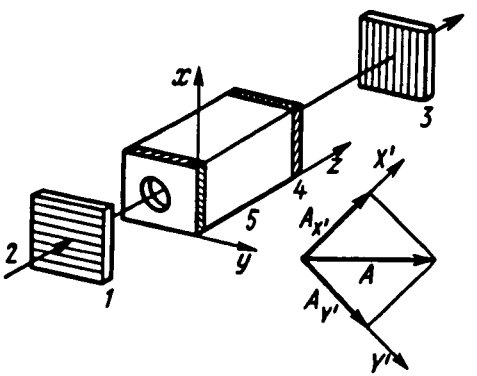


Рис. 5.3. Электрооптический модулятор с продольным полем

Сдвиг фаз зависит от разности коэффициентов преломления

$$\Gamma = \frac{n_{y'} - n_{x'}}{\lambda} l = \frac{n_0^3 r_{63} E_0}{\lambda} l = \frac{n_0^3 r_{63} U}{\lambda},$$

т. е. от приложенного напряжения $U = E_0 l$. Таким образом, изменяя напряжение на электродах, можно в соответствии с (5.19) осуществить амплитудную модуляцию света: $I = I_0 \sin^2 \pi n_0^3 r_{63} U/\lambda$.

Следовательно, полуволновое напряжение для луча, идущего вдоль внешнего поля (по оси z),

$$U_{\parallel} = \lambda/(2 r_{63} n_0^3). \qquad (5.29)$$

Модуляторы с продольным полем не всегда удобны для практического применения, поскольку часть кристалла закрыта электродами и не используется. Если же электроды образованы проводящим прозрачным покрытием, то они обладают заметными потерями. Кроме того, полуволновое напряжение в таких модуляторах довольно велико. Для кристаллов KDP оно составляет около 8 кВ (для света с длиной волны 550 нм). Следует отметить, что в модуляторах такого типа полуволновое напряжение (5.29) не зависит от длины кристалла, так как с увеличением длины кристалла уменьшается напряженность электрического поля. Поэтому необходимое напряжение $U = E_0 l$ остается постоянным.

Более чувствительными являются *модуляторы с поперечным полем*. В этих модуляторах можно достичь существенного уменьшения полуволнового напряжения за счет увеличения длины кристалла.

Один из вариантов модулятора с поперечным полем изображен на рис. 5.4. Кристалл, используемый в модуляторе, вырезают таким образом, чтобы ось z совпадала с направлением внешнего электрического поля, а кристаллографические оси x и y были направлены под углом 45° к передней границе кристалла. После приложения поля индикатриса поворачи-

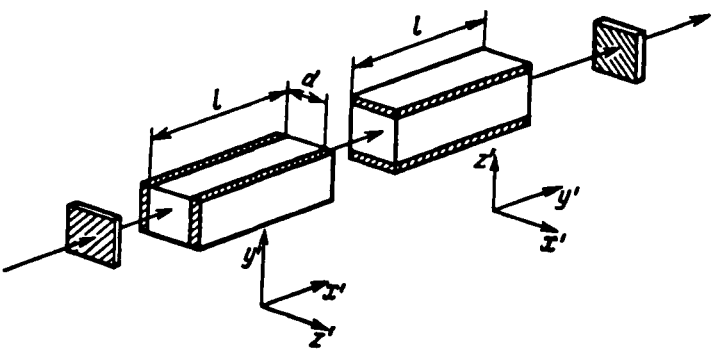


Рис. 5.4. Электрооптический модулятор с поперечный полем

вается вокруг оси z на 45°; при этом ось y' становится параллельной передней грани, а ось x' располагается вдоль направления распространения. Направление поляризации падающего на кристалл света составляет угол 45° с осью z. Поэтому первоначальная волна распадается на две одинаковые ортогональные компоненты, поляризованные вдоль осей y' и z'. После прохождения кристалла между ними образуется сдвиг фаз, зависящий от электрооптического эффекта.

Модулятор состоит из двух электрооптических кристаллов. Второй кристалл вырезают так, что оси x' и y' меняются местами. Расположение кристаллов и направления осей показаны на рис. 5.4. Второй кристалл используется для компенсации значительного естественного двупреломления ($B_0 = n_0 - n_e$).

Естественная анизотропия (B_0) также накладывает жесткие ограничения на допустимый угол расходимости луча, поэтому один кристалл в модуляторах с поперечным полем обычно не применяется. В модуляторе с двумя кристаллами влияние естественной анизотропии практически устранено за счет того, что анизотропию первого кристалла компенсирует анизотропия второго.

При прохождении луча через первый кристалл вертикальной компоненты запаздывание фазы составит $\varphi_{y'} = 2\pi n_{y'} l/\lambda$, а через второй кристалл $\varphi_{z'} = 2\pi n_{z'} l/\lambda$. На это же значение изменяется фаза горизонтальной компоненты в первом кристалле. Во втором кристалле для этой компоненты — $\varphi_{x'} = 2\pi n_{x'} l/\lambda$. Полный сдвиг фаз для первой компоненты $\varphi_{\text{вер}} = \varphi_{y'} + \varphi_{z'}$, а для второй $\varphi_{\text{гор}} = \varphi_{z'} + \varphi_{x'}$. Разность фаз между ортогональными компонентами на выходе второго кристалла

$$\varphi = \varphi_{\text{вер}} - \varphi_{\text{гор}} = 2\pi(n_{y'} - n_{x'})l/\lambda = 2\pi n_0^3 r_{63} E_0 l/\lambda.$$

В этом случае относительная разность оптических путей

$$\Gamma = \frac{n_{y'} - n_{x'}}{\lambda} E_0 l = n_0^3 \frac{r_{63}}{\lambda} \frac{l}{d} U,$$

полуволновое напряжение

$$U_{\perp} = \frac{\lambda}{2n_0^3 r_{63}} \frac{d}{l} = \frac{d}{l} U_{\parallel}. \qquad (5.30)$$

Таким образом, в модуляторе с поперечным полем для получения одинакового сдвига фаз необходимо напряжение, в l/d раз меньшее, чем в модуляторе с продольным полем. Для достаточно длинных кристаллов полуволновое напряжение (5.30) в этом случае оказывается на порядок ниже, чем в выражении (5.29) при продольном поле.

В устройствах для оптической обработки информации сочетание электрооптического модулятора с одноосным кристаллом называют *двоичной ячейкой.*

В зависимости от поляризации выходящий из модулятора луч проходит одноосный кристалл либо без отклонения (считывается «0»), либо с отклонением (считывается «1»). На рис. 5.5, *а* показан одноосный кристалл кальцита, ориентированный таким образом, что его ось анизотропии лежит в плоскости рисунка. Если вектор электрического поля поляризованного луча направлен нормально к оси анизотропии, то он, как обыкновенный луч, проходит по прямой линии. Если падающий луч имеет вертикальную поляризацию, то он смещается пропорционально толщине кристалла. При правильной ориентации кристалла кальцита угол между необыкновенными и обыкновенными лучами равен приблизительно 6° и смещение $D \approx 0{,}1d$.

Поворот плоскости поляризации осуществляется электрооптическим переключателем (модулятором). Прозрачный кристалл для этой цели разрезают перпендикулярно оптической оси *z*. На переднюю и заднюю поверхности кристалла напыляют полупрозрачные электроды. Если кристалл погружен в жидкость для согласования коэффициентов преломления, то электроды обеспечивают прозрачность порядка 93%. Напряжение, приложенное к электродам, создает в кристалле электрическое поле, направленное вдоль светового луча. При полуволновом напряжении разность фаз между обыкновенными и необыкновенными волнами на выхо-

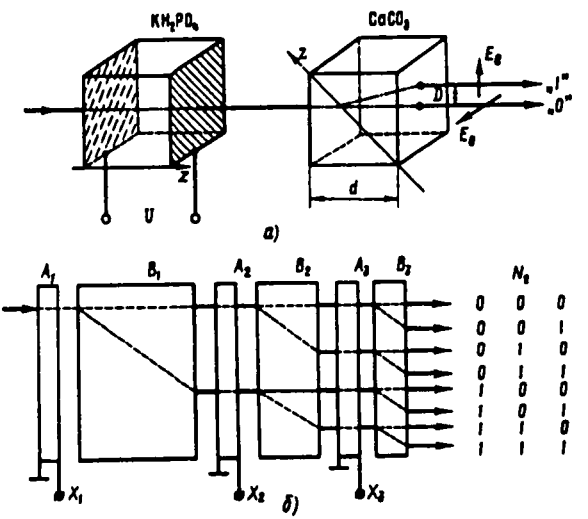


Рис. 5.5. Двоичная оптическая ячейка (*а*) и многопозиционный оптический каскад (*б*)

де из кристалла соответствует половине длины волны. Тогда поляризация суммарной волны на выходе модулятора оказывается повернутой на 90° относительно поляризации первоначальной волны, определяемой входным поляризатором.

Многопозиционный *оптический каскад*, или дефлектор, состоит из нескольких двоичных ячеек, на которые направляется линейно поляризованный луч. На рис. 5.5, *б* показано соединение трех таких ячеек (*А* — электрооптические кристаллы, *В* — кристаллы кальцита). Начальный луч поляризован в горизонтальном направлении. Толщина пластинок кальцита выбрана так, что $B_1 : B_2 : B_3 = 4 : 2 : 1$. При этом каждый каскад обеспечивает смещение между лучами, вдвое меньшее смещения в предыдущем каскаде. Если считать, что в отсутствие напряжения на модуляторе $X = 0$, а при полуволновом напряжении $X = 1$, то каждому сочетанию сигналов на входах модуляторов X_1, X_2, X_3 соответствует свой номер N_2 позиции отклоненного луча в двоичном исчислении. Для n каскадов число положений луча $N = 2^n$.

Преимущества оптической обработки информации могут быть реализованы полностью только при наличии *пространственных модуляторов света*. Так называют материалы или устройства, при прохождении через которые свет, до этого обладавший одинаковыми характеристиками по всему сечению светового потока, приобретает пространственное различие, определяемое свойствами модулирующего материала. Подобные оптические преобразователи можно разделить на два класса: фоточувствительные и электронночувствительные. Первые адресуются путем сканирования или пространственной модуляции коллимированного светового луча, вторые — с помощью сканируемого и модулированного электронного пучка, либо матрицы электродов, расположенных на модуляторе.

Известно большое количество пространственных модуляторов, в основу действия которых положен электрооптический эффект. Одним из них является модулятор света с *электронно-лучевой адресацией* на основе монокристалла дейтерированного дигидрофосфата калия KD_2PO_4 (DKDP). Вариант структуры такого устройства показан на рис. 5.6

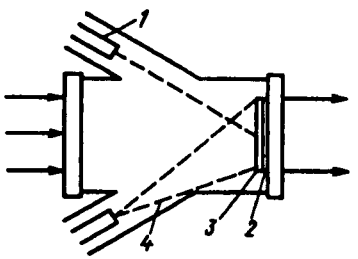


Рис. 5.6. Пространственный электрооптический модулятор с электронно-лучевой адресацией

Запись изображения на мишени *3* (кристалл DKDP) проводят сканирующим по ее поверхности электронным лучом, образующим на ней потенциальный рельеф. Электрод *2* заземлен, а на катод записывающей пушки *1* подан

высокий ускоряющий потенциал. При записи на элементарный участок мишени наносится отрицательный заряд.

Модуляция напряжения на кристалле осуществляется модуляцией тока сканирующего луча входной информацией.

Пространственное распределение поля вызывает соответствующее изменение показателя преломления мишени, которое модулирует поляризацию прошедшего через кристалл считывающего поляризованного света. Если использовать скрещенные поляризатор и анализатор, то возникает пространственное распределение амплитудной модуляции интенсивности считывающего света. В телевизионном режиме (30 кадр/с) считывание осуществлялось за время обратного хода сканирующего луча.

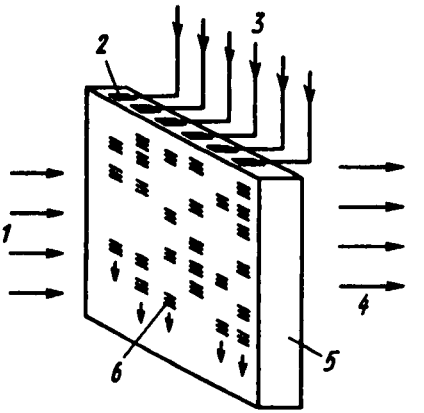


Рис. 5.7. Схема акустооптического пространственно-временно́го модулятора

Стирание информации осуществляется широким электронным пучком *4* при низких ускоряющих напряжениях. В результате потенциал передней поверхности кристалла возвращается к нулю.

Электрооптические коэффициенты кристаллов мало зависят от частоты внешнего электрического поля. Поэтому кристаллические модуляторы могут работать в очень широком диапазоне частот — от нуля до нескольких десятков гигагерц.

Одними из наиболее быстродействующих являются *акустооптические* пространственно-временны́е модуляторы. Двухмерный массив информации может быть сформирован в большеразмерном кристалле *5* (рис. 5.7), когда в каждом из n независимо адресуемых каналов (столбцов) входные электрические сигналы *3* возбуждают в пьезоэлектрических преобразователях *2* временну́ю последовательность m электрических импульсов (строк). Если длительность импульса считывающего светового пучка *1* не превышает несколько наносекунд, движущуюся последовательность акустических волн *6* (цугов) можно рассматривать как неподвижную многоэлементную дифракционную решетку. Поэтому число параллельно формируемых в промодулированном световом пучке *4* элементов изображения определяется произведением $n \times m$.

Одним из наиболее распространенных типов электрически управляемых пространственно-временны́х модуляторов света в настоящее время являются жидкокристаллические приборы. Это обусловлено, в первую очередь, их чрезвычайно высокой чувствительностью к электри-

ческим сигналам и высоким оптическим качеством тонких слоев жидких кристаллов. Элементарная ячейка модулятора представляет собой сэндвич, состоящий из слоя жидкого кристалла толщиной 3...30 мкм, заключенного между электродами на стеклянных подложках; один или оба электрода должны быть полупрозрачными. Были созданы модуляторы в виде матрицы из 128×128 элементов, использующие ориентационные эффекты и фазовые переходы в нематических кристаллах, а также в виде матрицы из 260×260 элементов на основе эффектов динамического рассеяния.

§ 5.9. Электроуправляемые микроволноводные элементы

При модуляции света в диапазоне СВЧ длина волны модулирующего поля становится сравнимой или меньше размеров кристалла в модуляторе и его нельзя больше рассматривать как сосредоточенную емкость. Так как происходит эффективное взаимодействие между модулирующей волной и световым лучом, то требуется согласование групповых скоростей двух полей в кристалле. В большинстве электрооптических материалов показатель преломления в СВЧ-области значительно больше, чем в оптической. Тем не менее фазовое согласование удается получить в полосковом волноводе *2*, который частично заполняется электрооптическим кристаллом *3* (рис. 5.8). Если остальное пространство заполнено воздухом, то фазовая скорость модулирующей волны в данной структуре возрастает по сравнению с фазовой скоростью в кристалле с электродами (см. рис. 5.4) примерно в $\sqrt{w/a}$ раз. Это и позволяет согласовать фазовые скорости модулирующей волны и светового луча *1* при соответствующем выборе геометрических размеров структуры.

Подобные модуляторы с шириной полосы частот до 3 ГГц были созданы на основе кристаллов ADP, KDP и танталата лития.

Широкополосность пластинчатого волновода позволяет использовать его и в сверхскоростном оптическом переключателе (рис. 5.9), в котором световой луч *4* перпендикулярен волноводу *3*, выполненному из $LiTaO_3$. При поглощении оптического импульса *1* длительностью 10 пс в пироэлектрическом кристалле *2* из $LiTaO_3 : Cu^{2+}$, наклеенном на конец прозрачного волновода *3*, генерируется короткий электрический импульс. По мере распространения по волноводу происходит изменение его показателя

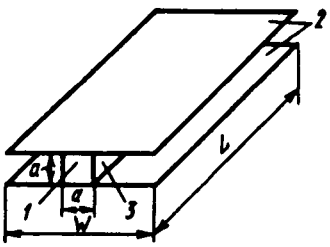


Рис. 5.8. Модулятор бегущей волны

преломления и фазы проходящего поперечного светового луча *4*. Это позволяет наблюдать распространение короткого светового луча вдоль волновода, помещенного между скрещенными поляризатором и анализатором.

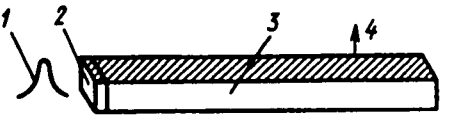


Рис. 5.9. Пикосекундный переключатель

Электрооптичеекий эффект получил широкое применение для модуляции, переключения, пространственного отклонения и сканирования оптических сигналов в пленочных волноводах и световодах. Концентрация оптического поля вблизи управляющих электродов обеспечивает выигрыш по управляющему полю в 10...100 раз по сравнению с объемным управляющим аналогом, обладающим, кроме того, и дифракционными эффектами. В качестве прозрачных материалов с высокой электрооптической эффективностью обычно используют оксид цинка, ниобат и танталат лития, а также полупроводниковые материалы типа $A^{III}B^{V}$.

В фазовом модуляторе (рис. 5.10) управляющее напряжение U прикладывается к электродам Э, изменяя показатель преломления оболочки. Связанное с этим изменение угла падения θ парциальных волн и количества отражений на активном участке создает дополнительный сдвиг фазы волны, пропорциональный изменению управляющего напряжения. Для уменьшения нежелательной амплитудной модуляции в качестве управляющего элемента используется участок оболочки, а не сердечник (полимерное волокно), по которому распространяется основная часть световой энергии.

В электрооптическом модуляторе (рис. 5.11, *а*) два канальных волновода *1* и *2* расположены параллельно на малом расстоянии друг от друга. В процессе распространения излучения происходит перекрытие полей волноводных мод, осуществляющих распределенную связь между ними. Перераспределение энергии между ними зависит от прикладываемого к управляющим электродам напряжения. При соответствующем его значении и длине области взаимодействия может быть обеспечена 100%-ная глубина модуляции. Рассмотренная структура эффективно работает на излучении только одного направления линейной поляризации.

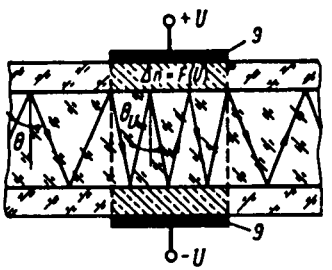


Рис. 5.10. Волоконно-оптический фазовый модулятор

Рабочие характеристики модуляторов с переменными .расстоянием между волноводами и показателем преломления практически не зависят от поляризации излучения. К таким структурам относится широкополосный модулятор на связанных волноводах *2*, работающий по

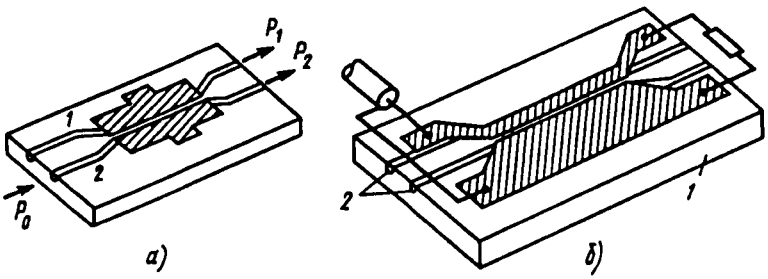


Рис. 5.11. Электрооптические модуляторы на связанных волноводах

схеме бегущей волны (рис. 5.11, *б*) и выполненный на кристалле *1* ниобата лития. На нем достигнута ширина полосы модулирующих частот 7,2 ГГц при управляющем напряжении 4,5 В и факторе качества 7,6 мВт/ГГц, определяемом как мощность возбуждения на единицу полосы частот при эквивалентной 84%-ной модуляции интенсивности излучения.

В пленочных волноводах для управления интенсивностью светового потока и перераспределения его на два направления оказывается эффективным модулятор на основе пространственной дифракции света (рис. 5.12). Он представляет собой сформированный на поверхности электрооптического кристалла *4* тонкий оптический волновод *3*, на поверхность которого нанесены гребенчатые электроды, расположенные под углом Брэгга $\theta_Б$ к направлению распространения света. Если между гребенчатыми электродами имеется разность потенциалов, то вектор напряженности электрического поля и соответствующие изменения показателя преломления имеют противоположные направления между соседними парами штырей. Глубина модуляции показателя преломления в образованной таким образом объемной фазовой дифракционной решетке пропорциональна приложенному напряжению. Интенсивности отклоненного *2* и неотклоненного *1* световых лучей также модулируются изменениями электрического напряжения. При некоторой амплитуде напряжения осуществляется как 100%-ная модуляция интенсивности света, так и перераспределение света между этими двумя направлениями.

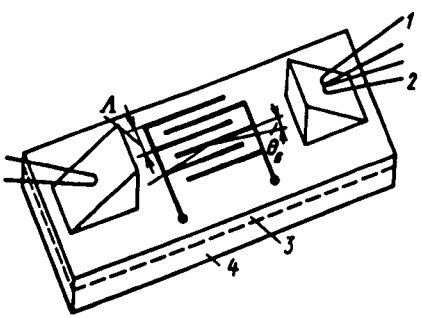


Рис. 5.12. Тонкопленочный дифракционный модулятор

Изменение показателя преломления необыкновенной волны

происходит вблизи поверхности кристалла и составляет примерно 0,002. Типичные значения углового отклонения луча лежат в пределах 1—2°, оптимальные амплитуды напряжения составляют от 5 до 30 В, полоса рабочих частот — единицы гигагерц, минимальная удельная мощность устройства управления модулятора — около 1 мВт/МГц.

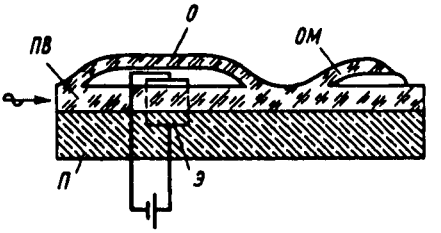


Рис. 5.13. Модулятор на направленных ответвителях

Малые мощности, требуемые для модуляции в интегральной оптике, связаны с тем, что свет распространяется в узких оптических волноводах малого сечения. Высокая концентрация света в особенно тонком электрооптическом слое достигнута использованием двойных гетероструктур из GaAs—$Al_xGa_{1-x}As$. В фазовом волноводном модуляторе с подобной структурой для сдвига фазы на 1 рад была достаточна удельная мощность 0,1 мВт/МГц.

Модулятор на направленных ответвителях с локальной связью показан на рис. 5.13. Между планарным волноводом ПВ и двумя микроволноводными локальными ответвителями О введен слой стекла со скошенными торцами (П — подложка). При подаче на электроды Э управляющего напряжения изменяется показатель преломления, а, следовательно, и сдвиг фаз распространяющихся в нем мод. На следующем ответвителе ОМ происходит отбор моды, приводящий к амплитудной модуляции сигнала.

Интерферометрический модулятор (рис. 5.14) образован тонкопленочным двухканальным синфазным разветвителем с электродами Э. Требуемая для 100%-ной модуляции разность фаз 90° между волнами создается воздействием управляющего напряжения (около 2 В) на оба плеча.

Дифракционный электрооптический переключатель (рис. 5.15) представляет собой планарный волновод из пленки $LiNb_xTa_{1-x}O_3$ на подложке П из ниобата лития, на который нанесена периодическая решетка электродов. Переходы на входные (A_1 и A_2) и выходные (B_1 и B_2) оптические волноводы, расположенные под углом дифракции Брэгга, осуществляют тонкопленочными линзами Л. В отсутствие напряжения на электродах свет распространяется из волновода A_1 в B_1. Под действием управляющего напряжения U в планарном волноводе образуется фазовая дифракционная ре-

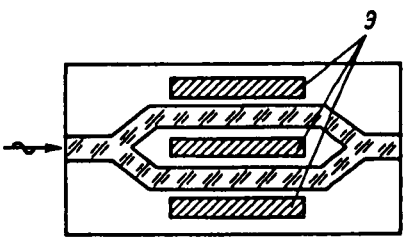


Рис. 5.14. Интерферометрический модулятор

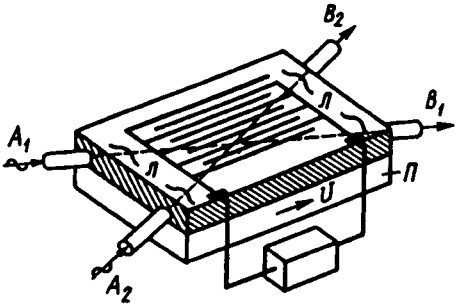


Рис. 5.15. Электрооптический переключатель

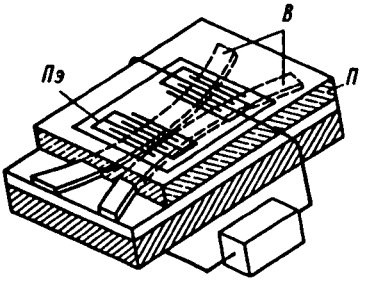


Рис. 5.16. Тонкопленочный акустооптический переключатель

шетка, дифрагируя на которой, свет из волновода A_1 попадает в волновод B_2. Напряжение U также позволяет сигнал, поступающий от волновода A_2, переключать из волновода B_2 в волновод B_1.

Акустооптический переключатель (рис. 5.16) функционирует на направленном ответвителе из планарных волноводов В с показателем преломления бóльшим, чем у подложки П из арсенида галлия. Длина области взаимодействия равна нечетному числу волн. Гребенчатые электроды на пьезоэлектрической пленке Пэ из оксида цинка обеспечивают формирование поверхностной звуковой волны и ввод ее в область взаимодействия, где осуществляется коммутация. Фазовое согласование обоих волноводов обеспечивает полную перекачку энергии.

В системах обработки оптической информации электрооптические модуляторы используют благодаря своим высоким скоростным характеристикам и компактности.

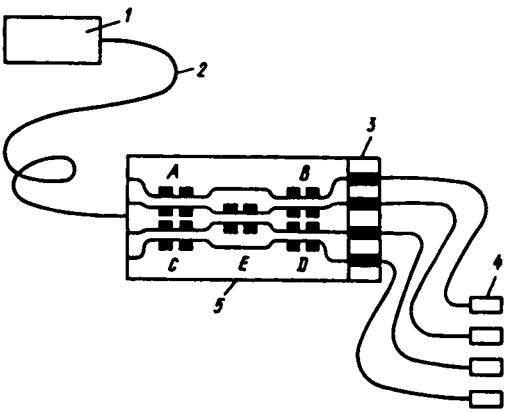


Рис. 5.17. Матрица волноводных модуляторов

На рис. 5.17 показана матрица переключающих элементов *5* из пяти модуляторов, сформированных на основе кристалла ниобата лития и включенных в коммутационный блок волоконно-оптической линии (*1* — излучатель; *2* — входной волновод; *3* — выходные световоды; *4* — фотоприемники; *A*, *B*, *C*, *D*, *E* — группы электродов переключателей). Длина мат-

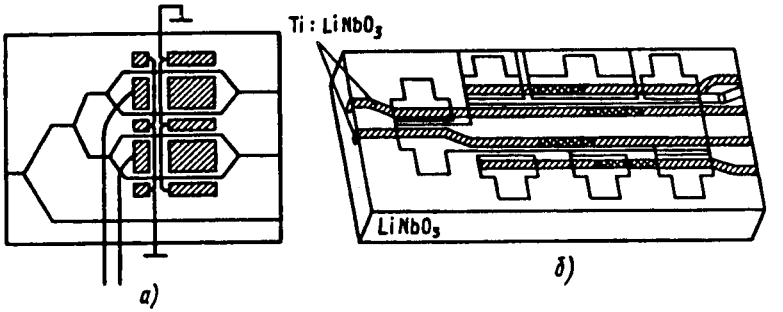


Рис. 5.18. Интегрально-оптические блоки быстродействующих АЦП

рицы 4 см, потери оптического тракта 6,8 дБ, управляющее напряжение не превышает 28 В. Матрица оптических переключателей на кристалле арсенида галлия имеет рабочее напряжение примерно вдвое меньшее, однако во столько же раз возрастают потери.

Интегрально-оптические блоки быстродействующих аналого-цифровых преобразователей (АЦП) используют несколько электрооптических модуляторов, сформированных на единой подложке. Оптический процессор на базе АЦП (рис. 5.18, *а*) включает систему нескольких параллельно включенных интерферометрических элементов. Каждый из элементов обеспечивает модуляцию светового сигнала по косинусоидальному закону $\cos^2 KU$ с различными периодами (постоянная K пропорциональна длине электрода). Выбор электродов кратной длины параллельно включенных интерферометрических элементов обеспечивает нужный разрядный код. Дискретизация преобразуемого электрического сигнала 1 ГГц, а разрядность преобразования — 6 бит.

На рис. 5.18, *б* показана конструкция блока интегрально-оптического

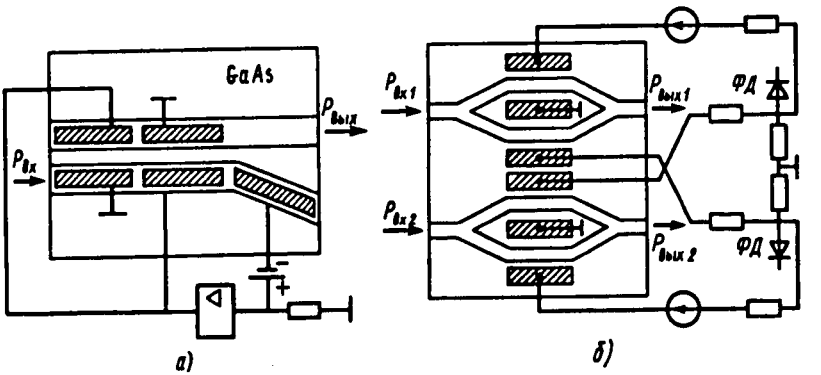


Рис. 5.19. Интегральные электрооптические бистабильные устройства

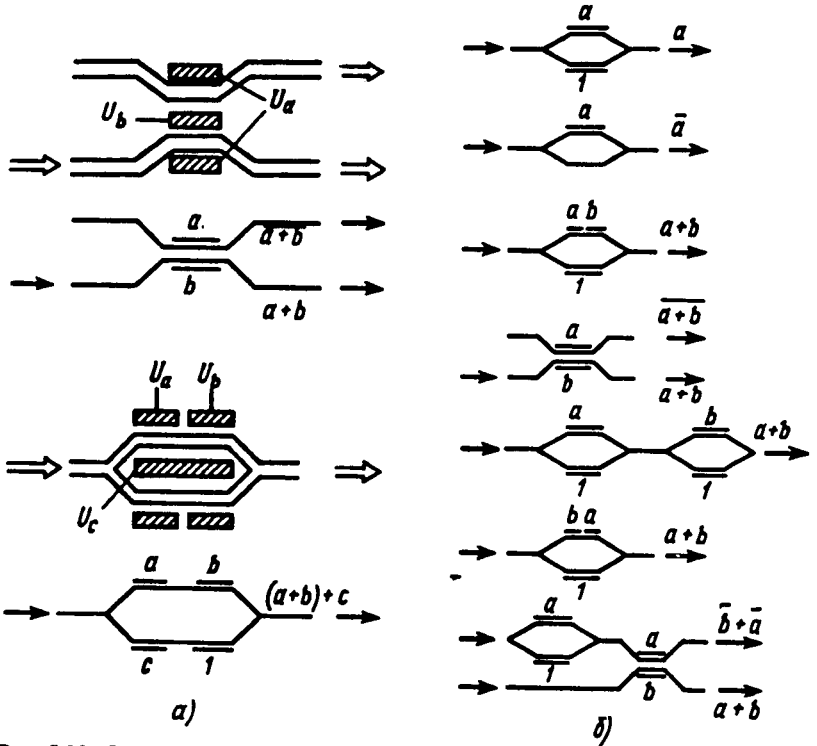


Рис. 5.20. Логические преобразователи на основе интегрально-оптических модуляторов

АЦП с модуляторами на связанных волноводах, обладающими меньшими оптическими потерями.

Принцип действия и технология производства волноводных модуляторов и их массивов лежит в основе изготовления оптических бистабильных устройств, являющихся основой интегрально-оптических логических схем и схем памяти. На рис. 5.19, *а* дана структура бистабильного устройства на основе интегрально-оптического модулятора на связных волноводах, а на рис. 5.19, *б* — структура мультивибратора на модуляторах интерференционного типа (ФД — фотодетектор). На рис. 5.20, *а* показаны логические преобразователи на основе этих двух типов модуляторов, а на рис. 5.20, *б* — функциональные схемы логических преобразователей на основе интегрально-оптических бистабильных устройств.

Электрооптический эффект нашел применение и в безлинзовом спектроанализаторе. Излучение гетеролазера Л вводится в планарный волновод ПВ и дифрагирует на тех участках волновода, где происходят возмущения показателя преломления, вызванные полем электродов Э

(рис. 5.21). Электрооптический эффект создает в волноводе ряд накладывающихся друг на друга «зонных пластинок», фокусирующих падающую волну в различных точках вдоль оси АВ. Электроды осуществляют пространственную модуляцию света, формирующую на линейке фотоприемников ФД дискретное фурье-преобразование приложенного к электродам напряжения. Функциональная гибкость электрооптического анализатора намного больше, чем акустооптического.

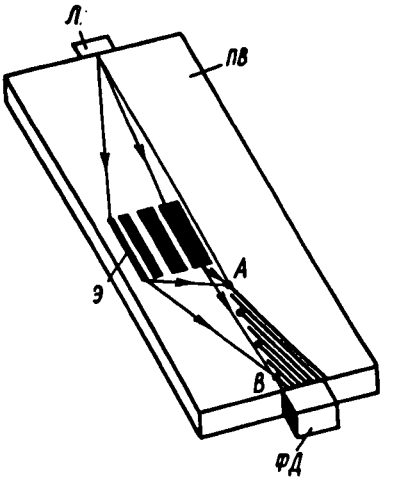


Рис. 5.21. Безлинзовый интегрально-оптический спектроанализатор

В заключение следует отметить, что свойства анизотропной среды по отношению к электромагнитным волнам зависят в общем случае не только от тензора ε_{ik}, но и тензора μ_{ik}. Несмотря на то, что в настоящей главе для определенности рассматривались главным образом электрическое поле и тензор ε_{ik}, все полученные результаты полностью справедливы и для тензора μ_{ik}.

Глава 6
Электромагнитное поле в гиротропной среде

Гиротропией (gyros — круг) или *оптической активностью* называют явление оптической вращательной способности, при котором происходит вращение плоскости поляризации электромагнитной волны.

Оптическая активность связана с зависимостью поляризации от пространственной структуры молекул и характера их расположения в решетке кристалла. Особенности пространственной структуры молекулы могут влиять на распространение электромагнитной волны тем больше, чем больше отличаются параметры волны (амплитуды и фазы) в разных точках молекулы. Поэтому они характеризуются отношением поперечника молекулы к длине волны. Для видимого света это отношение имеет порядок 10^{-3}. Несмотря на малость этой величины, именно с ней связана сущность естественной гиротропии. В таких средах уравнения связи имеют вид $\mathbf{D} = \varepsilon[\mathbf{E} + (\alpha\nabla)\times\mathbf{E}]$, $\mathbf{B} = \mu[\mathbf{H} + (\tilde{\alpha}\nabla)\times\mathbf{H}]$, где α — *тензор гирации* ($\tilde{\alpha}$ — транспонированный тензор). В этом случае связь между напряженностью и индукцией поля волны зависит от направления распространения и скорости волны. Например, для плоской волны $\nabla = -jk$ и $\dot{\mathbf{D}} = \varepsilon[\dot{\mathbf{E}} - j(\alpha k)\times\dot{\mathbf{E}}]$.

Гиротропию, возникающую под воздействием магнитного или электрического поля, называют *вынужденной* или *индуцированной*. В этом случае соотношение между индукцией и напряженностью поля часто не зависит от направления распространения и скорости волны и характеризуется комплексными тензорами проницаемости

$$\underline{\varepsilon} = \underline{\varepsilon}_1 + j\varepsilon_0\mathbf{a}\times, \quad \underline{\mu} = \underline{\mu}_1 + j\mu_0\mathbf{G}\times, \tag{6.1}$$

т. е.

$$\dot{\mathbf{D}} = \underline{\varepsilon}_1\dot{\mathbf{E}} + j\varepsilon_0\mathbf{a}\times\dot{\mathbf{E}}, \quad \dot{\mathbf{B}} = \underline{\mu}_1\dot{\mathbf{H}} + j\mu_0\mathbf{G}\times\dot{\mathbf{H}},$$

где **a**, **G** — *векторы гирации*.

В некоторых случаях, например при наличии постоянного внутримолекулярного или межмолекулярного поля и в отсутствие внешних электрических и магнитных полей, среды могут обладать проницаемостями типа (6.1). Их гиротропию называют *собственной*.

Гиротропными называют среды, в которых хотя бы одна из величин **a**, **G** или α отлична от нуля. При этом необходимо иметь в виду, что в некоторых гиротропных средах не для всякого направления происходит вращение плоскости поляризации плоской волны и есть гиротропные среды, в которых вообще невозможно вращение плоскости поляризации ни для какого направления (кристаллы *планальных* классов средних сингоний). Про такие среды можно сказать лишь то, что в них существуют направления распространения однородной плоской волны, вдоль которых волна приобретает эллиптическую поляризацию и при отсутствии диссипации энергии.

Впервые поворот плоскости поляризации был обнаружен в 1811 г. при прохождении света через немагнитный кристалл (кварц). Однако наиболее сильно гиромагнитные явления проявляются в ферромагнитных материалах. Угол поворота плоскости поляризации в прозрачных тонких ферромагнитные пленках на несколько порядков выше, чем в неферромагнитных материалах.

Решающее значение для технического использования гиромагнитных эффектов имело исследование свойств и овладение технологией изготовления ферритов, т. е. ферромагнитных диэлектриков. Их слабая электрическая проводимость (в $10^{11}...10^{13}$ раз меньше, чем у ферромагнитных металлов) обеспечивает относительно малые потери и затухание электромагнитных волн, позволяя в наиболее чистом виде реализовать взаимодействие электромагнитных колебаний с намагниченной средой. С этим связано широкое применение многочисленных ферритовых устройств в технике СВЧ и в вычислительных устройствах. Использование ферритов позволяет создать такие элементы электрической цепи, которые не удовлетворяют принципу взаимности и отличаются высокой скоростью управления режимом цепи: быстродействующие переключатели, циркуляторы, модуляторы, перестраиваемые по частоте фильтры, ослабители, смесители, генераторы, преобразователи частоты, параметрические усилители и др. По этой причине свойства гиротропных сред в дальнейшем рассматриваются применительно к ферритам.

§ 6.1. Магнитная проницаемость намагниченного феррита

Гиромагнитные свойства обнаруживаются при одновременном воздействии на феррит постоянного $\mathbf{H}_0$ и переменного $\mathbf{h}$ магнитных полей. Поэтому напряженность результирующего поля

$$\mathbf{H} = \mathbf{H}_0 + \mathbf{h}. \tag{6.2}$$

Аналогично, намагниченность

$$\mathbf{M} = \mathbf{M}_0 + \mathbf{m} \tag{6.3}$$

имеет постоянную $\mathbf{M}_0$ и переменную $\mathbf{m}$ составляющие.

В дальнейшем изложении будем предполагать, что переменные магнитное поле и намагниченность достаточно малы и между ними имеется линейное соотношение

$$m_i = \chi_{ik} h_k, \tag{6.4}$$

не зависящее от амплитуды переменного поля, причем χ_{ik} — тензор магнитной восприимчивости для электромагнитной волны.

Как известно из электродинамики, на элемент объема с магнитным моментом $\mathbf{M}$ действует со стороны магнитного поля пара сил, т. е. вра-

щающий механический момент $\mathbf{T} = \mu_0 \mathbf{M} \times \mathbf{H}_{эф}$. Если рассматриваемый элемент объема обладает механическим моментом количества движения **J**, то, как известно из механики, конец этого вектора будет перемещаться под воздействием вращающего момента **T** со скоростью $\mathrm{d}\mathbf{J}/\mathrm{d}t = \mathbf{T} = \mu_0 \mathbf{M} \times \mathbf{H}_{эф}$. Ферромагнитные свойства вещества связаны главным образом со спиновым магнитным моментом. Поэтому гиромагнитное отношение $\gamma = -M/J = e/m = 1{,}76 \times 10^{11}$ К/кг, и предыдущее уравнение принимает вид

$$\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}t} = -\gamma \mu_0 \mathbf{M} \times \mathbf{H}_{эф}. \qquad (6.5)$$

Из уравнения (6.5) непосредственно следует вывод о том, что в случае постоянного магнитного поля вектор магнитного (и механического) момента электрона прецессирует вокруг направления поля с угловой скоростью

$$\omega_H = \gamma \mu_0 H_{эф}, \qquad (6.6а)$$

называемой *частотой гиромагнитного резонанса*. Если напряженность магнитного поля выражать в амперах, деленных на метр, то $\omega_H = 3{,}5 \cdot 10^4 H_{эф}$ c^{-1}.

Вектор $\mathbf{H}_{эф}$ соответствует суммарному магнитному полю, действующему на магнитный момент частицы,

$$\mathbf{H}_{эф} = \mathbf{H} + \mathbf{H}_{ан} \qquad (6.6б)$$

и включает кроме вектора внешнего поля **H** вектор эффективного поля $\mathbf{H}_{ан}$ от обменного и дипольного взаимодействия, размагничивания (нескомпенсированные заряды доменов на поверхности) и анизотропии. Последнее слагаемое объясняет происхождение так называемого естественного ферромагнитного резонанса, наблюдаемого при отсутствии внешнего магнитного поля. Резонансная частота $\omega_{ест} = \gamma \mu_0 H_{ан}$, обусловленная внутренними размагничивающими полями (вследствие неоднородностей в распределении намагниченности в феррите), лежит в области дециметровых волн (500... 3000 МГЦ).

Магнитные потери, всегда имеющие место в реальных ферромагнитных средах, обусловлены взаимодействием спинов с окружающей средой. Магнитные потери приводят к тому, что свободная прецессия намагниченности в реальных средах затухает, и при отсутствии внешнего переменного поля устанавливается равновесное состояние, соответствующее статической намагниченности M_0. Электромагнитная энергия, связанная с прецессией намагниченности, передается в конечном счете кристаллической решетке ферромагнетика, преобразуясь в энергию ее тепловых колебаний.

Учет потерь в ферритах приводит к комплексным значениям компонент тензоров магнитной восприимчивости и проницаемости, причем они остаются конечными при ферромагнитном резонансе. Для оценки их значений и зависимости от частоты и подмагничивающего поля исходное уравнение (6.5) должно быть дополнено членом, учитывающим потери и уменьшающим энергию прецессии электрона. Из различных форм записи диссипативного члена рассмотрим выражение, выведенное в предположении действия на намагниченность кроме поля $\mathbf{H}_{эф}$ некоторого эффективного поля «сил трения», антипараллельного и пропорционального скорости изменения намагниченности:

$$\frac{d\mathbf{M}}{dt} = -\gamma\mu_0 \mathbf{M} \times \mathbf{H}_{эф} + \frac{\delta}{M} \mathbf{M} \times \frac{d\mathbf{M}}{dt}, \qquad (6.7)$$

где δ — безразмерный параметр, характеризующий потери.

Обычно потери столь малы, что второй член уравнения (6.7) пренебрежимо мал по сравнению с первым. Тогда во втором члене можно приближенно считать, что $d\mathbf{M}/dt = -\gamma\mu_0 \mathbf{M} \times \mathbf{H}_{эф}$, и уравнение (6.7) переписать в виде

$$\frac{d\mathbf{M}}{dt} = -\gamma\mu_0 \mathbf{M} \times \mathbf{H}_{эф} - \frac{\gamma\mu_0}{M} \delta \mathbf{M} \times \mathbf{M} \times \mathbf{H}_{эф}, \qquad (6.8)$$

где величина $\delta \ll 1$ определяется экспериментальным путем.

Уравнение (6.8) было предложено в 1935 г. Л. Д. Ландау и Е. М. Лифшицем и носит их имя, как и его другая форма

$$\frac{d\mathbf{M}}{dt} = -\gamma\mu_0 \mathbf{M} \times \mathbf{H} + \eta\left(\mathbf{H} - \frac{\mathbf{M}}{M^2}\mathbf{H}\mathbf{M}\right), \qquad (6.9)$$

где $\eta = \gamma\mu_0\delta M \ll 1$. Уравнение Ландау—Лифшица составлено в предположении, что прецессия намагниченности и при наличии потерь происходит при неизменной длине вектора $\mathbf{M}$.

В дальнейшем будем предполагать, что феррит намагничен однородным внешним полем до насыщения, т. е. магнитные моменты всех атомов параллельны (обменные силы можно не учитывать) и направлены вдоль приложенного поля $\mathbf{H}_0 = H_0\mathbf{e}_z$, а вектор намагниченности $\mathbf{M}_0 = M_0\mathbf{e}_z$ ориентирован одинаково с вектором поля $\mathbf{H}_0$:

$$\mathbf{M}_0 \times \mathbf{H}_0 = 0. \qquad (6.10)$$

Будем также считать, что переменные составляющие $h \ll H_0$ и $m \ll M_0$. Тогда, подставив (6.2) и (6.3) в уравнение (6.9), пренебрегая малыми членами второго порядка и учитывая (6.10), найдем

$$\frac{\mathrm{d}\mathbf{m}}{\mathrm{d}t}+\gamma\mu_0(\mathbf{M}_0\times\mathbf{h}+\mathbf{m}\times\mathbf{H}_0)=\eta\left(\mathbf{H}_0+\mathbf{h}-\mathbf{M}_0\frac{\mathbf{H}_0\mathbf{M}_0+\mathbf{H}_0\mathbf{m}+\mathbf{M}_0\mathbf{h}}{M_0^2}-\right.$$

$$\left.-\mathbf{m}\frac{\mathbf{H}_0\mathbf{M}_0}{M_0^2}\right)\equiv\eta\left(\mathbf{h}-\mathbf{H}_0\frac{m_z}{M_0}-\frac{\mathbf{M}_0}{M_0}h_z-\frac{H_0}{M_0}\mathbf{m}\right).$$

Переходя к комплексной форме векторов **m** и **h** и проектируя полученное уравнение на оси координат, нетрудно получить сначала

$$(j\omega+\omega_r)\dot{m}_x+\omega_H\dot{m}_y=\eta\dot{h}_x+\omega_M\dot{h}_y,$$

$$-\omega_H\dot{m}_x+(j\omega+\omega_r)\dot{m}_y=-\omega_M\dot{h}_x+\eta\dot{h}_y,$$

$$(j\omega+2\omega_r)\dot{m}_z\approx 0,$$

а затем

$$\dot{m}_x=\chi_0\frac{\omega_0^2+j\omega\omega_r}{\omega_0^2-\omega^2+2j\omega\omega_r}\dot{h}_x+j\chi_0\frac{\omega\omega_H}{\omega_0^2-\omega^2+j2\omega\omega_r}\dot{h}_y,$$

$$\dot{m}_y=-j\chi_0\frac{\omega\omega_H}{\omega_0^2-\omega^2+j2\omega\omega_r}\dot{h}_x+\chi_0\frac{\omega_0^2+j\omega\omega_r}{\omega_0^2-\omega^2+j2\omega\omega_r}\dot{h}_y, \qquad (6.11)$$

$$\dot{m}_z\approx 0,$$

где $\omega_r=\eta H_0/M_0$; $\omega_M=\gamma\mu_0 M_0$; $\chi_0=M_0/H_0$;

$$\omega_0=\sqrt{\omega_H^2+\omega_r^2}=\omega_H\sqrt{1+\delta^2} \qquad (6.12)$$

— резонансная частота.

Переменная составляющая магнитной индукции $\mathbf{b}=\mu_0(\mathbf{h}+\mathbf{m})=\mu\mu_0\mathbf{h}$. Отсюда с учетом (6.11) тензор магнитной проницаемости

$$\underline{\mu}=\begin{bmatrix}\underline{\mu}_r & -j\underline{\mu}_a & 0\\ j\underline{\mu}_a & \underline{\mu}_r & 0\\ 0 & 0 & \mu_z\end{bmatrix}, \qquad (6.13)$$

где

$$\underline{\mu}_r=1+\chi_0\frac{\omega_0^2+j\omega\omega_r}{\omega_0^2-\omega^2+j2\omega\omega_r}=\mu'-j\mu''; \qquad (6.14)$$

$$\underline{\mu}_a=-\chi_0\frac{\omega\omega_H}{\omega_0^2-\omega^2+j2\omega\omega_r}=\mu_a'-j\mu_a'';$$

$$\mu_z=1.$$

Поглощение энергии высокочастотного поля в феррите характеризует мнимые составляющие компонент тензора магнитной проницаемости

$$\mu'' = \chi_0 \frac{\omega\omega_r(\omega_0^2 + \omega^2)}{(\omega_0^2 - \omega^2)^2 + 4\omega^2\omega_r^2};$$
$$\mu_a'' = \chi_0 \frac{2\omega^2\omega_H\omega_r}{(\omega_0^2 - \omega^2)^2 + 4\omega^2\omega_r^2}. \qquad (6.15)$$

Нетрудно видеть, что зависимость мнимых компонент от частоты и намагничивающего поля H_0 носит резонансный характер.

При резонансной частоте (6.12) они достигают максимальных значений:

$$\mu''_{max} = \chi_0 \frac{\omega_0}{2\omega_r} = \frac{\chi_0}{2\delta}\frac{\omega_0}{\omega_H} = \frac{\chi_0}{2}\sqrt{1 + \frac{1}{\delta^2}};$$

$$\mu''_{a\,max} = \chi_0 \frac{\omega_H}{2\omega_r} = \frac{\chi_0}{2\delta}.$$

Ширина $2\Delta\omega$ резонансной кривой (6.15) на половинном уровне может быть найдена из условия

$$\mu'' = \chi_0 \frac{(\omega_0 \pm \Delta\omega)[\omega_0^2 + (\omega \pm \Delta\omega)^2]\omega_r}{[\omega_0^2 - (\omega_0 \pm \Delta\omega)^2]^2 + 4(\omega_0 \pm \Delta\omega)^2\omega_r^2} = \frac{\chi_0\omega_0}{4\omega_r} = \frac{\mu''_{max}}{2},$$

откуда для ферритов с малыми потерями ($\Delta\omega \ll \omega_0$) $\Delta\omega = \omega_r = \delta\omega_H$.

Величина ω_r характеризует, таким образом, затухание в феррите. Ее удвоенное значение равно ширине кривых μ'' и μ_a'' на половинном уровне. Если частота фиксирована, то можно говорить о зависимости тензора от напряженности поля H_0. Ширину кривых мнимых компонент намагниченности по половинному уровню поля находят из условия $\gamma\mu_0\Delta H = \omega_r : 2\Delta H = 2\omega_r/(\gamma\mu_0) = 2\delta H_0$.

Параметр δ характеризует относительное смещение резонансной частоты (6.12) от ее значения (6.6) в отсутствие потерь и равен нормированной ширине резонансных кривых ($\delta = \Delta\omega/\omega_H \approx \Delta H/H_0$). С уменьшением потерь и соответственно δ помимо сужения кривых происходит рост пиков мнимых составляющих. Потери тем меньше, чем однороднее феррит. Так, для иттриевого феррита со структурой граната, в котором отсутствует беспорядок в распределении магнитных ионов по узлам решетки, была получена кривая шириной менее 0,4 А/см. Наиболее узкие резонансные кривые ферритов со структурой шпинели имеют ширину 8...16 А/см (марганцевый и марганцово-магниевый феррит).

Из формул (6.14) и соответствующих им графиков (рис. 6.1, *а*) видно, что вещественные составляющие компонент тензора магнитной проницаемости при резонансной частоте не претерпевают разрыва и не уходят в бесконечность (как это получается при пренебрежении потерями,

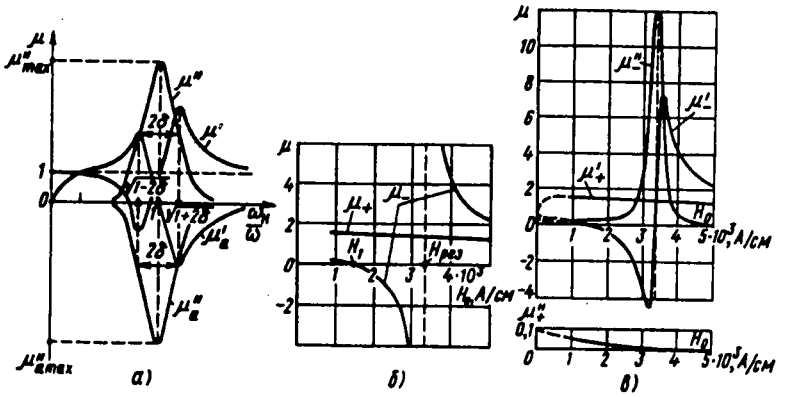


Рис. 6.1. Магнитная проницаемость намагниченного феррита:

а — вещественные и мнимые части компоненты тензора проницаемости в зависимости от отношения ω_H/ω; *б* — проницаемость для поляризованных по кругу волн без учета потерь; *в* — вещественные и мнимые части проницаемости для волн с круговой поляризацией с учетом потерь

когда $\delta = 0$), хотя и достигают максимальных значений вблизи этой частоты (при $\omega = \omega_H \sqrt{1 \pm 2\delta}$). В точке резонанса проницаемость μ_a'' всегда обращается в нуль, после чего становится отрицательной и в больших магнитных полях стремится к нулю. При сравнительно больших потерях проницаемость μ' нигде не принимает отрицательных значений.

В упрощенном расчете тензора (6.13) компонента μ_z получилась равной единице. Однако по своему физическому смыслу она определяет магнитную проницаемость феррита для обыкновенной волны. Поэтому в намагниченных ферритах она не зависит от напряженности внешнего поля, но зависит от намагниченности насыщения различных типов ферритов. Компонента μ_z у большинства ферритов несколько меньше единицы и определяется экспериментальным путем.

Структура тензора (6.13) показывает, что его можно представить в характерной для гиротропной среды форме (6.1), поскольку в данном случае

$$\underline{\mu}_1 = \begin{bmatrix} \underline{\mu}_r & 0 & 0 \\ 0 & \underline{\mu}_r & 0 \\ 0 & 0 & \mu_z \end{bmatrix}$$

и вектор гарации $\mathbf{G} = -j\underline{\mu}_a \mathbf{e}_z$.

Тензор (6.13) можно «диагонализировать», если заменить поперечные составляющие векторов электромагнитной волны их линейными комбинациями:

$$\dot{h}_+ = \dot{h}_x - j\dot{h}_y,\ \ \dot{h}_- = \dot{h}_x + j\dot{h}_y;$$

$$\dot{b}_+ = \dot{b}_x - j\dot{b}_y,\ \ \dot{b}_- = \dot{b}_x + j\dot{b}_y.$$

Тогда из выражения (6.13) следует

$$\dot{b}_\pm = (\underline{\mu}_r \pm \underline{\mu}_a)\dot{h}_\pm. \qquad (6.16)$$

Это означает, что для новых переменных тензор проницаемости

$$\underline{\mu} = \begin{bmatrix} \underline{\mu}_r + \underline{\mu}_a & 0 & 0 \\ 0 & \underline{\mu}_r - \underline{\mu}_a & 0 \\ 0 & 0 & \mu_z \end{bmatrix}$$

стал диагональным.

Нетрудно убедиться, что в уравнении (6.16) верхние знаки соответствуют параметрам волны с поперечным магнитным полем, поляризованным по кругу с правым вращением, нижние — параметрам такой же волны с левым вращением.

Действительно, в правополяризованной волне, аналогично (1.29), $\dot{h}_y = j\dot{h}_x = jh$, $h_z = 0$; вектор напряженности $\dot{\mathbf{h}}$ имеет единственную компоненту $h_+ = 2h$, поскольку $h_- = 0$.

В левополяризованной волне, где, аналогично (1.30), $\dot{h}_y = -j\dot{h}_x = -jh$ и $h_z = 0$, вектор $\dot{\mathbf{h}}$ имеет только компоненту $h_- = 2h$, поскольку $h_+ = 0$.

В случае пренебрежимо малых потерь ($\delta = 0$) зависимость от намагничивающего поля магнитной проницаемости для право- и левополяризованных волн

$$\mu_\pm = 1 + \chi_0 \frac{\omega_H}{\omega_H \pm \omega} \qquad (6.17)$$

приведена на рис. 6.1, *б* при частоте 9375 МГц и намагниченности насыщения 16 мТл. Из рис. 6.1, *б* отчетливо видно различие в распространении поперечных волн с правым и левым вращением вдоль вектора гиротропии. Резонансный характер имеет проницаемость только той поперечной волны, направление вращения которой совпадает с направлением собственной прецессии намагниченности.

Это различие сохраняется и для феррита с потерями. Из приведенных на рис. 6.1, *в* графиков составляющих магнитной проницаемости

видно, что для волны с левым вращением в точке резонанса мнимая составляющая достигает большого значения ($\mu_-'' \approx \chi_0/\delta$). Для волны с правым вращением зависимость мнимой части проницаемости, как и вещественной, не имеет резонансного характера, меняется монотонно, оставаясь всюду малой, в том числе и при резонансе ($\mu_+'' \approx \delta\omega_M/\omega$).

§ 6.2. Электромагнитные волны в гиротропной среде

При рассмотрении свойств волн в безграничном феррите исходят из решения уравнений Максвелла

$$\operatorname{rot}\mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}; \quad \operatorname{div}\mathbf{D} = 0;$$

$$\operatorname{rot}\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}; \quad \operatorname{div}\mathbf{B} = 0, \tag{6.18}$$

в которых должны быть учтены гиротропные свойства среды. При этом имеют в виду процессы, которые складываются из постоянной и быстропеременной компонент $\mathbf{H} = \mathbf{H}_0 + \mathbf{h}$, $\mathbf{E} = \mathbf{E}_0 + \mathbf{e}$, $\mathbf{B} = \mathbf{B}_0 + \mathbf{b}$, $\mathbf{M} = \mathbf{M}_0 + \mathbf{m}$.

Если возбуждающие поля и токи изменяются по гармоническому закону, то переменные составляющие изменяются также по гармоническому закону. Тогда для постоянных составляющих и комплексных амплитуд могут быть составлены две независимые системы уравнений, совпадающие по форме с (6.18), причем для первых составляющих система распадается, в свою очередь, на системы уравнений электростатики и магнитостатики. При дальнейшем рассмотрении процессов в ферритах представляет интерес лишь система уравнений для гармонических составляющих

$$\operatorname{rot}\dot{\mathbf{h}} = j\omega\dot{\mathbf{d}}, \quad \operatorname{rot}\dot{\mathbf{e}} = -j\omega\dot{\mathbf{b}}. \tag{6.19}$$

Входящие в уравнения (6.19) величины являются макроскопическими — усредненными в пространстве и во времени значениями соответствующих микроскопических величин. Эти уравнения должны быть, как известно, дополнены соотношениями, связывающими $\mathbf{D}$ и $\mathbf{B}$ с $\mathbf{E}$ и $\mathbf{H}$. Последние могут быть получены теоретически или экспериментально, и в макроскопической электродинамике рассматриваются как заданные.

Разделение микроскопической и макроскопической задач при интегрировании уравнений Максвелла в некоторых случаях невозможно, и тогда обе задачи нужно решать совместно. Такое положение имеет место в ферромагнитных металлах при высоких частотах, когда глубина проникновения поля в металл становится сравнимой или меньшей, чем размеры доменов. В ферритах, проводимость которых в $10^8 \ldots 10^{13}$ раз меньше

электропроводности металлов, глубина проникновения значительно превышает размеры доменов. В этом случае разделение указанных задач возможно. Для феррита без потерь, намагниченного в направлении оси z,

$$\dot{b}_x = \mu_r\mu_0\dot{h}_x - j\mu_a\mu_0\dot{h}_y;$$

$$\dot{b}_y = \mu_r\mu_0\dot{h}_y + j\mu_a\mu_0\dot{h}_x; \qquad (6.20)$$

$$\dot{b}_z = \mu_z\mu_0\dot{h}_z;$$

$$\dot{d} = \varepsilon\dot{e}. \qquad (6.21)$$

Линейность соотношений (6.20) и (6.21) носит, как отмечалось, неуниверсальный характер и справедлива только при определенных условиях.

Связь между переменными составляющими может быть принята линейной, если эти компоненты достаточно малы. От постоянных компонент поля зависят лишь коэффициенты в линейных соотношениях.

Для плоской волны, направление которой задано волновым вектором $\mathbf{k}$, уравнения (6.19) упрощаются:

$$\mathbf{k}\times\dot{\mathbf{h}} = -\omega\varepsilon\dot{\mathbf{e}}, \qquad \mathbf{k}\times\dot{\mathbf{e}} = \omega\mu\mu_0\dot{\mathbf{h}}, \qquad (6.22)$$

откуда

$$\mathbf{k}\times\dot{\mathbf{h}}\times\mathbf{k} \equiv k^2\dot{\mathbf{h}} - \mathbf{k}(\mathbf{k}\dot{\mathbf{h}}) = \omega^2\varepsilon\mu\mu_0\dot{\mathbf{h}} \qquad (6.23)$$

или

$$(k^2\delta_{ik} - k_ik_k - \omega^2\varepsilon\mu_{ik}\mu_0)\dot{h}_k = 0. \qquad (6.24)$$

Волновое число может быть найдено из условия совместимости уравнений (6.24), т. е. условия, когда определитель системы

$$\left|k^2\delta_{ik} - k_ik_k - \omega^2\varepsilon\mu_{ik}\mu_0\right| = 0. \qquad (6.25)$$

Например, если волновой вектор $\mathbf{k}$ образует угол θ с направлением намагничивающего поля $\mathbf{H}_0$ (ось z), а плоскость ($\mathbf{k}$, $\mathbf{H}_0$) составляет угол 90° с осью x, решение уравнения (6.25) с учетом (6.13) для феррита без потерь будет иметь вид

$$k_{1,2}^2 = \frac{1}{2}\omega^2\varepsilon\mu_z\mu_0\left[\frac{(\mu_r^2/\mu_z^2 - \mu_a^2/\mu_z^2 - \mu_r/\mu_z)\sin^2\theta + 2\mu_r/\mu_z}{(\mu_r/\mu_z - 1)\sin^2 + 1} \pm\right.$$

$$\left. \pm \frac{\sqrt{(\mu_a^2/\mu_z^2 - \mu_r^2/\mu_z^2 + \mu_r/\mu_z)^2\sin^4\theta + 4\mu_a^2\cos^2\theta/\mu_z^2}}{(\mu_r/\mu_z - 1)\sin^2 + 1}\right], \qquad (6.26)$$

где θ — угол между направлениями распространения волны и намагничивающего поля (между векторами $\mathbf{k}$ и $\mathbf{H}_0$).

Полученные значения волновых чисел k_1 и k_2 соответствуют двум волнам, распространяющимся независимо друг от друга в одном направлении $\mathbf{k}$ с фазовыми скоростями

$$v_{1,2} = \omega / k_{1,2} = 1 / \sqrt{\varepsilon \mu_{эф\,1,2} \mu_0}\,, \tag{6.27}$$

где

$$\mu_{эф\,1,2} = k_{1,2}^2 / (\omega^2 \varepsilon \mu_0) \tag{6.28}$$

— эффективная магнитная проницаемость среды для волн, распространяющихся в направлении **k**.

Из выражения (6.28) следует, что магнитные свойства среды не одинаковы для различных волн и направлений. В частности, только для одной из этих волн с индексом *1* имеет место резонанс, т. е. возможны бесконечные значения $\mu_{эф}$. Если подставить в уравнения (6.26) и (6.28) значения компонент тензора магнитной проницаемости (6.14) и принять $\omega_r = 0$, то для этой волны резонансная частота

$$\omega_{рез} = \omega_H \sqrt{1 + \frac{M_0}{H_0} \sin^2 \theta} \tag{6.29}$$

будет зависеть и от подмагничивающего поля, и от направления распространения.

Выражения для составляющих напряженности поля плоской волны, распространяющейся вдоль плоскости *zy* ($k_x = 0$), можно получить из (6.22) и (6.24):

$$\begin{gathered}
H_z = A e^{-jkR}; \\
H_y = \frac{k_y^2 - \gamma_z^2}{k_y k_z} A e^{-jkR}; \\
H_x = -j \frac{\omega^2 \varepsilon \mu_a \mu_0}{k^2 - \omega^2 \varepsilon \mu_r \mu_0} \frac{k_y^2 - \gamma_z^2}{k_y k_z} A e^{-jkR}; \\
E_z = -j \frac{k_y}{\omega \varepsilon} \frac{\omega^2 \mu_a \mu_0 \varepsilon}{k^2 - \omega^2 \varepsilon \mu_r \mu_0} \frac{k_y^2 - \gamma_z^2}{k_y k_z} A e^{-jkR}; \\
E_y = j \frac{k_z}{\omega \varepsilon} \frac{\omega^2 \mu_a \mu_0 \varepsilon}{k^2 - \omega^2 \varepsilon \mu_r \mu_0} \frac{k_y^2 - \gamma_z^2}{k_y k_z} A e^{-jkR}; \\
E_x = -\frac{\gamma_z}{\omega \varepsilon} \frac{\gamma_z}{k_y} A e^{-jkR},
\end{gathered} \tag{6.30}$$

где $\gamma_z = \omega \sqrt{\varepsilon \mu_z \mu_0}$.

Два решения (6.26) позволяют получить в безграничной ферритовой среде два класса плоских волн со сложной, но не поперечной структурой, определяемой формулами (6.30). Фазовые скорости (6.27) этих волн зави-

сят не только от направления распространения, но и от постоянного магнитного поля, намагниченности и частоты.

Имеющие широкое техническое применение в ферритах явления вращения плоскости поляризации, двойного лучепреломления, продольного и поперечного ферромагнитного резонанса являются характерными проявлениями особенностей распространения волн в гиротропной среде.

§ 6.3. Волны в продольно намагниченном феррите

Если электромагнитная волна распространяется в направлении, совпадающем с направлением намагничивающего поля, то $\theta = 0$ и согласно (6.26) волновые числа

$$k_{1,2}^2 = \omega^2 \varepsilon \mu_{эф\,1,2} \mu_0 = \omega^2 \varepsilon (\mu_r \pm \mu) \mu_0.$$

Следовательно, в данном случае эффективные магнитные проницаемости $\mu_{эф\,1,2} = \mu_r \pm \mu_a = \mu_\pm$.

Компоненты векторов поля (6.30)

$$\begin{gathered}
H_{y\,1,2} = H_{y\pm} = A_{1,2}\, e^{-jk_{1,2}z}; \\
H_{x\,1,2} = H_{x\pm} = \pm jA_{1,2}\, e^{-jk_{1,2}z}; \\
H_z = 0, \quad \dot{E}_z = 0; \\
E_{y\,1,2} = E_{y\pm} = \pm j \frac{k_{1,2}}{\omega\varepsilon} A_{1,2}\, e^{-jk_{1,2}z}; \\
E_{x\,1,2} = E_{x\pm} = \frac{k_{1,2}}{\omega\varepsilon} A_{1,2}\, e^{-jk_{1,2}z}.
\end{gathered} \tag{6.31}$$

Обе волны являются плоскими, поперечными и поляризованными по кругу в противоположных направлениях.

Волна любой поляризации, распространяющаяся вдоль постоянного магнитного поля, распадается на две волны с круговой поляризацией. Соотношение между их амплитудами зависит от условий возбуждения. В частности, если обе волны образуются из волны с линейной поляризацией, то амплитуды вторичных волн будут одинаковы. Тогда из (6.31)

$$H_{y\pm} = H_m e^{-jk_\pm z}, \quad E_{y\pm} = -z_\pm H_{x\pm},$$
$$H_{x\pm} = \pm jH_m e^{-jk_\pm z}, \quad E_{x\pm} = z_{x\pm} H_{y\pm},$$

где

$$z_\pm = \sqrt{\mu_\pm \mu_0 / (\varepsilon_r \varepsilon_0)} = \sqrt{(\mu_r \pm \mu_a)\mu_0 / (\varepsilon_r \varepsilon_0)}$$

— волновое сопротивление среды для право- и левополяризованных волн.

Мгновенные значения магнитных составляющих волны с правой поляризацией

$$h_{y+}(t) = H_m \sin(\omega t - k_+ z), \quad h_{x+}(t) = -H_m \cos(\omega t - k_+ z);$$

с левой поляризацией

$$h_{y-}(t) = H_m \sin(\omega t - k_- z), \quad h_{x-}(t) = H_m \cos(\omega t - k_- z).$$

Составляющие суммарной волны

$$h_y(t) = h_{y+} + h_{y-} = 2H_m \cos\frac{k_- - k_+}{2} z \sin\left(\omega t - \frac{k_- + k_+}{2} z\right);$$

$$h_x(t) = h_{x+} + h_{x-} = 2H_m \sin\frac{k_- - k_+}{2} z \sin\left(\omega t - \frac{k_- + k_+}{2} z\right),$$

напряженность результирующего магнитного поля волны

$$h(t) = \sqrt{h_x^2 + h_y^2} = 2H_m \sin\left(\omega t - \frac{k_- + k_+}{2} z\right).$$

Следовательно, по мере распространения волны вдоль оси z в феррите без потерь ее амплитуда $2H_m$ достается неизменной. Однако вследствие изменения соотношения между составляющими h_x и h_y вектор напряженности результирующего магнитного поля непрерывно поворачивается по мере распространения волны. Угол поворота плоскости поляризации волны относительно оси y

$$\psi = \operatorname{arctg}\left(-\frac{h_x}{h_y}\right) = \frac{k_- - k_+}{2} z. \tag{6.32}$$

Направление поворота плоскости поляризации оказывается не зависящим от направления волны и связано лишь с направлением постоянного магнитного поля. Рассмотренное явление поворота плоскости поляризации продольной волны называют *эффектом Фарадея.*

Вращение плоскости поляризации можно рассматривать как результат различия скоростей распространения компонент волн с правым и левым вращением ($v_+ \neq v_-$).

С помощью выражения (6.32) можно показать, что угол поворота (в градусах на единицу длины)

$$\psi / z \approx 180\sqrt{\varepsilon}\,\omega_M / (2\pi) = 360\gamma\sqrt{\varepsilon} M_0, \tag{6.33}$$

если эффект Фарадея рассматривать (как обычно и бывает на практике) в области слабых магнитных полей, где потери, связанные с ферромагнитным резонансом, малы. В этой области $\omega \gg \omega_H = \gamma\mu_0 H_0$; $\omega \gg \omega_M = \gamma\mu_0 M_0$. Из соотношения (6.33) следует, что угол поворота ψ

не зависит от частоты, мало зависит от диэлектрической проницаемости феррита и определяется главным образом намагниченностью феррита.

Направление вращения плоскости поляризации волны зависит только от направления намагниченности феррита. В области постоянных полей, меньших его резонансного значения, плоскость поляризации поворачивается по часовой стрелке относительно направления постоянного магнитного поля ($\psi > 0$). В области больших постоянных полей угол поворота ψ меняет знак.

§ 6.4. Волны в поперечно намагниченном феррите

Волновые числа двух волн, распространяющихся перпендикулярно постоянному полю, могут быть получены из (6.26) при $\theta = 90°$:

$$k_1^2 = \omega^2 \varepsilon\varepsilon_0 \mu_{эф1} \mu_0 = \omega^2 \varepsilon\varepsilon_0 \frac{\mu_r^2 - \mu_a^2}{\mu_r} \mu_0;$$

$$k_2^2 = \omega^2 \varepsilon\varepsilon_0 \mu_{эф2} \mu_0 = \omega^2 \varepsilon\varepsilon_0 \mu_z \mu_0.$$

В этом случае эффективные проницаемости

$$\mu_{эф1} = \frac{\mu_r^2 - \mu_a^2}{\mu} = \mu_1, \quad \mu_{эф2} = \mu_z.$$

Структура поля определяется уравнениями (6.30). Так, для первой волны

$$\dot{H}_{x1} = A_1 e^{-jk_1 y}, \quad \dot{H}_{y1} = -j \frac{\mu_a}{\mu_r} A_1 e^{-jk_1 y},$$

$$\dot{E}_{z1} = A_1 \sqrt{\frac{\mu_1 \mu_0}{\varepsilon\varepsilon_0}} e^{-jk_1 y},$$

для второй волны

$$\dot{H}_{z2} = A_2 e^{-jk_2 y}, \quad \dot{E}_{x2} = -A_2 \sqrt{\frac{\mu_z \mu_0}{\varepsilon_r \varepsilon_0}} e^{-jk_2 y}.$$

Первую из волн называют *необыкновенной*. Она плоская, но не поперечная, так как имеет составляющую магнитного поля H_y вдоль направления распространения, сдвинутую по фазе относительно поперечной составляющей H_x на 90°. С уменьшением намагниченности ($\mu_a/\mu_r \to 0$) продольная составляющая поля исчезает, и волна превращается в поперечную. Постоянная распространения необыкновенной волны зависит также от напряженности намагничивающего поля. Вторую волну назы-

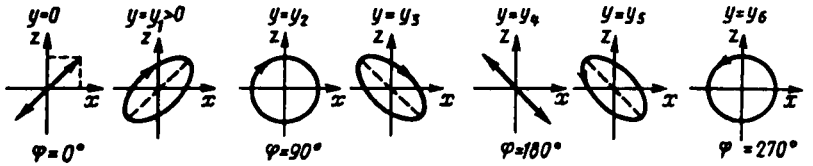


Рис. 6.2. Изменение поляризации по мере распространения поперечной волны в гиротропной среде

вают *обыкновенной*. Она чисто поперечная и имеет постоянную распространения, не зависящую от постоянного магнитного поля. Поэтому она никак не реагирует на гиротропные свойства среды и ведет себя так же, как волна в изотропной среде.

Волна произвольной поляризации, распространяющаяся перпендикулярно направлению постоянного поля, распадается в феррите на две волны. Амплитуды обыкновенной и необыкновенной волн одинаковы, если плоскость поляризации исходной волны наклонена под углом 45° к намагничивающему полю. Вследствие разности скоростей обыкновенной и необыкновенной волн меняется по мере распространения разность фаз между ними (φ) и поляризация суммарной волны (рис. 6.2). Преобразование линейной поляризации в эллиптическую в поперечно намагниченной среде носит название *эффекта Коттона—Мутона* или *эффекта Фохта.*

При одинаковых амплитудах обеих волн угол наклона плоскости поляризации составляет ±45°. Значение этого угла становится другим, если амплитуды обыкновенной и необыкновенной волн различны.

Одно из различий между рассмотренными волнами заключается в том, что только необыкновенной волне присуще явление поперечного ферромагнитного резонанса, сопровождающееся резонансным поглощением при резонансной частоте (6.29) $\omega_{\perp} = \omega_H \sqrt{1 + M_0/H_0}$ или при определенном магнитном поле

$$H_{\perp} = \sqrt{\frac{\omega^2}{\gamma^2 \mu_0^2} + \frac{M_0^2}{4}} - \frac{M_0}{2} = \sqrt{H_{\parallel}^2 + \frac{M_0^2}{4}} - \frac{M_0}{2}. \qquad (6.34)$$

Резонансная напряженность поля (6.34) всегда меньше, чем напряженность поля в случае продольного резонанса $H_{\parallel} = \omega/(\gamma\mu_0)$ при $\theta = 0$.

§ 6.5. Невзаимные явления в волноводах с ферритами

Волноводы, полностью заполненные ферритом, как правило, не применяются, так как они имеют большие потери. Кроме того, отражения

на границе феррита усложняют согласование и определяют узкополосность параметров. Этих недостатков можно избежать при частичном заполнении волновода. При этом сохраняется эффективное управление параметрами распространяющейся волны и появляется эффект невзаимной постоянной распространения, т. е. зависимость ее от направления распространения. В полностью заполненном волноводе и безграничной ферритовой среде это явление не имеет места. Невзаимные явления приводят к нарушению принципов взаимности и представляют большой практический интерес.

Рассмотрим в качестве примера прямоугольный волновод, содержащий поперечно намагниченный ферритовый стержень (рис. 6.3, *а*).

Несоблюдение в данной системе принципа взаимности позволяет использовать ее для создания различных вентильных систем. Если для прямой волны набег фазы на длине l обозначить через $\varphi_1 - \Delta\varphi_1/2$, то для той же волны, движущейся в обратном направлении, набег фазы будет $\varphi_1 + \Delta\varphi_1/2$. В этом и заключается явление *невзаимного фазового сдвига.* В отсутствие потерь невзаимный фазовый сдвиг определяется разностью волновых чисел прямой и обратной волн H_{10}:

$$\Delta k_1 = k_{1\text{пр}} - k_{1\text{обр}} = \frac{\Delta\varphi_1}{l} = \frac{-4\mu_a}{1+\mu_r}\frac{\pi}{a}\frac{\pi r^2}{ab}\sin^2\frac{2\pi t}{a}. \qquad (6.35)$$

Невзаимный фазовый сдвиг $\Delta\varphi_1$ пропорционален отношению объемов феррита и волновода $\pi r^2/(ab)$ и зависит от параметров феррита μ_a и μ_r. Для слабых полей можно показать, что получается линейная связь между невзаимным фазовым сдвигом и намагниченностью феррита. Поэтому перемена направления постоянного магнитного поля изменяет знак невзаимного фазового сдвига (знак перед μ_a). Характер зависимости

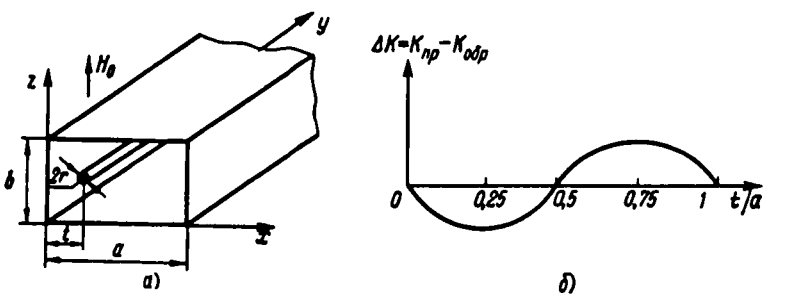


Рис. 6.3. Положение ферритового цилиндра в прямоугольном волноводе (*а*) и невзаимный фазовый сдвиг (*б*)

(6.35) от положения ферритового образца (рис. 6.3, *б*) объясняется структурой поля в волноводе. При расположении цилиндра в середине волновода ($t = a/2$) как у прямой, так и у обратной волн имеется только одна магнитная составляющая H_x. Поэтому свойства системы для этих волн одинаковы ($k_{1пр} = k_{1обр}$) и невзаимный фазовый сдвиг равен нулю. При перемещении пластины в левую половину сечения волновода поле волны в плоскости *xy*, перпендикулярной намагничивающему полю, становится эллиптически поляризованным, так как появляется составляющая H_y, сдвинутая на 90° относительно H_x. Различие знаков составляющих H_y у прямых и обратных волн приводит к различию направления вращения их плоскости поляризации относительно оси *z*. Вблизи $t = a/4$ компоненты H_x и H_y создают поле с поляризацией, наиболее близкой к круговой (относительно направления подмагничивающего поля). Прямая волна имеет левое вращение плоскости поляризации, а обратная — правое. Для них, как отмечалось, феррит имеет различные свойства, и невзаимный фазовый сдвиг получается наибольшим. При дальнейшем смещении цилиндра к боковой стенке волновода невзаимный фазовый сдвиг спадает до нуля по мере уменьшения составляющей поля H_x. Изменение знака невзаимного фазового сдвига на другой половине сечения волновода ($t > a/2$) связано с изменением направлений вращения магнитных полей прямой и обратной волн.

Для знакомства с *невзаимным лучепреломлением* в волноводе с ферритом рассмотрим одновременное распространение двух волн низших типов H_{10} и H_{01}. В целях большей наглядности в дальнейшем предполагаем, что ферритовый стержень расположен в точке $z = b/2$, где у волны типа H_{01} имеется лишь одна составляющая магнитного поля H_z и отсутствует невзаимный фазовый сдвиг, определяемый выражением (6.35). Тогда для волны типа H_{01} набег фазы φ_2 будет одним и тем же независимо от направления распространения ($\Delta\varphi_2 = 0$). Волны типов H_{10} и H_{01} могут возникнуть одновременно при распространении по волноводу волны с направлением электрического поля *E*, показанные на рис. 6.4.

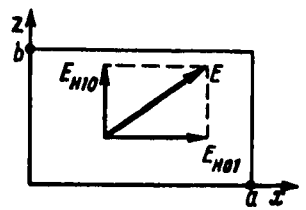


Рис. 6.4. Разложение волны с произвольно направленным поперечным полем *E* на волны типа H_{01} и H_{10}

В процессе распространения волн в прямом направлении ($y > 0$) между волнами типов H_{10} и H_{01} образуется непрерывно увеличивающийся фазовый сдвиг $\varphi_2 - (\varphi_1 - \Delta\varphi_1)$. При распространении волн в обратном направлении фазовый сдвиг $\varphi_2 - (\varphi_1 + \Delta\varphi_1)$ между ними связан с невзаимными свойствами волновода для первой волны. В общем случае

структура результирующей обратной волны сохраняется аналогичной структуре прямой волны, но период ее изменения будет другим. Однако при некоторых условиях структура результирующей обратной волны меняется.

В качестве примера на рис. 6.5, *а* изображено изменение поляризации результирующей прямой волны при условии, что в начальной плоскости ($y = 0$) волны обоих типов (H_{10} и H_{01}) совпадают по фазе и имеют одинаковые амплитуды. На рис. 6.5, *б* показана поляризация обратной волны при выполнении того же условия и при соблюдении равенства $\varphi_2 - (\varphi_1 + \Delta\varphi_1) = 0$. В этом случае отсутствует набег фазы между обратными волнами типов H_{10} и H_{01}, и результирующая волна остается линейно поляризованной с неизменным положением плоскости поляризации в отличие от эллиптически поляризованной результирующей прямой волны с непрерывным изменением направления поляризации.

Следовательно, в волноводе фазовые соотношения между двумя волнами (типов H_{10} и H_{01}) для прямого и обратного направлений неодинаковы и двойное лучепреломление в волноводе с ферритом оказывается невзаимным.

Невзаимность структуры поля прямых и обратных волн используется в вентиле на смещении поля. Для прямоугольного волновода такой вентиль представляет собой участок волновода с поперечно намагниченной пластиной, параметры которой подобраны таким образом, чтобы структура поля прямой волны сильно отличалась от структуры поля обратной волны (рис. 6.6, *а*). Если в месте концентрации поля обратной волны расположить проводящую пленку, то обратная волна будет сильно поглощаться. В то же время в прямом направлении система будет иметь малые потери.

Вентиль подобного типа выполняют в виде отрезка прямоугольного волновода, в котором размещена ферритовая пластина, покрытая с одной

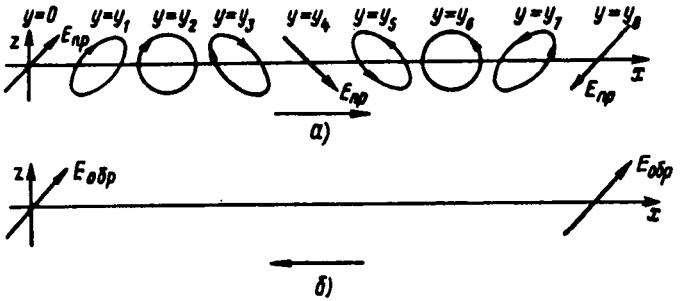


Рис. 6.5. Изменение поляризации прямой (*а*) и обратной (*б*) волн в отсутствии фазового сдвига между компонентами обратных волн типа H_{10} и H_{01}

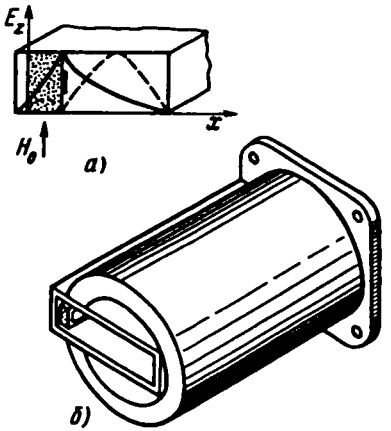


Рис. 6.6. Вентиль на смещении поля:

а — распределение электрического поля в поперечном сечении волновода (пунктирная линия — прямая волна, сплошная — обратная); *б* — расположение постоянного магнита

стороны поглощающим слоем. Надетый на волновод магнит создает требуемое поле в поперечном направлении (рис. 6.6, *б*). Обратная волна является волной поверхностного типа, если феррит несколько отодвинут от боковой стенки волновода. Необходимое для этого намагничивающее поле будет в несколько раз меньше поля поперечного ферромагнитного резонанса. Это значительно упрощает магнитную систему вентиля на смещении поля по сравнению с резонансным вентилем.

Остановимся на явлении *невзаимного поглощения* в волноводе с ферритом. Поглощение энергии в волноводе с ферритом определяется мнимой частью волнового числа $\underline{k} = k' - jk''$. Для цилиндрического ферритового стержня она может быть найдена как мнимая часть выражения (6.35) после подстановки в него комплексных значений компонент тензоров магнитной и диэлектрической проницаемостей.

Анализ полученных выражений показывает, что в общем случае поглощение оказывается неодинаковым для прямой *1* и обратной *2* волн (рис. 6.7) и достигает максимума при намагничивающем поле и частоте, соответствующим поперечному ферромагнитному резонансу. Наибольшее различие в затуханиях прямой и обратной волн достигается при расположени цилиндра на расстоянии $t \approx a/4$ от боковой стенки волновода в области круговой поляризации магнитного поля СВЧ.

Невзаимное поглощение при ферромагнитном резонансе лежит в основе работы резонансного вентиля. При полосковой конструкции резонансного вентиля (рис. 6.8) вращающееся магнитное поле в области ферритового образца создается с помощью двух параллельных разомкнутых на конце шлейфов длиной $\Lambda/8$ и $3\Lambda/8$. Входная прово-

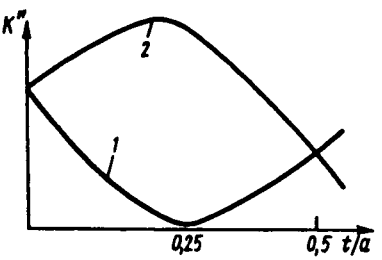


Рис. 6.7. Зависимость мнимой компоненты волнового числа от положения ферритового образца

димость первого шлейфа имеет чисто емкостный, а второго — индуктивный характер. Взаимная компенсация обеих реактивностей предотвращает появление отраженных волн в основной линии. Круговое вращение суммарного магнитного поля обеспечивается при условии, когда в области разветвления кроме обычного продольного тока I_z и поперечного магнитного поля H_x в основной линии появляется поперечный ток I_y и продольное относительно оси основной линии магнитное поле H_z, сдвинутое по фазе на 90° относительно H_x.

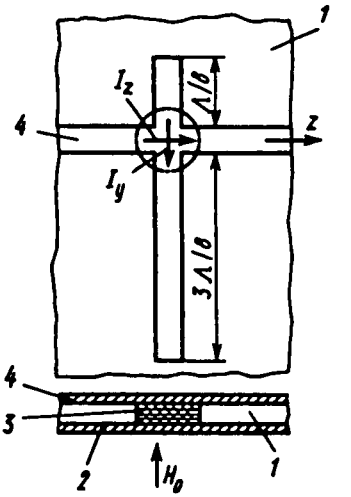


Рис. 6.8. Резонансный вентиль:

1 — диэлектрическая подложка; *2* — заземленное основание; *3* — ферритовый цилиндр; *4* — полосковый проводник

Полосковая конструкция трехплечного ферритового циркулятора выполняется на основе тройника в плоскости H, в центр которого помещен поперечно намагниченный ферритовый цилиндр, окруженный диэлектрической подложкой (рис. 6.9). Внутри ферритового цилиндра возбуждается стоячая волна. Силовые линии электрического поля перпендикулярны металлизированному основанию; силовые линии магнитного поля замыкаются за пределами ферритового диска. Это поле можно рассматривать как сумму двух волн, вращающихся навстречу друг другу вокруг оси цилиндра. При возбуждении плеча *1* и отсутствии постоянного магнитного поля H_0 плоскость узла напряжения ($E = 0$) повернута относительно этого плеча на 90° (рис. 6.10, *а*), а пучность располагается напротив плеча *1*. Поэтому энергия, поступающая из плеча *1*, делится поровну между плечами *2* и *3* (циркулярный эффект отсутствует).

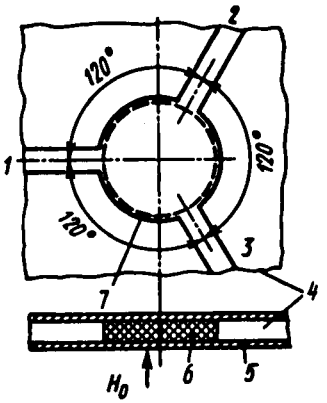


Рис. 6.9. Микрополосковый циркулятор:

1, 2, 3 — входные полосковые проводники; *4* — диэлектрическая подложка; *5* — заземленный экран; *6* — феррит; *7* — металлический диск

При наличии надлежащего поперечного поля H_0 вследствие различия магнитных проницаемостей и фазовых скоростей волн с правым и левым вращением плоскость $E = 0$ совмещается с центром плеча *3* (рис. 6.10, *б*). Это приводит к тому, что это плечо не возбуждается, и

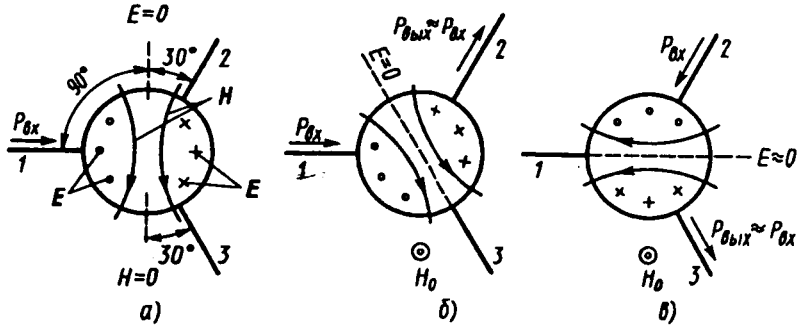


Рис. 6.10. Структура поля в феррите Y-циркулятора

энергия из плеча *1* передается только в плечо *2*. Если сигнал подан со стороны плеча *2*, то по тем же причинам энергия переходит только в плечо *3* (рис. 6.10,*в*).

В полосковой конструкции перестраиваемого фильтра используются две перекрещивающиеся линии, в месте пересечения которых помещена ферритовая сфера, намагничиваемая управляющим полем H_0 (рис. 6.11). Ферритовая сфера является резонатором, изготовляемым обычно из монокристалла иттриево-железного граната (ЖИГ). Собственная добротность такого резонатора около $3 \cdot 10^3$, а резонансная частота основного типа колебаний $\omega_0 = \gamma \mu_0 H_0$ определяется напряженностью поля подмагничивания H_0 и не зависит от размеров сферы. Поэтому резонатор может быть сделан очень малым. На частотах, не совпадающих с резонансной, резонатор ведет себя как изотропный образец и оказывает незначительное влияние на связанную с ним линию передачи. Лишь в узкой полосе частот вблизи частоты гиромагнитного резонанса связь резонатора с линией передачи резко увеличивается, так как появляется составляющая магнитной индукции, продольная относительно возбуждающей линии передачи (электрическое поле здесь минимально, поскольку точка пересечения расположена на расстоянии Λ/4 от разомкнутых концов полосковых линий).

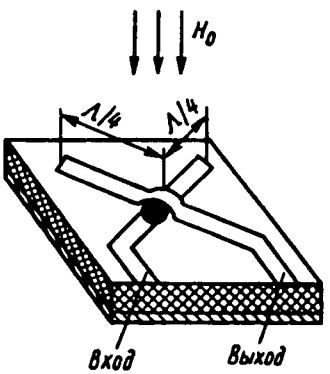


Рис. 6.11. Микрополосковый фильтр

Планарная конструкция и интегральная технология традиционных волноводных устройств с ферритами позволили существенно уменьшить их габариты и массу. Переход к магнитостатическим и

спиновым колебаниям дает возможность еще на несколько порядков уменьшить размеры ферритовых элементов сверхвысоких частот.

§ 6.6. Магнитостатистические и спиновые волны

Рассмотрение тензора магнитной проницаемости (6.13) и волн (6.30) в намагниченном феррите предполагало однородное намагничивание среды, не учитывало обменное и дипольное взаимодействия, существенные в намагниченных ферритовых образцах с размерами значительно меньшими, чем длина электромагнитной волны. Однако может существовать неоднородная прецессия намагничивания, зависимость которой от координат характеризуется длинами волн значительно меньшими, чем длина электромагнитной волны или размеры образца, когда условия на границах не имеют существенного значения и определяющим процессом является обменное взаимодействие. Такой вид прецессии намагниченности называют *спиновыми волнами*. Неоднородную прецессию, обусловленную главным образом дипольным взаимодействием или полем размагничивания в образцах ограниченных размеров, называют *магнитостатическими волнами*. Магнитостатические и спиновые волны относят к «медленным» волнам с очень малыми (по сравнению с «обычными» электромагнитными волнами) длинами волн. Они представляют собой волны прецессии магнитного момента электрона. Частоты собственных колебаний в ограниченных образцах (резонаторах) с неоднородной прецессией значительно ниже, чем частоты «обыкновенных» электромагнитных колебаний (в образцах с однородной прецессией намагниченности). Это дает возможность при исследовании «медленных» волн или «низкочастотных» колебаний пренебречь в уравнениях Максвелла током смещения, т. е. исходить из уравнения магнитостатики $\mathrm{rot}\,\mathbf{h} = 0$.

Рассмотрим сначала собственную прецессию, связанную со спиновыми волнами в изотропном однородном феррите без потерь ($\eta = 0$), намагниченном до насыщения M_0 полем H_0 в направлении оси z. Намагниченность (6.3) включает переменную составляющую

$$\dot{\mathbf{m}} = \dot{\mathbf{m}}_0 e^{j(\omega t - \mathbf{kR})}, \tag{6.36}$$

соответствующую бегущей спиновой волне с волновым вектором $\mathbf{k}$, а эффективное поле (6.6б)

$$\mathbf{H}_{эф} = \mathbf{H}_0 + \mathbf{h}_{обм} + \mathbf{h}_d \tag{6.37}$$

включает размагничивающее дипольное поле $\mathbf{h}_d$ и эффективное поле обменного взаимодействия $\mathbf{h}_{обм} = \mathbf{h}_\lambda + \mathbf{h}_q$. «Молекулярное» поле или поле Вейсса $\mathbf{h}_\lambda = -\lambda \mathbf{M}_0$, параллельное $\mathbf{M}_0$, можно не учитывать, как не даю-

щее вклада в уравнение (6.5) движения магнитного момента. «Неоднородный» член обменного взаимодействия

$$\mathbf{h}_q = q\nabla^2\mathbf{m} \text{ или } \mathbf{h}_q = -qk^2\dot{\mathbf{m}} = -qk^2\mathbf{m}_0\,e^{-j\mathbf{kR}}. \qquad (6.38)$$

Размагничивающее поле, связанное с изменением намагниченности в пространстве, должно удовлетворять уравнениям магнитостатики $\operatorname{rot}\mathbf{h}_d = 0$, $\operatorname{div}\mathbf{h}_d = -\operatorname{div}\mathbf{m}$. Откуда $\operatorname{grad}\operatorname{div}\mathbf{h}_d - \nabla^2\mathbf{h}_d = 0$, $\nabla^2\mathbf{h}_d = -\operatorname{grad}\operatorname{div}\mathbf{m}$.

Следовательно, для плоской волны

$$\mathbf{h}_d = -k^{-2}\mathbf{k}(\mathbf{m}_0\mathbf{k})e^{-j\mathbf{kR}}. \qquad (6.39)$$

Подставив уравнение (6.3) с учетом (6.36)—(6.39) в уравнение движения (6.5) и пренебрегая членами второго порядка малости, получим

$$-\frac{d\mathbf{m}}{dt} = \gamma\mu_0(\mathbf{M}_0\times\mathbf{h}_q + \mathbf{M}_0\times\mathbf{h}_d + \mathbf{m}\times\mathbf{H}_0).$$

Откуда, переходя к комплексной форме и проектируя полученное уравнение на оси координат, получим

$$\begin{gathered}\left(-j\omega + \omega_M\frac{k_xk_y}{k^2}\right)\dot{m}_{0x} + \left(\omega_H + \omega_{обм} + \omega_M\frac{k_y^2}{k^2}\right)\dot{m}_{0y} = 0;\\ \left(\omega_H + \omega_{обм} + \omega_M\frac{k_y^2}{k^2}\right)\dot{m}_{0x} + \left(j\omega + \omega_M\frac{k_xk_y}{k^2}\right)\dot{m}_{0y} = 0,\end{gathered} \qquad (6.40)$$

где

$$\omega_{обм} = \gamma q\mu_0k^2M_0 = qk^2\omega_M. \qquad (6.41)$$

Равенство нулю определителя системы (6.40) позволяет найти собственную частоту спиновой волны:

$$\omega^2 = (\omega_H + \omega_{обм})(\omega_H + \omega_{обм} + \omega_M\sin^2\theta), \qquad (6.42)$$

где $\theta = \arcsin\dfrac{\sqrt{k_x^2 + k_y^2}}{k}$ — угол между направлением распространения волны и направлением постоянного намагничивающего поля ($\mathbf{e}_z$).

Спектр спиновых волн в неограниченной среде показан на рис. 6.12 в виде зависимости собственной частоты спиновых волн от волнового числа, определяемой выражениями (6.41) и (6.42). Формулу (6.42) можно применять и для образцов ограниченных размеров, если они значительно больше длины спиновых волн $\lambda_c = 2\pi/k$, и в формуле (6.6а) принять $H_{эф} = H_{0i}$, т. е. исходить из внутреннего постоянного поля (учесть раз-

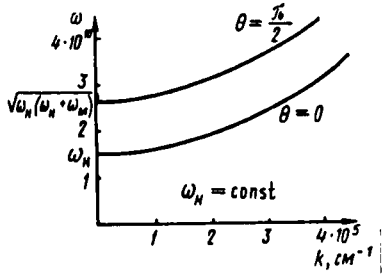


Рис. 6.12. Спектр спиновых волн в неограниченной среде

магничивание). Возбуждение спиновых волн в однородном высокочастотном поле может происходить при наличии различных магнитных неоднородностей феррита. Взаимодействие спиновых волн с тепловыми колебаниями решетки приводит к тому, что в феррите всегда возникает некоторое статистическое равновесие спиновых волн с различными $\mathbf{k}$ и θ.

В трехсантиметровом диапазоне для ЖИГ групповая скорость спиновых волн $1{,}6\cdot10^5$ см/с меньше скорости звуковых волн в том же кристалле и значительно меньше скорости магнитостатических волн. Малая групповая скорость спиновых волн используется при создании линий задержки.

Полученные дисперсионные соотношения для «медленных» электромагнитных волн ($k \gg \omega\sqrt{\varepsilon}/c$) показывают наличие области, для которой характерна ограниченность частоты с ростом k и даже практическая независимость ее от волнового числа. Из выражений (6.41) и (6.42) следует, что это происходит в том случае, когда роль обменного взаимодействия, обусловленная непараллельностью спинов, становится пренебрежимо малой ($\omega_{обм} \ll \omega_H$) и неоднородная прецессия определяется дипольным взаимодействием. Волны с такими свойствами и принято называть *магнитостатическими*. Согласно (6.42) их собственные частоты

$$\omega^2 = \omega_H(\omega_H + \omega_M \sin^2\theta). \qquad (6.43)$$

Эти волны, как следует из (6.14) и (6.43) при $\omega_r = 0$, существуют лишь в области отрицательных значений $\mu_r = -\mathrm{ctg}^2\,\theta$.

Таким образом, частота магнитостатических волн в неограниченной среде лежит в пределах

$$\omega_H \leqslant \omega \leqslant \sqrt{\omega_H(\omega_H + \omega_M)},$$

а волновые числа — в пределах от $\sim10^2$ до 10^5 см$^{-1}$. Нижний предел значений k следует из условия применимости магнитостатического приближения, а верхний — из условия пренебрежения энергией обменного взаимодействия. Приведенная оценка справедлива и для более сложных волновых процессов в ограниченных телах, если под k понимать величину, обратную расстоянию, на котором амплитуда и фаза переменной намагниченности существенно изменяются.

Реализация интегральных схем СВЧ на основе элементов с магнитостатическими волнами (МСВ) стала возможной благодаря быстрому

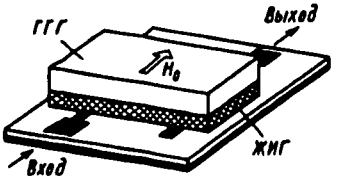


Рис. 6.13. Линия передачи на магнитостатических волнах

прогрессу технологии изготовления тонких пленок из ферромагнитных материалов с малыми потерями. В этом случае для магнитодиэлектрического слоя используется эпитаксиально выращенный слой ЖИГ на подложке из гадолиний-галлиевого граната (ГГГ). Предпосылкой для конструирования различных функциональных элементов интегральных схем (фильтров, линий задержки, усилителей и т. д.) является малая скорость распространения МСВ, высокая степень дисперсии, а также существенная зависимость параметров от направления и напряженности постоянного поля H_0. Формирование необходимого распределения поля H_0 возможно в виде полосок из пермаллоя, что обеспечивает конструктивную совместимость и магнитной системы с объемными интегральными схемами. Поэтому применение МСВ можно рассматривать как новое перспективное направление функциональной электроники СВЧ.

Важным достоинством устройств на МСВ является планарность конструкции и возможность их изготовления по хорошо отработанной интегральной технологии. На основе МСВ могут быть созданы магнитостатические функциональные аналоги всех ферритовых приборов СВЧ, но имеющие на порядок меньше размеры.

Проще всего возбуждать МСВ с помощью несимметричной полосковой линии, на поверхность которой нанесена слоистая структура ЖИГ-ГГГ (рис. 6.13). В зависимости от ориентации внешнего магнитного поля в пленке ЖИГ возникает три типа волны.

Поверхностная МСВ образуется, когда вектор поля смещения H_0 расположен в плоскости пленки перпендикулярно волновому вектору (рис. 6.14, *а*). Для нее характерно сосредоточение энергии поля вблизи одной из поверхностей ферритовой пластины и невзаимность структуры поля (при изменении направления фазовой скорости волна сосредоточена на противоположной поверхности пленки).

Прямая объемная МСВ образуется, когда вектор $\mathbf{H}_0$ направлен перпендикулярно плоскости пленки и волновому вектору (рис. 6.14, *б*). Она изотропна в плоскости пленки, может распространяться в любом направлении, обычно многомодовая, при этом наиболее легко возбуждаются и доминируют низшие типы волн (моды). Групповая скорость этих волн положительна (нормальная дисперсия: задержка сигнала возрастает с частотой).

Обратная объемная МСВ возникает, если вектор $\mathbf{H}_0$ расположен в плоскости пленки параллельно волновому вектору (рис. 6.14, *в*). Она характеризуется многомодовым распространением, при котором энергия распределяется в объеме ферритовой пластины с противоположно на-

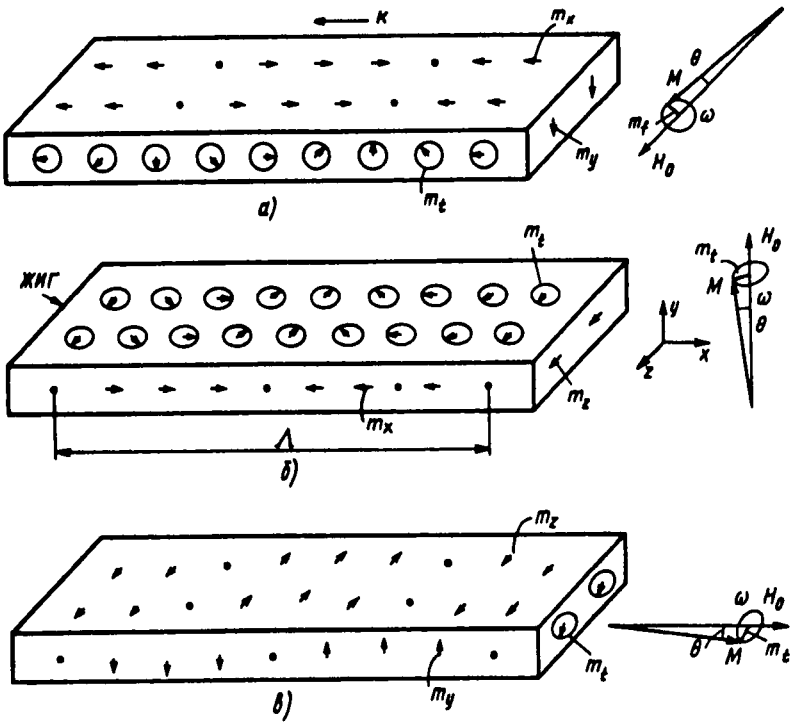


Рис. 6.14. Структура поля магнитостатических волн в ферритовой пленке

правленными фазовой и групповой скоростями (с убывающей зависимостью задержки сигнала от частоты).

Возможность управления дисперсией МСВ с помощью внешнего магнитного поля позволяет осуществить оптимальную аналоговую обработку информации непосредственно в СВЧ-диапазоне как во временно́й (линии задержки), так и в частотной (фильтры, резонаторы) области, а также путем изменения траектории движения волн. Схема одного из вариантов линии задержки приведена на рис. 6.13.

Схема однополосного резонатора на поверхностной МСВ показана на рис. 6.15. Она сформирована на пленке ЖИГ

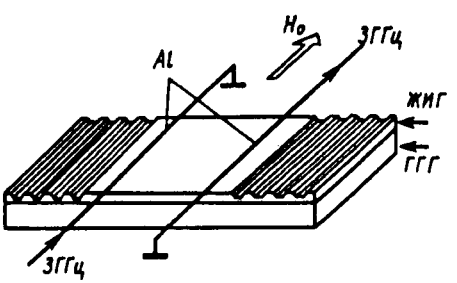


Рис. 6.15. Однополосный резонатор

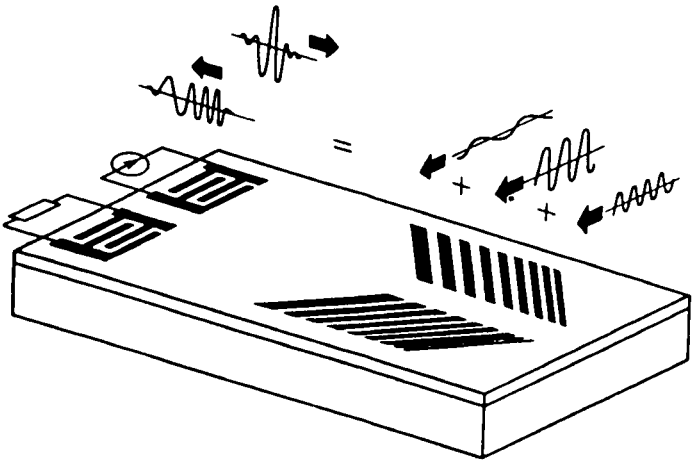


Рис. 6.16. Трансверсальный фильтр

толщиной 16 мкм. Две решетки образованы 50-ю бороздками глубиной 1,7 мкм, расстояние между которыми — 75 мкм, а между внутренними крайними обеих решеток — 1,5 мм. Расстояние между алюминиевыми микрополосковыми преобразователями (шириной 30 мкм и длиной 6 мм) — 1,2 мм. Нагруженная добротность при резонансной частоте 3 ГГц равна 600.

Структура трансверсального фильтра приведена на рис. 6.16. Такие фильтры позволяют получить импульсную характеристику (АЧХ) практически любой формы. Фильтр содержит последовательно соединенные элементы задержки, с каждого из которых отводится сигнал, умножаемый на заданный весовой коэффициент. Для этого используют отражение наклонно падающих МСВ от решетчатых структур, образованных металлизированными полосками или канавками. Весовой коэффициент преобразуемого сигнала определяют конфигурация и расположение решеток. В выходном преобразователе осуществляется суммирование отфильтрованных компонент сигнала. В рассматриваемом примере входной и выходной преобразователи имеют встречно-штыревую конструкцию.

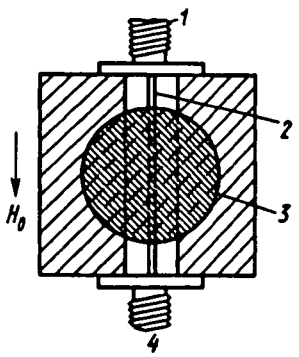


Рис. 6.17. Шумоподавитель

Нелинейные свойства линий МСВ позволяют применять их для увеличения отношения сигнал/шум. На рис. 6.17 линия передачи сигнала и шума выполнена в виде полосковой линии *2* (ширина проводника 25 мкм, толщина подложки 635 мкм). На нее

накладывается пленка *3* со структурой ЖИГ—ГГГ диаметром 21 мм и толщиной 36 мкм. Затраты энергии на возбуждение поверхностной МСВ в пленке ЖИГ приводят к затуханию сигнала между входом *1* и выходом *4*. Относительный уровень потерь снижается по мере роста уровня сигнала и достижения порога насыщения для МСВ. За счет этого отношение сигнал/шум увеличивается для сигналов, превышающих порог насыщения. Затухание слабых сигналов (ниже пороговых значений) составило около 35 дБ, а сильных (выше пороговых значений) — 10 дБ. Увеличение отношения сигнал/шум вблизи 4 ГГц равно 10 дБ. Разработаны также ограничители, генераторы и магнетометры на МСВ.

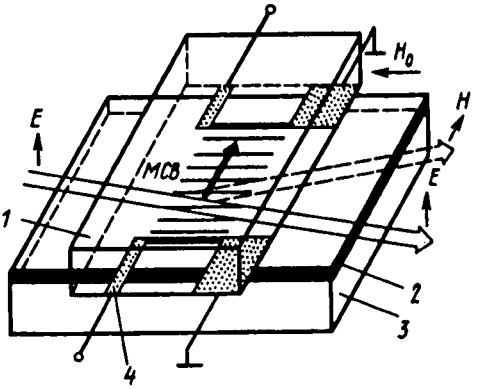


Рис. 6.18. Интегральный магнитооптический модулятор-переключатель

Применение МСВ позволяет повысить эффективность управления светом в интегральных магнитооптических устройствах и уменьшить их размеры. В структуре, изображенной на рис. 6.18, поверхностная МСВ возбуждается металлическими электродами *4*, нанесенными на пластинку из сапфира *1*, которая прижата к магнитооптической волноводной пленке *2* алюмоиттриевого граната. Волновод формируется на подложке *3* из ГГГ. Высокочастотное (от 1 до 20 ГГц) магнитное поле МСВ индуцирует в магнитооптической пленке изменения показателя преломления, формирующие бегущую дифракционную решетку, на которой происходит отклонение световых волн. Напряженность постоянного магнитного поля H_0 изменяется от 160 до 1800 А/см. При этом была получена дифракция волны 1,153 мкм типа *E* с шириной полосы модуляции 500 МГц.

Раздел третий

ЭЛЕКТРОМАГНИТНЫЕ ЯВЛЕНИЯ В НЕЛИНЕЙНЫХ СРЕДАХ И УСТРОЙСТВАХ

Глава 7
Электромагнитные явления в сегнетоэлектриках

Линейные среды характеризуются независимостью макроскопических параметров ε, μ и σ от интенсивности электромагнитного поля. В настоящем разделе рассматриваются среды, отличающиеся зависимостью макроскопических параметров от напряженности поля. Такие среды называют нелинейными. Они находят все большее применение в технике. К типичным нелинейным средам относят ферромагнетики и сегнетоэлектрики, обладающие сходными зависимостями проницаемостей от поля.

Физические процессы и явления, происходящие в природе, как правило, носят нелинейный характер и описываются нелинейными математическими моделями. Линейные модели получаются обычно при некоторых упрощающих предположениях и не дают полного описания явлений, а лишь указывают тенденции в развитии процессов.

Нетрудно убедиться, что в общем случае нелинейность должна больше проявляться в сильных полях. Для этого достаточно разложить, например, смещение D как функцию напряженности E в ряд:

$$D = (D)_{E=0} + \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{\partial^n D}{\partial E^n} \right)_{E=0} E^n.$$

Если ограничиться конечными значениями диэлектрической проницаемости ($D = 0$ при $E = 0$), то

$$\varepsilon = \left(\frac{\partial D}{\partial E} \right)_{E=0} + \sum_{n=2}^{\infty} \frac{1}{n!} \left(\frac{\partial^n D}{\partial E^n} \right)_{E=0} E^{n-1}.$$

Таким образом, диэлектрическая проницаемость ε не зависит от напряженности E в том случае, когда можно пренебречь всеми членами разложения, кроме первого. Тогда $\varepsilon = (\partial D / \partial E)_{E=0} = \text{const}$.

Для большинства сред такая линеаризация в обычных условиях оправдана. Если нельзя пренебречь вторым слагаемым разложения, то линеаризация может быть осуществлена только в пределах ограниченного участка нелинейной характеристики путем ее кусочно-линейной аппроксимации (рис. 7.1):

$$D \approx D_0 + E(\partial D/\partial E) = D_0 + \varepsilon_d E,$$

где D_0 — смещение, определяемое координатой точки пересечения с осью ординат касательной к кривой $D = f(E)$; $\partial D/\partial E$ — первая производная, принимаемая постоянной в пределах рассматриваемого участка.

Очевидно, что подобная замена действительной нелинейной среды с проницаемостью ε эквивалентной линейной средой с динамической проницаемостью $\varepsilon_d = \partial D/\partial E = \text{const}$ допустима только на определенном участке $E_1 < E < E_2$, ширина которого зависит от степени нелинейности параметров среды и допустимой погрешности в расчете.

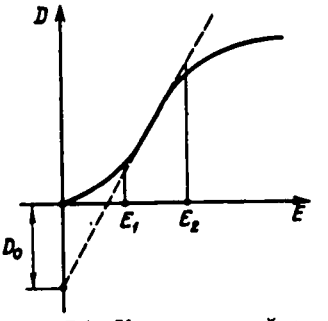


Рис. 7.1. Кусочно-линейная аппроксимация нелинейной среды на участке $E_1 < E < E_2$

Изложенный метод рассмотрения процессов в нелинейных средах особенно удобен в тех случаях, когда возможные непериодические и периодические изменения $\delta\varepsilon$ остаются малыми по сравнению с интервалом E_2—E_1. Подобная линеаризация была использована при анализе движения электромагнитной волны с амплитудой $h \ll H_0$ в феррите. Нелинейные свойства диэлектриков будут изложены на примере сегнетоэлектриков, у которых особенности таких сред проявляются наиболее отчетливо.

Постоянный член $D_0 = (D)_{E=0}$ существует только при наличии спонтанной поляризации в отсутствие внешнего электрического поля. Такие кристаллы называют *пироэлектрическими*. Пироэлектрическими могут быть кристаллы лишь тех групп симметрии, которые образуются одной осью и проходящими через нее плоскостями симметрии.

Спонтанная поляризация пироэлектриков очень мала по сравнению с молекулярными полями и в обычных условиях не создает внешнего электрического поля. Дело в том, что образец обладает хотя и малой, но не равной нулю проводимостью. Благодаря наличию внутри спонтанно поляризованного кристалла электрического поля появляются токи, образующие на поверхности тела свободные заряды, уравновешивающие поле в образце. Такую же роль играют ионы, оседающие на поверхность тела из воздуха. Поэтому пироэлектрические свойства можно наблюдать лишь при нагревании образца, когда его спонтанная поляризация меняется. Важно отметить, что пироэлектрики обладают спонтанной поляризацией во всем интервале температур вплоть до температуры плавления.

§ 7.1. Сегнетоэлектрики

Сегнетоэлектриками называют вещества, которые обладают самопроизвольным (спонтанным) электрическим моментом в отсутствие внешнего поля в некотором температурном интервале.

При нагревании до температуры (точки) Кюри сегнетоэлектрик испытывает фазовый переход, сопровождающийся изменением симметрии кристаллической решетки и исчезновением спонтанной поляризации.

Сегнетоэлектрики обладают следующими характерными особенностями: 1) резкой зависимостью диэлектрической проницаемости от напряженности поля; 2) высокими значениями относительной диэлектрической проницаемости, достигающими нескольких тысяч; 3) отставанием поляризации от приложенного напряжения, т. е. наличием *диэлектрического гистерезиса*; 4) резкой зависимостью диэлектрической проницаемости от температуры с аномально большим максимумом в точке Кюри.

Сегнетоэлектрики применяют в различных областях техники в качестве элементов, обладающих высокой диэлектрической проницаемостью и резко выраженными нелинейными и пьезоэлектрическими свойствами.

В электрических цепях с сосредоточенными параметрами сегнетоэлектрики используют в качестве составной части их нелинейных элементов — конденсаторов.

Керамические конденсаторы с резко выраженными нелинейными свойствами называют *варикондами*.

Налажено промышленное производство и *позисторов*. Так называют керамические терморезисторы с большим положительным температурным коэффициентом сопротивления (ПТКС). Они были открыты, когда начали исследовать электропроводность сегнетоэлектриков на основе титаната бария. В отличие от известных ранее терморезисторов с отрицательным температурным коэффициентом сопротивления (термисторов), равным 3...5%/К, позисторы в определенном интервале температур характеризуются значениями ПТКС, доходящими до 100%/К и более. При изменении температуры в пределах до 10 К сопротивление позистора может изменяться в миллионы раз. Температура в области резкого изменения сопротивления совпадает с точкой Кюри используемого сегнетоэлектрика и может изменяться в широких пределах.

Титанат бария проявляет свойства полупроводников при добавлении микроскопического количества примесей, причем лишь в узких пределах. Например, в случае добавления сурьмы при атомном содержании добавки ниже 0,2...0,3% относительно атомов бария сопротивление материала уменьшается обратно пропорционально концентрации добавки, а при концентрации сурьмы выше 0,3% удельное сопротивление материала резко возрастает.

Аномально большие ПТКС связаны с тем, что на поверхности зерен керамического полупроводникового титаната бария присутствуют в избытке кислород и примеси, захватывающие электроны из близлежащих слоев как акцепторы и образующие обедненный слой (потенциальный барьер). При температуре выше точки Кюри это приводит к высокому

сопротивлению керамики. При температуре ниже этой точки происходит спонтанная поляризация и возникает электрическое поле, под влиянием которого снижается высота потенциального барьера на границе зерен и сопротивление керамики.

§ 7.2. Сегнетоэлектрики в приборах СВЧ-диапазона

Сегнетоэлектрики представляют интерес для техники СВЧ главным образом благодаря тому, что их диэлектрическая проницаемость зависит от напряженности приложенного электрического поля. Это позволяет выполнить конденсатор с электрически управляемой емкостью, волновод с электрически управляемой фазовой скоростью, электрически перестраиваемые фильтры, переключатели, параметрически управляемые усилители. По сравнению с приборами на основе ферритов подобные устройства более экономичны по энергопотреблению, а по сравнению с полупроводниковыми устройствами СВЧ обладают большей технологичностью, электрической и радиационной стойкостью.

Возможность создания пригодных для широкого технического использования приборов СВЧ появилась после применения пленки сегнетоэлектрического материала, образованной непосредственно на диэлектрической подложке с высокой теплопроводностью.

Подобная конструкция позволила обеспечить малую зависимость диэлектрической проницаемости структуры от температуры в достаточно широком интервале изменения при сохранении заметной зависимости диэлектрической проницаемости от приложенного напряжения. Хороший тепловой контакт пленки с теплопроводящей подложкой особенно важен при использовании нелинейного элемента на высоком уровне мощности. В качестве диэлектриков с большой теплопроводностью применяют оксиды бериллия и магния, хотя предпочтительнее первый, поскольку его теплопроводность в три раза выше. Материалом подложки может быть также нитрид бора, сапфир (оксид алюминия) и алмаз.

Сегнетоэлектрики в диапазоне СВЧ в основном используют в параэлектрической фазе, поскольку при температурах несколько выше температуры фазового перехода диэлектрические потери существенно меньше*, чем при температурах ниже точки Кюри, а зависимость диэлектрической проницаемости от напряжения еще достаточно велика. В планарных варикондах на основе твердых растворов $Ba_xSr_{1-x}TiO_3$ можно смещать точку Кюри от 0 до 400 К, меняя содержание бария x.

В диапазоне СВЧ широкое применение получили сегнетокерамики на основе системы из титанатов бария и стронция с температурой Кюри –8° С.

---

* «Выключается» ориентационная поляризация.

Сегнетоэлектрические пленки изготовляют также из материалов типа ВК-8 и КН1-5. Толщина пленки обычно колеблется от 2 до 30 мкм. Сегнетокерамика обладает высокой механической и термической прочностью, радиационной и химической стойкостью, высокой плотностью и позволяет проводить металлизацию из различных металлов многими способами.

Сегнетоэлектрические элементы СВЧ можно разделить на сосредоточенные и распределенные. Основу большинства *сосредоточенных* элементов составляют планарные конденсаторы с пленкой сегнетоэлектрика. Выполнение активной области в виде тонкого слоя, неразрывно связанного с подложкой из теплопроводящего диэлектрика, заставляет отказаться от традиционного размещения обкладок конденсатора на двух противоположных поверхностях нелинейного материала и перейти к размещению обкладок на одной поверхности пленки. Во избежание зазоров между пленкой и электродами их наносят напылением или вжиганием. Планарные конденсаторы на основе сегнетоэлектрической пленки с диэлектрической подложкой получили название *планарных варикондов*. Конструкция планарных варикондов приспособлена к использованию при их изготовлении технологии интегральных схем СВЧ. Они легко включаются в состав как гибридных, так и монолитных интегральных схем на основе микрополосковых и щелевых (копланарных) линий передачи, а также в обычные полосковые линии. Планарные вариконды могут быть сформированы на подложке в технологическом цикле, едином для изготовления всей монолитной интегральной схемы.

Применение планарного вариконда возможно для электрической перестройки контуров, управления фазой сигнала с помощью перестраиваемых фильтров, электрического управления характеристиками периодических структур на основе линий передачи, перестройки частоты генераторов. Общими характерными чертами всех управляющих устройств СВЧ с сегнетоэлектриками являются взаимность (независимость характеристик от направления распространения сигнала и от полярности управляющего напряжения) и непрерывность регулировочных характеристик, основанная на непрерывности зависимости диэлектрической проницаемости от напряженности электрического поля. На рис. 7.2 в качестве примера показано расположение сегнетоэлектрической пленки на диэлектрической подложке в

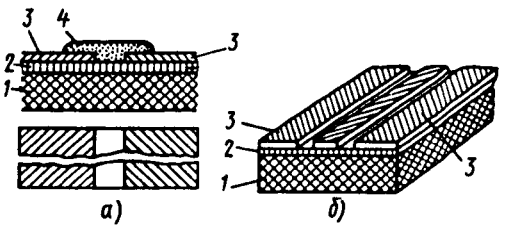


Рис. 7.2. Конструкции элементов СВЧ с сегнетоэлектрической пленкой:

1 — подложка; *2* — пленка сегнетоэлектрика; *3* — электроды; *4* — защитное покрытие

планарном конденсаторе (*а*) и копланарной линии (*б*).

Для практического использования *распределенных* структур на основе сегнетоэлектрической пленки в линиях передачи наибольший интерес представляют прямоугольные волноводы, щелевые линии и диэлектрические структуры с волнами поверхностного типа. К одному из подобных устройств относят волноводный фазовращатель, изображенный на рис. 7.3. Он состоит из отрезка прямоугольного волновода *2* с фазосдвигающей структурой, образованной сегнетоэлектрической пленкой *5* и диэлектрической подложкой *1* с повышенной теплопроводностью. На пленку нанесены металлические электроды *4*, к которым подводится управляющее напряжение через отрезок четвертьволновой коаксиальной линии *3*. Изменение управляющего напряжения приводит к изменению диэлектрической проницаемости сегнетоэлектрической пленки и фазы проходящей волны.

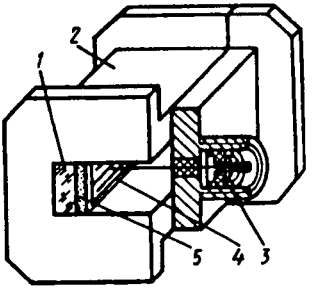


Рис. 7.3. Конструкция волноводного фазовращателя на сегнетоэлектрической пленке

Основные ограничения быстродействия сегнетоэлектрических приборов на СВЧ налагаются схемой управления, а не физическими явлениями в нелинейном материале (у сегнетоэлектриков в парафазе вплоть до частот $10^{11}...10^{12}$ Гц заметная дисперсия отсутствует). Например, даже в случае сегнетоэлектрической пластины с нанесенными на нее электродами в виде вставленных одна в другую гребенок емкость достигает 1000 пФ при управляющем напряжении до 200 В. Энергия переключения при этом составляет 20 мкДж, а время коммутации — 10...100 нс.

§ 7.3. Оптические свойства сегнетоэлектриков

Диэлектрические восприимчивость и проницаемость, как и форма и ориентация оптической индикатрисы (5.14), зависят не только от электрического поля, но и от механических напряжений. Следовательно, в общем случае изменение поляризационных констант (5.21) можно представить в виде

$$\Delta a_{ij} = k_{ij}T + r_{ijk}E_k + R_{ijkl}E_kE_l + p_{ijkl}\sigma_{kl}, \qquad (7.1)$$

где первый член характеризует температурные изменения поляризационных констант; k_{ij} — термический коэффициент показателей преломления (симметричный тензор второго ранга); второй и третий члены описывают электрооптические эффекты, причем тензор третьего ранга r_{ijk} характери-

зует линейный электрооптический эффект (эффект Поккельса); тензор четвертого ранга R_{ijkl} — квадратичный эффект (эффект Керра); четвертый член соответствует упругооптическому эффекту, т. е. изменению оптической индикатрисы под воздействием механических деформаций (p_{ijkl} — упругооптические постоянные).

Электрооптический эффект. Из уравнения (7.1) следует, что в точке Кюри при переходе из параэлектрической в сегнетоэлектрическую фазу поляризационные константы испытывают изменения, обусловленные возникновением спонтанной поляризации. Это явление называют *спонтанным электрооптическим эффектом.*

В сегнетоэлектриках, которые в параэлектрической фазе являются изотропными, после возникновения спонтанной поляризации вдоль оси [001] оптическая индикатриса приобретает вид эллипсоида вращения

$$(a_{11} + R_{12}^{*}P_z^2)(x^2 + y^2) + (a_{11} + R_{11}^{*}P_z^2)z^2 = 1.$$

Следовательно, при распространении луча вдоль направления поляризации (z) спонтанное двупреломление $\Delta n = \frac{1}{2} n_0^3 P_z^2 (R_{11}^{*} - R_{12}^{*})$ является квадратичным. В монокристаллах титаната бария $R_{11}^{*} - R_{12}^{*} = 0{,}15$ м^4/Кл^2.

В сегнетоэлектриках, которые в параэлектрической фазе анизотропны, первоначальная оптическая индикатриса (5.12) неравноосная. После возникновения спонтанной поляризации вдоль оси [100] уравнение индикатрисы приобретает вид

$$(a_{11} + R_{11}^{*}P_x^2)x^2 + (a_{22} + R_{12}^{*}P_x^2)y^2 + (a_{33} + R_{31}^{*}P_x^2)z^2 + 2r_{41}^{*}P_x yz = 1.$$

Таким образом, в сегнетоэлектриках этого типа в точке Кюри двупреломление возникает вследствие как линейного, так и квадратичного спонтанных электрооптических эффектов. Действительно, при этом поворот оптической индикатрисы на угол ξ вокруг оси [100]

$$\operatorname{tg}^2 \xi = \frac{2n_2^2 n_3^2 r_{41}^{*}}{n_3^2 - n_2^2} P_x$$

оказывается следствием линейного спонтанного электрооптического эффекта, а двупреломление при распространении луча по кристаллографическим осям

$$\Delta n_x = \frac{1}{2}(n_3^3 R_{31}^{*} - n_2^3 R_{21}^{*})P_x^2 = R_1^{*}P_x^2;$$

$$\Delta n_y = \frac{1}{2}(n_1^3 R_{11}^{*} - n_3^3 R_{31}^{*})P_x^2 = R_2^{*}P_x^2;$$

$$\Delta n_z = \frac{1}{2}(n_2^3 R_{21}^* - n_1^3 R_{11}^*) P_x^2 = R_3^* P_x$$

— следствием квадратичного спонтанного электрооптического эффекта ($n_1 = \sqrt{\varepsilon_x}$, $n_2 = \sqrt{\varepsilon_y}$, $n_3 = \sqrt{\varepsilon_z}$ — главные показатели преломления). В частности, в монокристаллах сегнетовой соли $R_1^* = 4{,}5$ м^4/Кл^2, $R_2^* = -10{,}8$ м^4/Кл^2, $R_3^* = 9{,}9$ м^4/Кл^2, $r_{41}^* = 6{,}3 \cdot 10^{-4}$ м^2/Кл.

При появлении внешнего электрического поля вдоль оси спонтанной поляризации P_c поляризация $P = P_c + P_{инд}$. Поэтому *индуцированный электрооптический эффект* можно рассматривать как продолжение спонтанного эффекта. Если приращение поляризационных констант вследствие спонтанного квадратичного эффекта вдоль i-го направления представить как $\Delta a_{ij} = R_{ij}^* P_j$, то во внешнем электрическом поле результирующее изменение поляризационных констант

$$\Delta a_{ij} = R_{ij}^* (P_c + P_{инд})^2 \approx R_{ij} P_c^2 + 2 P_{ij}^* P_c P_{инд} = \Delta a_c + \delta \Delta a_{ij}.$$

Следовательно, в данном случае индуцированный электрооптический эффект оказывается линейным, если пренебречь малым членом. Сравнив полученные приращения

$$\delta \Delta a_{ij} = 2 R_{ij}^* P_c P_{инд} = 2 R_{ij} \varepsilon_j \varepsilon_0 P_c E_j = r_{ij} E_j$$

с выражениями (5.21) нетрудно получить соотношение между индуцированным линейным электрооптическим эффектом сегнетоэлектрической фазы и спонтанным квадратичным эффектом $r_{ij} = 2 R_{ij}^* \varepsilon_j \varepsilon_0 P_c$.

Переход от монодоменной структуры сегнетоэлектрического кристалла к многодоменной сопровождается появлением гистерезисной петли индуцированного двупреломления (рис. 7.4).

До сих пор имелось в виду влияние на оптические свойства кристалла внешнего электрического поля, обычно постоянного или меняющегося с частотой, много меньшей частоты световой волны. При этом предполагалось, что показатели преломления не зависят от напряженности электрического поля световой волны. Это справедливо для излучения большинства традиционных источников света с напряженностью, пренебрежимо малой по сравне-

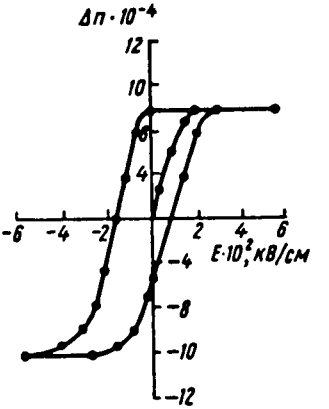


Рис. 7.4. Электрооптический гистерезис кристалла KDP при температуре –155,4 °C

нию с напряженностью внутреннего поля в среде (в атомах — примерно 10^9 В/см, в полупроводниках — около 10^7 В/см).

Появление квантовых генераторов и усилителей привело к широкому использованию световых волн с напряженностью до 10^8 В/см. При таких полях показатель преломления зависит от напряженности электрического поля световой волны, т. е. электрооптический эффект становится нелинейным.

Нелинейные оптические явления. Такие явления можно рассматривать как результат нелинейной зависимости поляризации среды от электрического поля. Обычно эту зависимость представляют с помощью разложения в ряд по степеням напряженности поля (вследствие малости оптических нелинейностей). Например, в чисто дипольном случае

$$\mathbf{P} = \chi^{(1)}\mathbf{E} + \chi^{(2)}\mathbf{EE} + \chi^{(3)}\mathbf{EEE} + \ldots, \qquad (7.2)$$

где первый член определяет обычную линейную восприимчивость, второй — нелинейную восприимчивость низшего порядка и т. д.

В общем случае

$$\mathbf{P} = \mathbf{P}_L + \mathbf{P}_{NL}, \qquad (7.3)$$

где с учетом запаздывания линейная компонента поляризации

$$\mathbf{P}_L = \int_0^\infty \chi^{(1)}(t')\mathbf{E}(t-t')\,\mathrm{d}t',$$

а нелинейная

$$\mathbf{P}_{NL} = \int_0^\infty \mathrm{d}t' \int_0^\infty \chi^{(2)}(t', t'')\mathbf{E}(t-t')\mathbf{E}(t-t'-t'')\,\mathrm{d}t'' +$$

$$+ \int_0^\infty \mathrm{d}t' \int_0^\infty \mathrm{d}t'' \int_0^\infty \chi^{(3)}(t', t'', t''')\mathbf{E}(t-t')\mathbf{E}(t-t'-t'')\mathbf{E}(t-t'-t''-t''')\,\mathrm{d}t''' + \ldots \qquad (7.4)$$

Если применить преобразование Фурье к выражению (7.4), то спектральная компонента поляризации на частоте ω_p

$$P_{NL\,i}(\omega_p) = \chi_{ijk}(\omega_p, \omega_q, \omega_r)E_j(\omega_q)E_k(\omega_r) +$$

$$+\chi_{ijkl}(\omega_p, \omega_q, \omega_r, \omega_s)E_j(\omega_q)E_k(\omega_r)E_s(\omega_s) + \ldots, \qquad (7.5)$$

где нелинейные восприимчивости различных порядков χ_{ijk}, χ_{ijkl} и т. п., как и линейная восприимчивость, определяются физическими свойствами материальной среды.

Из большого числа нелинейных оптических эффектов остановимся лишь на *генерации оптических гармоник*, тем более, что начало современного лазерного этапа развития нелинейной оптики обычно связывают с обнаруженной в 1961 г. генерацией второй гармоники рубиновым лазе-

ром в оптически прозрачных кристаллах. Появление второй гармоники можно рассматривать как следствие квадратичной компоненты в зависимости поляризации (7.2) от напряженности поля ($\sin^2 \omega t = \frac{1}{2} - \frac{1}{2}\cos 2\omega t$), точно так же третью гармонику — как следствие кубической зависимости ($\sin^3 \omega t = \frac{3}{4}\sin \omega t - \frac{1}{4}\sin 3\omega t$). Спектральная компонента поляризации (7.5) на частоте ω_p образована за счет комбинации двух полей с частотами ω_q и ω_r при учете младшего нелинейного члена в разложении (7.2) или трех (и более) полей при учете членов более высокого порядка. При этом могут возникать слагаемые не только с кратными гармониками, но и с суммарной или разностной частотой (биения между двумя световыми волнами), а также с нулевой частотой («выпрямление» или *детектирование* света за счет квадратичной нелинейности).

Очевидно, что наибольшего проявления нелинейных эффектов следует ожидать в веществах, обладающих большой поляризуемостью, в первую очередь в сегнето- и антисегнетоэлектрических кристаллах. Однако одной большой нелинейности среды оказывается недостаточно. Для получения значительных нелинейных эффектов необходимо выполнение *условия синхронизма*. Оно связано с тем, что мощная световая волна осуществляет временну́ю и пространственную модуляции поляризации и показателя преломления среды по закону бегущей волны.

В качестве примера рассмотрим процесс образования волны на удвоенной частоте. Если $n(\omega)$ — показатель преломления на частоте ω, c — фазовая скорость падающей волны в вакууме, то в кристалле фазовая скорость $v(\omega) = c/[n(\omega)] = \omega/k_1$. Фазовая скорость волны с частотой 2ω $v(2\omega) = c/[n(2\omega)] = 2\omega/k_2$, где $n(2\omega) = n(\omega) + \Delta n_\omega$ — показатель преломления на удвоенной частоте.

Источником волны с частотой 2ω является волна с частотой ω. Энергия волны с удвоенной частотой может увеличиваться только за счет энергии основной волны, так как других источников энергии нет. Поэтому при распространении в нелинейной среде не только фазы, но и амплитуды волн будут зависеть от координат:

$$\mathbf{E}_1(x,t) = \mathbf{E}_1(x)\sin[\omega t - k_1 x],$$

$$\mathbf{E}_2(x,t) = \mathbf{E}_2(x)\sin[2\omega t - k_2 x].$$

Исходными для расчета этих функций служат уравнения Максвелла, включающие нелинейную компоненту поляризации (7.2). Рассмотрим приближенное решение волновых уравнений в среде без потерь при ее слабой нелинейности, когда можно считать, что поляризация $P = \chi_1 E + \chi_2 E^2 = P_л + P_н$. В нашем случае нелинейная компонента

$$P_{\text{н}} = \chi_2 E^2 = \chi_2 E_1^2(x)\sin^2(\omega t - k_1 x) =$$

$$= \frac{1}{2}\chi_2 E_1^2[1 - \cos(2\omega t - 2k_1 x)] = P_0 + P_2.$$

Постоянная составляющая поляризации $P_0 = 0{,}5\chi_2 E_1^2$. Поляризация двойной частоты $P_2 = -P_0\cos(2\omega t - 2k_1 x)$ означает появление молекулярных диполей, осциллирующих и излучающих электромагнитные волны с удвоенной частотой и комплексной амплитудой $\dot{E}_2(x,t) = \dot{E}_2(x)\mathrm{e}^{j(2\omega t - k_2 x + \varphi)}$.

В изотропной среде без объемных зарядов поле второй гармоники должно удовлетворять волновому уравнению

$$\nabla^2 E - \mu\frac{\partial^2 D}{\partial t^2} = 0,$$

где $D = \varepsilon_0 E + P = \varepsilon_0 E + P_1 + P_0 + P_2 = \varepsilon E + P_0 + P_2$, причем $\varepsilon = \varepsilon_0 + \chi_1$. В случае однородной плоской волны, распространяющейся вдоль оси x, волновое уравнение приобретает вид

$$\frac{\partial^2 E_2}{\partial x^2} - \varepsilon\mu\frac{\partial^2 E_2}{\partial t^2} = \mu\frac{\partial^2 P_2}{\partial t^2}.$$

Для комплексной поляризации $\dot{P}_2(x,t) = -\dot{P}_0(x)\mathrm{e}^{j(2\omega t - 2k_1 x)}$ в условиях относительно малого изменения амплитуды $E_2(x)$ (на расстоянии, равном длине волны), когда $\partial^2\dot{E}_2/\partial t^2 \ll k_2\partial\dot{E}_2/\partial t$, волновое уравнение можно переписать в сокращенном виде

$$-j2k_2\frac{\partial\dot{E}_2}{\partial x} \approx 4\omega^2\mu P_0\,\mathrm{e}^{j(k_2 - 2k_1)x}.$$

Тогда

$$\dot{E}_2 \approx j\frac{2\omega^2\mu}{k_2}P_0\int_0^x \mathrm{e}^{j\Delta k x}\,\mathrm{d}x = j\frac{\chi_2 E_1^2}{n_2\Delta n}\mathrm{e}^{j\Delta k x/2}\sin\frac{\Delta k x}{2},$$

где

$$\Delta k = k_2 - 2k_1 = 2\frac{\omega}{c}(n_2 - n_1) = 2\frac{\omega}{c}\Delta n = \frac{4\pi}{\lambda}\Delta n.$$

Точное решение волнового уравнения должно учитывать связь между волнами с основной и удвоенной частотами, при которых происходит изменение амплитуды не только второй гармоники, но и первой. Это следует и из условия сохранения потока мощности в диэлектрике без потерь:

$$I = I_1 + I_2 = \frac{E_1 H_1}{2} + \frac{E_2 H_2}{2} = v_1\frac{\varepsilon_1 E_1^2}{2} + v_2\frac{\varepsilon_2 E_2^2}{2} = v_1\frac{\varepsilon_1 E_1(0)}{2} = \text{const},$$

откуда

$$E_1^2(x) = E_1^2(0) - \frac{n_2}{n_1} E_2^2(x),$$

если $E_2(0) = 0$. Результаты решения волнового уравнения приведены на рис. 7.5 для произвольных краевых условий.

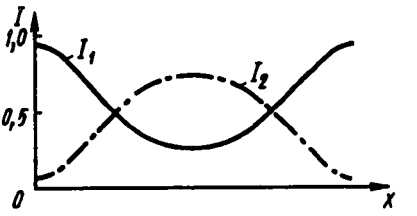


Рис. 7.5. Зависимость интенсивности света основной частоты и второй гармоники от пройденного пути

Таким образом, по мере распространения волн происходит периодическая передача энергии от волны с частотой ω волне с частотой 2ω и обратно. На поверхности напряженность $E_2(0)$ практически равна нулю, так как здесь поле определяется напряженностью падающей волны. Затем интенсивность $E_2^2(x)$ начинает расти. Вследствие различия фазовых скоростей $v(2\omega)$ и $v(\omega)$ характер взаимодействия волн по мере их распространения меняется: первоначальное усиление волны 2ω сменяется ее ослаблением. При этом энергия, накопленная сначала в поле этой волны, начинает переходить в энергию основной волны, пока снова не обратится в нуль.

Расстояние, на котором амплитуда второй гармоники достигает своего максимального значения

$$l_к = \frac{\lambda}{4\Delta n_\omega} = \frac{1}{4}\frac{\lambda}{n(2\omega) - n(\omega)} \qquad (7.6)$$

называется *когерентной длиной*. Длина $l_к$ определяет расстояние, на котором заканчивается передача энергии от основной волны, а энергия второй гармоники достигает максимума. Для кварца это расстояние очень мало (примерно 10^{-3} см) и энергия на частоте 2ω очень мала.

Один из способов увеличения энергии излучения второй гармоники заключается в увеличении энергии падающей волны. Однако и при использовании гигантских импульсов квантовых генераторов отношение E_2/E_1 остается еще очень малым. Увеличение этого отношения за счет значений σ и $\chi^{(2)}$ не эффективно, так как для известных пьезоэлектрических кристаллов эти значения не превышают их значений для KDP.

Наиболее результативной возможностью увеличения отношения E_2/E_1 является уменьшение разности $\Delta n_\omega = n(2\omega) - n(\omega)$, поскольку при этом возрастают когерентная длина (7.6) и путь, на котором передается энергия. Следовательно, наилучшие условия для передачи энергии возникают при $n(2\omega) = n(\omega)$ или

$$v(2\omega) = v(\omega), \qquad (7.7)$$

когда когерентная длина становится бесконечно большой, если не учитывать потери. При учете потерь когерентная длина имеет конечное значение, но при выполнении равенства (7.7) условия для передачи энергии по-прежнему остаются наилучшими. Расчет показывает, что при этом почти вся энергия падающей волны может перейти к волне с удвоенной частотой на длине, близкой к $l_к$. Если подобрать кристалл такой толщины, то на выходе из него энергия будет в основном сосредоточена в волне с частотой 2ω.

Условие равенства фазовых скоростей (7.7) и называют *условием синхронизма*. Если обе волны распространяются в одном направлении, то при $v(\omega) = \omega/k_1 = \omega_1/k_1 = 2\omega/k_2 = \omega_2/k_2 = v(2\omega)$ имеем:

$$2\omega_1 = \omega_2,\ \ 2\mathbf{k}_1 = \mathbf{k}_2,\ \ 2\hbar\omega_1 = \hbar\omega_2,\ \ 2\hbar\mathbf{k}_1 = \hbar\mathbf{k}_2. \qquad (7.8)$$

Следовательно, при возникновении волны с удвоенной частотой образование каждого кванта частотой 2ω происходит за счет энергии двух квантов частотой ω, т. е. «исчезают» два кванта $\hbar\omega_1$ и «появляется» один с энергией $\hbar\omega_2$. Процесс идет эффективно, если суммарный импульс двух исчезнувших квантов равен импульсу вновь образовавшегося кванта. Равенства (7.8) выражают законы сохранения энергии и импульса при образовании волн удвоенной частоты.

В изотропной среде практическое осуществление условий синхронизма $n(2\omega) = n(\omega)$ или $\varepsilon(2\omega) = \varepsilon(\omega)$ может иметь место только в том случае, когда одна из частот (ω или 2ω) находится в области поглощения. Поскольку при этом соответствующая волна будет быстро затухать, то изотропные среды малопригодны для генерации второй гармоники.

Более подходящими для этой цели являются те одноосные кристаллы, у которых волновые поверхности обыкновенных и необыкновенных волн для частот ω и 2ω пересекаются вне области поглощения (рис. 7.6). Например, в кристалле KDP в направлениях, образующих конус с углом $\theta_0 \approx 15°$ показатели $n_o(\omega)$ и $n_e(2\omega)$ совпадают. Если в кристалле KDP в этом направлении распространяется волна частотой ω с поляризацией, совпадающей с поляризацией обыкновенной волны, то для возникающей необыкновенной волны на частоте 2ω выполняется условие синхронизма. Поэтому когерентная длина достигает максимального значения. По сравнению с кварцем, в котором названные волновые поверхности не пересекаются, когерентная длина в KDP увеличи-

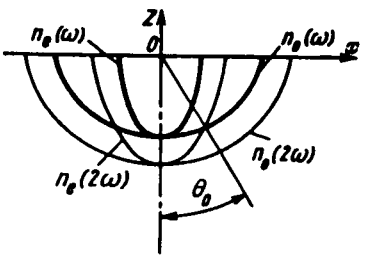


Рис. 7.6. Волновые поверхности обыкновенных и необыкновенных волн в кристалле KDP

вается более чем в 1000 раз, превышая 1 см. При этом удается преобразовать в энергию второй гармоники свыше 20% энергии падающей волны.

При облучении кристалла KDP рубиновым лазером (длина волны 694,3 нм) волна с частотой 2ω соответствует ультрафиолетовой области. При использовании газового лазера с длиной волны 1,058 мкм, приходящейся на инфракрасную область, вторая гармоника попадает в зеленый диапазон видимой области спектра.

При облучении кальцита и некоторых других кристаллов наблюдается генерация волны с утроенной частотой. Однако доля преобразуемой в третью гармонику энергии значительно меньше, чем при генерации второй гармоники, несмотря на выполнение условия синхронизма $\upsilon(3\omega) = \upsilon(\omega)$. Генерация волны с утроенной частотой — частный случай четырехволнового процесса с участием четырех разных фотонов. Подобные процессы идут наиболее эффективно, когда для фотонов выполняются условия $\omega_1 + \omega_2 + \omega_3 = \omega_4$, $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = \mathbf{k}_4$, выражающие законы сохранения энергии и импульса. При этом оказывается возможным как сложение волн и частот, так и их вычитание: $\omega_3 = \omega_1 \pm \omega_2$ или $\mathbf{k}_3 = \mathbf{k}_1 \pm \mathbf{k}_2$.

Разбиение на домены мешает оптическим применениям сегнетоэлектриков при температуре ниже точки Кюри. Поэтому многие кристаллы, в частности KDP и ADP, используют в параэлектрическом состоянии. Кристаллы ниобата лития $LiNbO_3$ могут быть получены и в однодоменном состоянии, поэтому их успешно применяют и при температурах ниже точки Кюри. Электрооптический эффект в этом кристалле на порядок выше, чем в KDP, ранее занимающем первое место среди кристаллов, в которых возможно выполнение условий синхронизма.

В лазерах ЛТИ-701 генерация второй гармоники обеспечивается введением в оптический резонатор кристалла иодата лития $LiIO_3$, достаточно стойкого к лазерному излучению. Кристалл располагается вблизи диэлектрического зеркала, прозрачного для второй гармоники и полностью отражающего на основной длине волны.

В заключение необходимо отметить, что под действием управляющего внешнего электрического поля в сегнетоэлектрике, являющемся источником излучения, может происходить модуляция как амплитуды, так и частоты когерентного излучения.

§ 7.4 Фотосегнетоэлектрические явления

Ранее было рассмотрено влияние на оптические свойства сегнетоэлектрика поляризации и фазового состояния сегнетоэлектриков. Оказывается, что во многих сегнетоэлектриках освещение может, в свою оче-

редь, изменять температуру Кюри, спонтанную поляризацию и другие макроскопические свойства. Эти явления связаны с наличием у подобных материалов полупроводниковых свойств. Первым таким материалом был титанат бария, обладающий слишком широкой запрещенной зоной (около 3 эВ). В настоящее время класс сегнетоэлектриков-полупроводников насчитывает большое количество соединений, среди которых имеются как широкозонные материалы (ниобаты лития, бария, стронция, соединения типа $A_2^VB_3^{VI}$), так и узкозонные (например, соединения типа $A^{IV}B^{VI}$). Фотоактивность этих соединений находит практическое применение в голографии и системах объемной оптической памяти. Кристаллы, изменяющие свои сегнетоэлектрические свойства при их оптическом возбуждении в собственной или примесной областях спектра, называют *фотосегнетоэлектриками.*

Вклад свободной энергии электронной подсистемы в свободную энергию решетки, взаимодействие электронов с сегнетоэлектрической модой приводят к ряду принципиально новых физических явлений в сегнетоэлектриках-полупроводниках. В частности, при их освещении в области собственной фоточувствительности происходит увеличение концентрации свободных электронов за счет неравновесных носителей заряда и носителей заряда в ловушках. Изменение распределения энергии между компонентами термодинамической системы сопровождается сдвигом точки Кюри и изменением физических свойств сегнетоэлектрика. Эти явления называют *фотосегнетоэлектрическими.* К ним относят сдвиг температуры Кюри, изменение спонтанной поляризации, температурного гистерезиса при увеличении концентрации электронов. Фазовый переход первого рода между пара- и сегнетоэлектрической фазами сопровождается скачком ширины запрещенной зоны. Ширина запрещенной зоны изменяется также под воздействием внешнего электрического поля. Особенно это заметно в параэлектрической области, где отсутствует спонтанная поляризация и происходит изменение только индуцированной поляризации, определяемой напряженностью электрического поля. Увеличивая ширину запрещенной зоны сегнетоэлектрика, поле соответственно *сдвигает край собственного оптического поглощения* в сторону коротких волн. Этот эффект в сегнетоэлектрике SbSI на порядок больше эффекта Келдыша—Франца (см. § 9.13). В сегнетоэлектрической области вдали от точки Кюри электроны не дают существенного вклада в свободную энергию.

Для сегнетоэлектрика ограниченных размеров условие конечности энергии системы требует полного экранирования спонтанной поляризации на поверхности. В идеальном диэлектрике экранирование спонтанной поляризации может происходить за счет внешней среды (адсорбция ионов и других заряженных частиц). В сегнетоэлектриках-полупровод-

никах экранирование осуществляется своими (внутренними) электронами и дырками. В связи с этим происходит приповерхностный изгиб энергетических зон, связанный с экранированием. В него вносят свой вклад и поверхностные электронные уровни. Наличие на поверхности сегнетоэлектрика слоя с особыми диэлектрическими свойствами («диэлектрического зазора») делает малоэффективным влияние на экранирование электродов, даже если они закорочены.

Следовательно, в реальном сегнетоэлектрике всегда существует внутреннее поле, значение и распределение которого определяются условиями экранирования спонтанной поляризации в объеме и на поверхности кристалла.

Фотосегнетоэлектрические явления можно подразделить на две основные группы. К первой относят явления, связанные с возбуждением неравновесных носителей заряда и оптической перезарядкой уровней в объеме сегнетоэлектрика, в котором отсутствует внутреннее поле. Ко второй группе относят явления, связанные с влиянием равновесных и неравновесных носителей заряда на экранирование спонтанной поляризации. Реально наблюдаемые фотосегнетоэлектрические явления могут быть связаны с двумя совершенно различными механизмами. Однако чаще всего влияние освещения на физические свойства сегнетоэлектрика обусловлено проявлением обоих механизмов.

Возможность существования монодоменного кристалла конечных размеров определяется наряду с другими условиями глубиной экранирования и концентрацией поверхности уровней. Благодаря этому, изменяя концентрацию носителей заряда в объеме и на поверхностных уровнях, можно превращать монодоменный кристалл в многодоменный, кристалл с *с*-доменом (поляризация перпендикулярна поверхности) — в кристалл с *а*-доменами (поляризация параллельна поверхности).

Зависимость доменной структуры от освещения и соответственно носителей заряда называют *фотодоменным эффектом.*

Одним из следствий влияния электронов на спонтанную поляризацию и фотодоменного эффекта является *фотодеформационный эффект*: влияние освещения на деформацию сегнетоэлектрика и, в частности, спонтанную деформацию при фазовом переходе.

При освещении короткозамкнутого сегнетоэлектрика возникает в направлении спонтанной поляризации стационарный ток, который называют *фотовольтаическим.* Напряженность электрического поля, образующегося вследствие заряда емкости кристалла фотовольтаическим током, в высокоомных сегнетоэлектриках-полупроводниках может достигать 10 кВ/см. В режиме разомкнутых электродов в этом направлении возникает фотонапряжение, которое в некоторых случаях доходит до 100 кВ, превышая на четыре порядка величину, соответствующую ширине запрещенной зоны. Возникновение стационарного фотовольтаического тока и

обусловленного им фотонапряжения называют *аномальным фотовольтаическим эффектом* в сегнетоэлектриках.

К фотосегнетоэлектрическим явлениям относится обнаруженный сравнительно недавно (1966 г.) эффект влияния интенсивного освещения на двупреломление в сегнетоэлектриках (на эффект Поккельса). Изменение двупреломления обусловлено изменением спонтанной поляризации под воздействием неравновесных носителей заряда или возбужденных электронных состояний, возникающих при освещении. Причиной изменения двупреломления могут быть также сильные электрические поля, возникающие при аномальном фотовольтаическом эффекте в высокоомном сегнетоэлектрике (например, в $LiNbO_3 : Fe$).

Эффект фотоиндуцированного изменения двупреломления иногда называют *оптическим искажением* или *фоторефрактивным эффектом*. Применение этого эффекта для записи объемных голограмм основано на локальном изменении двупреломления (главным образом за счет изменения показателя преломления необыкновенного луча) при освещении кристалла сфокусированный лучом лазера. Это изменение у ниобата и танталата лития достигает 10^{-3}, а время «памяти» эффекта колеблется от миллисекунд у титаната бария до месяцев у ниобата лития. Объемная модуляция показателя преломления обеспечивает высокое разрешение (до 10^3 лин/мм).

Основным условием заметного изменения двупреломления является достаточно высокая энергия излучения. Изменение двупреломления зависит от экспозиции. В ниобате лития при малых экспозициях (порядка 0,01 с) это изменение пропорционально произведению времени экспозиции на интенсивность излучения, а при длительных экспозициях (порядка нескольких минут) — пропорционально корню квадратному из этого произведения, причем при мощностях 20 мВт происходит насыщение.

Ниобат лития $LiNbO_3$ относят к кислородно-октаэдрическим сегнетоэлектрикам. Его валентная зона образована перекрытием кислородных орбиталей, а зона проводимости — орбиталей ионов ниобия. Эффект оптического искажения в ниобате лития имеет примесную природу. Например, в кристаллах, легированных железом, имеют место локальные уровни Fe^{2+} и Fe^{3+}, причем первые являются донорами, а вторые — ловушками. Оптическое искажение связывают с фотовозбуждением электронов с уровней Fe^{2+} в зону проводимости с последующим захватом ловушки Fe^{3+}. Поэтому чувствительность к оптическому искажению зависит от концентрации ловушек Fe^{3+} и не зависит от концентрации доноров Fe^{2+}.

Большое практическое значение имеет уменьшение дифракционной эффективности подобной голограммы, обусловленное ее разрушением в результате стирающего действия считывающего луча. Однако разрушение наблюдается только в том случае, когда исходная эффективность достигает максимального значения. Если эффективность меньше макси-

мального значения, то при считывании происходит сначала рост дифракционной эффективности, а затем ее спад. Рост эффективности объясняется тем, что в процессе считывания в объеме голограммы одновременно существуют восстановленная волна и волна нулевого порядка дифракции. Интерференция этих волн и продолжает запись голограммы в объеме кристалла (пока дифракционная эффективность не достигнет максимума).

Самоусиление (автографическая запись) может быть использовано в системах хранения информации. Например, если в объеме кристалла записать большое количество голограмм малой эффективности, то при последующем считывании наиболее часто и длительно «спрашиваемые» голограммы усилились бы в большей степени, что соответствовало бы закреплению наиболее необходимой информации. Таким образом, возникает возможность дифференцирования и селекции информации во всем записанном массиве по степени ее практической значимости.

§ 7.5. Обращение волнового фронта

Нелинейные явления в светочувствительной среде лежат в основе динамической регистрации волновых полей или динамической голографии. При этом среда реагирует на излучение непосредственно в момент его воздействия. Динамический эффект возникает, например, когда на кристалл ниобата лития V падают две плоские волны R и S, образующие стоячую волну с распределением интенсивности излучения $I(x)$, показанным на рис. 7.7. Под воздействием этой стоячей волны в ниобате лития возникает соответствующая модуляция показателя преломления. При этом максимумы распределения показателя преломления (обозначены на рисунке буквами z_1, z_2, z_3) оказываются смещенными относительно максимумов распределения интенсивности поля стоячей волны — вниз или вверх в зависимости от направления оптической оси кристалла C. При сдвиге на четверть периода максимумы интенсивности поля стоячей волны попадают на «склоны» периодического распределения показателя преломления. В результате условия прохождения волн R и S через кристалл становятся неравноправными. Одна из них (в данном случае волна R) отражается как от более плотной среды, а другая (волна S) — как от менее плотной. В такой ситуации одна из компонент отстает по фазе от другой, как бы сохраняющей первоначальную фазу. Та компонента, которая по-

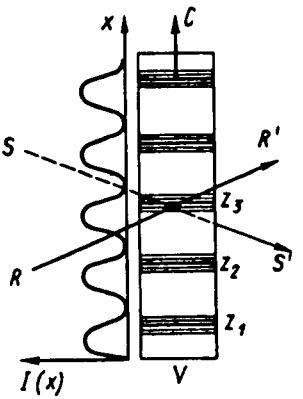


Рис. 7.7. Формирование динамической голограммы

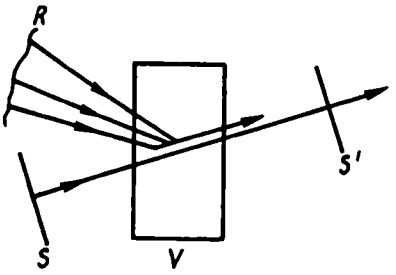


Рис. 7.8. Схема коррекции волнового фронта

теряла половину периода, сложится с другой с обратным знаком, и интенсивность поля соответствующей выходящей из кристалла суммарной волны уменьшится. Интенсивность поля другой покидающей кристалл волны, в которую входит компонента, не изменившая фазу при отражении от периодической структуры, соответственно увеличится. Поэтому по мере записи голограммы энергия выходящих из кристалла волн R' и S' начинает перераспределяться — волна R' усиливается, а волна S' ослабляется.

Рассмотренный эффект можно использовать для коррекции формы волновых фронтов лазерного излучения (рис. 7.8). При направлении на объемную светочувствительную нелинейную среду V излучения неправильной по форме интенсивной волны R лазера одновременно со слабой по интенсивности, но правильной по форме волной S, можно добиться того, чтобы энергия волны перешла полностью в волну S' с исправленным и усиленным волновым фронтом, в том случае, когда будет обеспечен на четверть периода сдвиг структуры голограммы относительно структуры показателя преломления. К сожалению, в этом методе коррекции волновых фронтов лазерного излучения трудно избежать нелинейных искажений.

Более удачным оказывается использование *обращения волновых фронтов* (рис. 7.9). В этом случае первичная слабая плоская волна A_1 проходит слева направо через оптически неоднородное тело усилителя У. На выход усилителя поступает усиленная волна сложной формы A_y, искаженная оптическими неоднородностями. Динамическая трехмерная голограмма $Г$ формируется в нелинейной среде вследствие смешения четырех волн. К первым трем относятся: волна A_y, плоская опорная волна R и распространяющаяся навстречу ей волна R_M, отраженная от зеркала M. Голограмму можно характеризовать изофазными поверхностями d_1, d_2, d_3 ... (в действительности возникают две голограммы: каждый из пучков накачки, взаимодейст-

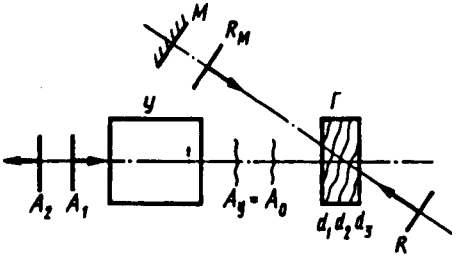


Рис. 7.9. Схема обращения волнового фронта динамической голограммой

вуя с пробным пучком, записывает свою голограмму, от которой при считывании отражается другой опорный пучок) . При считывании такой голограммы волной R_M формируется волна A_0 (четвертая), обращенная относительно волны, записанной на голограмме. Волна A_0 совпадает по форме с объектной волной, падающей на голограмму, но распространяется в противоположном направлении. После прохождения через усилитель в обратном направлении волна A_0 приобретает искажения противоположного знака и превращается в плоскую волну A_2, усиленную и распространяющуюся навстречу исходной волне A_1. В результате обращения все искажения волнового фронта, возникшие после первичного прохождения среды усилителя как бы «вывернутся наизнанку», исправляя идеальным образом фазовые неоднородности среды. В результате волновой фронт A_2 полностью восстановит волновой фронт A_1.

В установках термоядерного синтеза усиливаемым пучком A_1 может служить световая волна, рассеиваемая мишенью при ее импульсном лазерном нагреве.

Компенсация фазовых искажений в результате обращения волнового фронта используется также в интерферометре, работающем в режиме усреднения по времени, а также при изготовлении фотошаблонов для полупроводниковых интегральных схем. В последнем случае первичным усиливаемым пучком может служить слабая волна, просвечивающая шаблон. Устройство обращения позволяет не только устранить искажения, вносимые лазерным усилителем и спекл-эффектом, но и полностью отказаться от сложной проекционной оптики и необходимости физического контакта фотошаблона с подложкой.

Двухпроходная геометрия используется во многих случаях обращения волнового фронта в оптических системах (световой луч сопрягается по фазе и отражается так, чтобы пройти через ту же нелинейную среду второй раз). В этом случае двухпроходная геометрия может быть применена и для самонаведения излучения, например, при освещении «затравочным» лазером орбитального спутника та часть рассеянного им света, которая, возвращаясь в обращающую лазерную систему, может быть усилена и послана обратно на спутник. При этом обращающее зеркало способно не только компенсировать атмосферную турбулентность (вызывающую «мерцание» звезд), но и позволяет непрерывно удерживать на нем освещающий пучок. Усиленный лазерный пучок спутник может, в свою очередь, направить на ракету и вывести ее из строя, если интенсивность излучения будет для этого достаточна.

При четырехволновом смешении обращенный пучок можно считать результатом алгебраического перемножения трех входных пучков (пробного и двух встречных пучков накачки) в нелинейной среде. Поэтому, если один или все входные пучки кодированы путем изменения их пространственной структуры или их интенсивности во времени, то обращен-

ный луч будет кодирован таким образом, что его модуляция окажется пропорциональной произведению всех информационных модуляций взаимодействующих пучков. Такая система может быть использована для эффективной оптической связи со спутником. Ввод в обращенный пучок кодированного сообщения реализуется включением и выключением одного или нескольких входных пучков. Например, если маломощный лазер-бакен установлен на спутнике и направлен на находящуюся на Земле станцию с лазерным усилителем и обращающим кодирующим устройством, то система не только компенсирует турбулентность и отслеживает движение космического объекта, но и отправляет назад закодированную информацию.

Обращающее зеркало позволяет устранить искажения сигналов в оптическом волокне, обусловленные различием фазовых скоростей различных мод. Для этого достаточно искаженный обращенный волновой фронт сигнала пропустить через идентичный кусок волокна (рис. 7.10). Таким образом по световоду можно передавать с хорошим качеством и трехмерные изображения. В данном случае в обращающем зеркале происходит смешение трех входных световых пучков и выходящего пучка, которые имеют одинаковую несущую частоту, т. е. одинаковую длину волны. Такое смешение называют *вырожденным*.

Обращающее зеркало можно применять и для компенсации того уширения длительности импульсов в волоконно-оптических линиях, которое вызвано разными скоростями распространения различных частотных компонент импульса. При этом используется *почти вырожденное четырехволновое смещение*, когда частота падающего пробного пучка слегка сдвинута относительно частоты обоих пучков накачки. Здесь происходит непрерывное смещение интерференционной картины, наводимой пробным пучком и одним из встречных пучков накачки, создающее в среде взаимодействия движущееся обращающее волновой фронт зеркало. При отражении от зеркала второго пучка накачки его частота приобретает доплеровский сдвиг. Однако сдвиг частоты обращенного пучка отно-

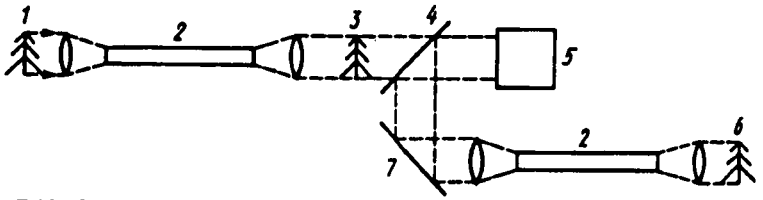


Рис. 7.10. Схема компенсации искажений в многомодовой волоконно-оптической линии:
1 — изображение объекта на входе; *2* — оптическое волокно; *3* — искаженное изображение объекта; *4* — полупрозрачное зеркало; *5* — обращающее зеркало; *6* — восстановленное изображение объекта; *7* — обычное зеркало

сительно частоты пучков накачки должен быть равен по значению, но противоположен по знаку сдвигу частоты пробного пучка (по закону сохранения энергии). В результате частотный спектр пробного пучка симметрично инвертируется. Если в середине волоконно-оптической линии расположить такое зеркало, то после попадания на него расширившегося импульса и обращения волнового фронта произойдет изменение порядка следования частотных компонент импульса: имеющие меньшую скорость компоненты покинут зеркало раньше, чем имеющие бóльшую скорость. Восстановление импульса путем сужения его до исходного состояния будет обеспечено за счет того, что на второй половине трассы быстрые компоненты догонят медленные.

Фотосегнетоэлектрические явления лежат в основе работы и многих типов оптически управляемых модуляторов.

§ 7.6. Оптически управляемый пространственно-временнóй модулятор

Одним из модуляторов для оперативной оптической обработки информации является многослойная структура, использующая монокристалл силикоселенита $Bi_{12}SiO_{20}$. Он является как электрооптическим, так и фотопроводящим. Темновое удельное объемное сопротивление кристалла больше 10^{13} Ом·см, а кратность фототока превышает 10^4.

Кристаллы силикоселенита, как и других соединений (ZnS, ZnSe, $Bi_{12}GeO_{20}$), пригодных для этого устройства, имеют кубическую структуру без центра инверсии, обладающую оптической изотропией в отсутствие внешних воздействий и оптической активностью. Отсутствие анизотропии является их важным преимуществом, поскольку при осуществлении оптимальной модуляции не требуется компенсации естественного двулучепреломления. Если к кристаллу приложить электрическое поле $\mathbf{E}_0$ вдоль оси [001], то для света, распространяющегося вдоль этого направления, новые главные показатели преломления

$$n_x = n_0 - \frac{1}{2} n_0^3 r_{41} E_0;$$

$$n_y = n_0 + \frac{1}{2} n_0^3 r_{41} E_0,$$

где n_0 — показатель преломления изотропного кристалла.

Поэтому на выходе из кристалла толщиной d разность фаз обыкновенного и необыкновенного лучей

$$\varphi = 2\pi \frac{n_y - n_x}{\lambda} d = \frac{2\pi}{\lambda} n_0^3 r_{41} U,$$

где U — приложенное напряжение.

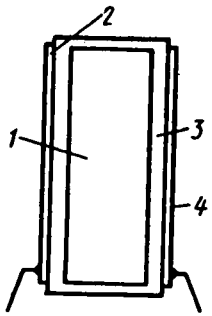


Рис. 7.11. Схема пространственно-временно́го модулятора с кристаллом силикоселенита

Модулятор (рис. 7.11) состоит из тонкой пластины электрооптического кристалла *1* (нормаль к поверхности совпадает с осью [001]), на которую наносят диэлектрические покрытия *3* (парилен, получающийся после полимеризации параксилена), и двух электродов (например, из SnO, InO, Pt). Один из них прозрачный *4*, а второй *2* либо прозрачный, либо отражающий. Последнее предпочтительнее, так как двукратное прохождение света позволяет в два раза уменьшить приложенное напряжение по сравнению с полуволновым (39 кВ) и в то же время исключают влияние оптической активности кристалла. Толщину слоев находят из условий электрической прочности и наибольшего светопропускания. Толщина диэлектрического покрытия обычно 3...10 мкм, кристаллической пластины — 100...200 мкм.

Этапы работы модулятора, называемого PROM (Pockels Readont Optical Modulator), изображены на рис. 7.12. Из первоначального распределения потенциала (показано на рис. 7.12, *а* толстой линией) видно, что бóльшая часть напряжения внешнего источника приложена к кристаллу. Затем кристалл освещается стирающим светом (равномерное освещение) от разряда ксеноновой лампы. Генерируемые при этом свободные носители заряда дрейфуют по кристаллу и компенсируют поле внутри него до нуля (рис. 7.12, *б*); весь перепад потенциала приходится теперь только на париленовые слои. После этого сначала закорачивают внешние электроды (рис. 7.12, *в*), а затем проводят переполяризацию путем изменения полярности приложенного напряжения (рис. 7.12, *г*). Благодаря этому падение напряжения на кристалле становится почти вдвое больше напряжения источника ($2U_0 = 0{,}5U_{\lambda/2}$).

Для записи кристалл освещается светом, модулированным записываемым изображением. При этом используется синий свет с энергией квантов, соответствующей собственному поглощению полупроводника. На освещенных участках генерируются свободные носители заряда, которые дрейфуют во внешнем электрическом поле и концентрируются в ловушках (уровнях захвата) на границе раздела кристалла и диэлектрика, экранируя внешнее поле. Это вызывает уменьшение напряжения на освещенной поверхности кристалла.

Напряжение на кристалле определяется экспозицией (чтобы снизить напряжение на кристалле в *e* раз, требуется плотность энергии порядка 5 мкДж/см^2). Благодаря этому распределение яркости в записываемом изображении преобразуется в соответствующее распределение двулучепреломления и фазового сдвига.

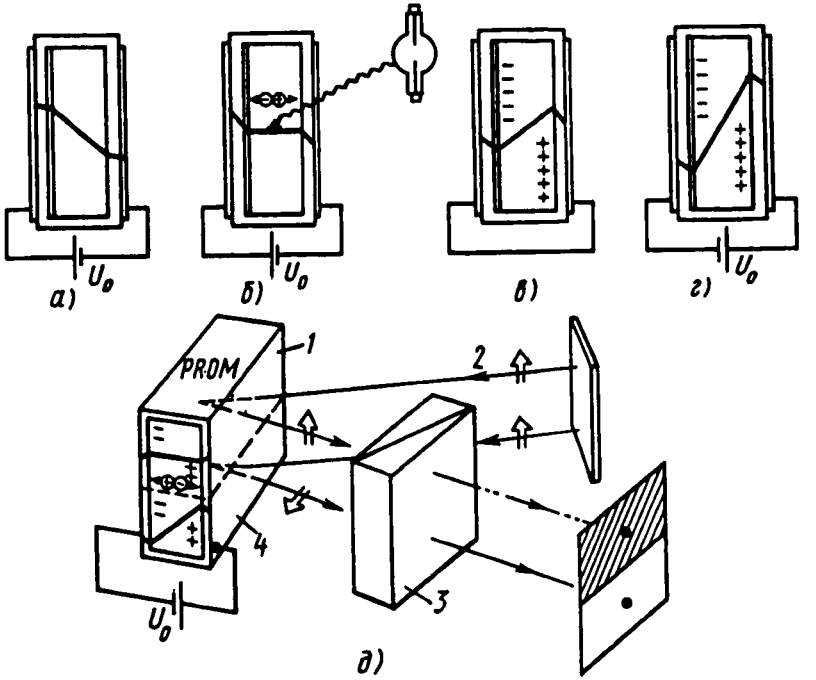


Рис. 7.12. Этапы работы системы:

а — первоначальная поляризация; *б* — стирание записи разрядом ксеноновой лампы; *в* — закорачивание электродов; *г* — переполяризация, *д* — считывание; *1* — область, освещенная при записи; *2* — считывающий луч с вертикальной поляризацией; *3* — анализатор; *4* — область, неосвещенная при записи

Считывание методом отражения через скрещенные поляризатор и анализатор (рис. 7.12, *д*) дает негативное изображение относительно записывающего света. Обычно считывают лучом гелий-неонового лазера (633 нм). В этой области кристалл также фоточувствителен, но носителей заряда генерируется незначительное количество (чувствительность в 10^4 раз меньше, чем к свету гелий-кадмиевого лазера с $\lambda = 440$ нм). Это позволяет при считывании пропускать через прибор значительно бóльшую световую энергию, чем требуется для записи, причем без снижения качества записанного изображения.

Изменяя внешнее напряжение, можно менять контраст считываемого изображения. Таким образом можно проводить операцию, называемую *вычитанием*. Если, например, записанное изображение имеет низкий контраст, то для его усиления постоянное напряжение нужно вычесть из напряжения на кристалле. Если приложить при считывании положительное напряжение, большее максимального напряжения на кристалле, то

можно получить позитивное изображение. Если же оно не достигнет максимума, то все точки кристалла с этим напряжением станут черными, образуя контур постоянной интенсивности. Стирание записанного изображения осуществляется равномерным освещением кристалла синим светом при замкнутых электродах.

Время хранения записанной информации определяется временем релаксации слоев пространственных зарядов, оно может составлять несколько часов. К разрушению записанного изображения приводит и поверхностная проводимость кристалла, поскольку происходит деградация потенциального рельефа.

Время однократной записи, связанное с подвижностью, приблизительно равно 10^{-5} с. Однако частота циклов «запись—стирание» ограничивается энергией, выделяющейся в приборе за один цикл. Она определяется диэлектрическими потерями ($\operatorname{tg}\delta = 0{,}013$) и плотностью энергии разделенного заряда на поверхности кристалла-диэлектрика (около $4\cdot10^{-4}$ Дж/см^2). Это дает возможность осуществить при комнатной температуре только 10 циклов в секунду.

Приборы PROM могут быть использованы в качестве преобразователей изображения из некогерентного (например с ЭЛТ) в когерентное, в качестве управляемых пространственных фильтров, формирователей страниц голографического запоминающего устройства, оптических усилителей, элементов оптических корреляторов и для таких оптических операций, как инвертирование, управление контрастом, выделение эквидистант и т. д.

§ 7.7. Самовоздействие световых импульсов

Нелинейность среды проявляется в зависимости комплексного показателя преломления $\underline{N} = n' - jn''$ от интенсивности света $I = E^2/(2z_c) = a_0^2/(2z_c)$. Самовоздействие световых импульсов заключается в том, что мощная световая волна сама определяет в общем случае как закон дисперсии фазовой скорости $\upsilon = c/n'$, так и коэффициент затухания $\alpha = \omega n''/c$.

Ограничимся эффектами, обусловленными нелинейностью действительной части показателя преломления, т. е. *дисперсионным самовоздействием*, представляющим наибольший интерес в оптике ультракоротких импульсов.

В изотропных средах

$$n' = n_1(\omega) + n_2' I = n_1(\omega) + n_2' a_0^2/2 = n_1(\omega) + b a_0^2/\beta_0^2.$$

Коэффициент фазы в среде с линейной дисперсией ($k_2 = \partial^2\beta/\partial\omega^2 = 0$)

$$\beta(\omega, I) = \omega n'/c = \beta_1(\omega) + \beta_2(I) \simeq \beta(\omega_0) + \left(\frac{\partial\beta}{\partial t}\right)_{\omega=\omega_0} (\omega - \omega_0) + ba_0^2.$$

В этом случае уравнение (3.67) приобретает вид

$$\frac{\partial \dot{A}_0(z, \omega - \omega_0)}{\partial z} = -j\left[\left(\frac{\partial\beta}{\partial\beta}\right)_{\omega_0} (\omega - \omega_0) + ba_0^2\right]\dot{A}_0(z, \omega - \omega_0).$$

После обратного преобразования Фурье и перехода к системе координат, движущейся с групповой скоростью $u = (\partial\omega/\partial\beta)_{\omega_0}$, получим

$$\frac{\partial \dot{a}_0(\eta, z)}{\partial z} = -jba_0^2\dot{a}_0(\eta, z).$$

Решение этого уравнения имеет вид

$$\dot{a}_0(\eta, z) = \dot{a}_0(\eta, 0)\,e^{-jba_0^2 z} = a_0(\eta)\,e^{-jba_0^2(\eta)}.$$

Отсюда следует, что огибающая импульса распространяется без изменения формы с групповой скоростью u в то время, как фаза импульса $\varphi = -bza_0^2(\eta) = -\beta_0 n_2' z I_0(\eta) = -\dfrac{2\pi}{\lambda} n_2' z I_0(z)$ уменьшается пропорционально пройденному расстоянию и интенсивности $I_0(\eta)$. Максимальный сдвиг фаз $\varphi_{max} = -bza_0(0) = -b\rho_0^2 z$.

Следовательно, в нелинейной среде возникает *фазовая самомодуляция* и уширение спектра импульса. Для гауссовского импульса максимальное симметричное смещение частоты

$$\delta\omega_{max} = \left(\frac{\partial\omega}{\partial\eta}\right)_{max} = -bz\left(\frac{\partial a_0^2}{\partial\eta}\right)_{max} = 0{,}43\varphi_{max}\Delta\omega_0,$$

если не учитывать нелинейную релаксацию (в волоконных световодах время релаксации не превышает 10^{-14} с).

Фазовая самомодуляция в среде с практически безинерционной электронной нелинейностью относится к наиболее удобному на сегодняшний день методу создания сверхбыстрой модуляции коротких импульсов. Возможно использование для этих целей одномодовых световодов. Применение прозрачных световодов значительной длины со стабильной геометрией позволяет осуществить устойчивую нелинейную модуляцию в большом интервале частот при пиковой мощности в единицы ватт.

Проявления нелинейности среды существенно возрастают, если возникает оптическая обратная связь, реализуемая, в частности, в интерферометре Фабри—Перо. При этом возникает *эффект бистабильности.*

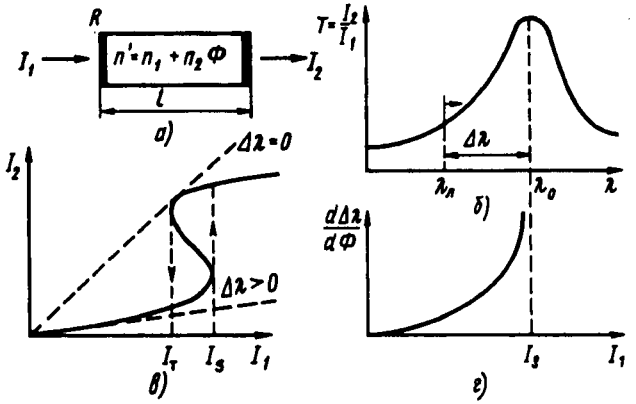


Рис. 7.13. Формирование оптической бистабильности в нелинейном резонаторе

При пропускании монохроматического светового луча через плоскопараллельный нелинейный образец (эталон Фабри—Перо), образующий резонатор (рис. 7.13, *а*), с ростом интенсивности света Ф происходит уменьшение оптической длины резонатора $(n_1 + n_2\Phi)l$ и затягивание в резонанс с внешним излучением (рис. 7.13, *б*).

Пусть сначала при малой интенсивности света на входе оптическая длина резонатора имеет значение, соответствующее области между пиками кривой Эйри (см. рис. 1.5), т. е. пропускание мало. По мере увеличения интенсивности пучка I_1 оптическая длина и порядок интерференции постоянно меняются, а пропускание медленно возрастает. При этом интенсивность на выходе увеличивается по закону, близкому к линейному, определяя малый начальный наклон прямой I_2 (рис. 7.13, *в*). В определенный момент за счет существенного увеличения положительной обратной связи между показателем преломления и оптической длиной эти параметры начнут возрастать, различие между фазами прямых и обратных волн будет все меньше, а интенсивности I_2 и Ф будут расти все быстрее. При некотором критическом значении $I_1 = I_s$ эффект обратной связи резко возрастает, что соответствует самому крутому участку кривой Эйри. При этом излучение внутри резонатора, примерно в $\frac{1+R}{1-R}$ раз больше приходящего, что обусловливает возрастание скорости приближения к резонансу

$$\frac{\mathrm{d}(\Delta\lambda)}{\mathrm{d}I_1} = T(\lambda)\left(\frac{T_{\max}}{2n_2 l} - I_1\frac{\mathrm{d}T}{\mathrm{d}\lambda}\right)^{-1}.$$

При некотором значении $\Delta\lambda = \lambda_0 - \lambda_л$ это приводит к изгибу кривой $I_2 = f(I_1)$ с крутизной больше единицы ($\lambda_л$ — длина волны лазера). По мере увеличения I_1 увеличивается $dT/d\lambda$, знаменатель в приведенном выражении стремится к нулю, а скорость приближения к резонансу стремится к бесконечности (рис. 7.13, *г*), что определяет высокое быстродействие использующих это явление устройств (около 1 пс).

После перехода эталона во включенное состояние пик пропускания резонатора быстро свипирует через частоту лазерного излучения, пока не наступает равновесие, при котором этот пик, настроенный на другую (от первоначальной) сторону относительно частоты излучения, стабилизируется в этом новом положении отрицательной обратной связью. Увеличение входной интенсивности приводит сначала к незначительному изменению выходной, поскольку возникающий сдвиг фаз уменьшает уровень пропускания резонатора, т. е. устройство функционирует как оптический ограничитель. При дальнейшем увеличении входной интенсивности возможно включение следующих резонансных пиков, т. е. возникновение мультистабильности. При уменьшении входной интенсивности внутрирезонаторное поле удерживает резонатор во включенном состоянии при уровнях входной интенсивности, меньших тех, которые требовались для включения устройства. Только после существенного уменьшения I_1 вблизи второго порогового значения $I_т$ небольшое изменение I_1 приводит к значительному изменению показателя преломления и ослаблению пропускания. Поэтому, если на расчетной кривой $I_2 = f(I_1)$ образуется отрицательный наклон (*S*-образная нелинейность), то возникает петля гистерезиса и бистабильность, в пределах которой одной и той же интенсивности I_1 соответствуют два значения I_2. Такая дисперсионная бистабильность является прямым следствием наличия обратной связи.

Гистерезис можно использовать в элементах памяти, поддерживая любое из устойчивых состояний выбором промежуточного значения постоянной интенсивности.

Процесс переключения эталона имеет две существенные особенности: 1) большое пропускание позволяет легко получить бистабильность при малых интенсивностях I_1; 2) форму зависимости $T = \varphi(I_1)$ можно изменять с помощью изменения начальных условий и, в частности, $\Delta\lambda$. Ширина петли гистерезиса зависит от длины интерферометра, длины волны падающего света $\lambda_л$ и свойств вещества, заполняющего резонатор.

Для оптического переключателя, а не элемента памяти, гистерезис не нужен. Конструкцию интерферометра и оптическую схему можно изменить так, что будет устранена петля гистерезиса, но сохранен ограниченный крутой участок, на котором незначительным изменениям интенсивности пучка на входе соответствуют большие изменения пропускания.

Форма кривой $I_2 = f(I_1)$ напоминает зависимость тока коллектора электронного транзистора от тока базы. Характеристика такого вида лежит в основе оптического транзистора, называемого *трансфазером*. В нем два световых пучка, слегка различающиеся углами падения, фокусируются в одной точке на передней грани нелинейного кристалла, например, антимонида индия. Один из пучков (пучок смещения) имеет относительно большую и неизменную интенсивность немного ниже крутого участка кривой пропускания. Другой пучок (управляющий пучок) имеет значительно меньшую интенсивность. Однако, когда он добавляется к постоянному пучку, рабочая точка переходит на участок кривой, где пропускание максимально. При этом незначительная модуляция второго пучка вызывает сильное увеличение пропускания прибора подобно тому, как небольшое изменение тока базы вызывает в транзисторе увеличение тока коллектора.

Нелинейные оптические эффекты, возникающие в луче лазера, могут привести к локальному повышению показателя преломления на оси луча в области максимальной интенсивности. При этом световой луч начинает втягиваться в область, где поле максимально. Это явление получило название *самофокусировки*. Самофокусировка возникает в том случае, когда показатель преломления растет с интенсивностью оптического сигнала. При этом различие в значениях показателя преломления внутри и вне светового пучка конечных размеров приводит к образованию в ранее однородной среде оптического волновода.

В качестве первого приближения рассмотрим двухмерный вариант решения волнового уравнения, предполагая, что показатель преломления $n = n_0 + n_2 E^2$, волновое число $k_0 = \omega n_0 \sqrt{\varepsilon_0 \mu_0}$, напряженность электрического поля $\dot{E} = \dot{E}_0(x, z)\mathrm{e}^{-jk_0 z}$, напряженность E_0 медленно меняется вдоль направления распространения луча ($\partial^2 E_0/\partial x^2 \gg \partial^2 E_0/\partial z^2 \ll k_0 E_0$). Если пренебречь членами квадратичными по $n_2 E_0^2$, то из уравнения (3.2) получим

$$\frac{\partial^2 \dot{E}_0}{\partial x^2} - 2jk_0 \frac{\partial \dot{E}_0}{\partial z} + \omega^2 \varepsilon_0 \mu_0 (n^2 - n_0^2)\dot{E}_0 \approx$$

$$\approx \frac{\partial^2 \dot{E}_0}{\partial x^2} - j2k_0 \frac{\partial \dot{E}_0}{\partial z} + 2k_0^2 \frac{n_2}{n_0} E_0^2 \dot{E}_0 = 0.$$

Если ввести нормированную координату $q = z/(2k_0)$ и обозначение $\varkappa = 2k_0^2 n_2/n_0$, то полученное уравнение можно преобразовать в нелинейное уравнение Шредингера:

$$\frac{\partial^2 \dot{E}_0}{\partial x^2} = -j\frac{\partial \dot{E}_0}{\partial q} + \varkappa E_0^2 \dot{E}_0 = 0.$$

Непосредственной подстановкой можно убедиться, что уравнение при $n_2 > 0$ имеет решение

$$\dot{E}_0 = \eta \frac{\exp\{j[(\xi^2 - \eta^2)q + \xi x - \varphi]\}}{\mathrm{ch}[\eta(x - x_0) + 2\eta\xi q]} \sqrt{\frac{2}{\varkappa}},$$

где ξ, η, x_0, φ — произвольные параметры.

В частном случае, когда $\xi = \varphi = 0$, полученное выражение описывает сжатие луча (вдоль координаты x) шириной $1/\eta$, с волновым числом $k = k_0 + \eta^2/2k_0$, причем интеграл $\int_{-\infty}^{\infty} E_0 \, \mathrm{d}x = 2\pi\sqrt{2/\varkappa}$ (независимо от параметров луча).

В более общем случае ($\xi \neq 0$) происходит также отклонение оси пучка относительно оси z на угол ξ/k_0. На этот же угол отклонен и волновой вектор $k = k_0 + (\eta^2 - \xi^2)/(2k_0)$. В целом стационарное решение задачи описывает распределение поля в сфокусированном луче, аналогичное распределению поля в тонкопленочном волноводе, в котором профиль показателя преломления совпадает с профилем интенсивности волны.

Анализ нестационарного решения уравнения Шредингера (например методом обратного рассеяния) показывает, что при $\pi < \sqrt{\varkappa/2}E_0(x, 0) < 3\pi$ начальное поперечное распределение преобразуется в устойчивый пучок с напряженностью $E_0(x, \infty) = \pi\sqrt{2/\varkappa}$ и мощностью, не зависящей от параметров пучка. В отличие от двухмерной задачи расчет трехмерной самофокусировки в параксиальном приближении приводит к значительному сжатию поперечного профиля пучка в точку, когда начальная мощность пучка превысит критическую. Это и объясняет возникновение оптического разрушения активной среды в мощных квантовых генераторах и усилителях.

§ 7.8. Нелинейные эффекты в диспергирующих средах

В линейных средах различные пространственно-временные спектральные составляющие волновых полей (монохроматические волны) распространяются без искажений, не взаимодействуя друг с другом. При этом волновое поле описывается линейным уравнением

$$\nabla^2 E - \frac{1}{v^2}\frac{\partial^2 E}{\partial t^2} - \mathcal{L}_1(E) = 0,$$

где $\mathcal{L}_1$ — линейный оператор, учитывающий дисперсию в среде.

Применимость линейной модели реальной среды зависит в первую очередь от отношения амплитуды волны E или B к напряженности внутреннего поля $E_в$ или $B_в$, действующего, например, на электроны атомов и молекул, спины и т. д. Линейное уравнение предполагает бесконечно малое отношение $E/E_в$ или $B/B_в$. Если это отношение имеет конечное значение, то в волновом уравнении необходимо удерживать нелинейные члены:

$$\nabla^2 E - \frac{1}{v^2}\frac{\partial^2 E}{\partial t^2} - \mathcal{L}_1(E) = \mathcal{L}_2(E^2) + \mathcal{L}_3(E^3) + \ldots,$$

где $\mathcal{L}_2$, $\mathcal{L}_3$, ..., $\mathcal{L}_n$ — нелинейные операторы. Учет нелинейных членов в волновом уравнении позволяет дать количественное описание качественно новых явлений. Простейший случай был рассмотрен ранее, когда распространялась волна частотой ω и нелинейность привела к последовательному возбуждению гармонических составляющих с частотами $n\omega$, где $n = 2, 3, 4, \ldots$. Изменение спектрального состава излучения означает искажение формы первоначального синусоидального профиля волны.

В среде без частотной дисперсии гармонические составляющие распространяются с одинаковыми скоростями и фазовые соотношения между гармониками сохраняются в процессе их распространения (выполняется условие фазового синхронизма). Если диссипация энергии мала, то затухание мало и нелинейные эффекты накапливаются пропорционально пройденному расстоянию. Поэтому в нелинейной недиспергирующей недиссипативной среде на достаточно больших расстояниях возникают сильные нелинейные искажения исходного профиля волны. В частности, гармоническая волна может превратиться в разрывную ударную волну.

В средах с сильной дисперсией вследствие различия фазовых скоростей волн на различных частотах соотношения между фазами гармоник быстро изменяются в пространстве, нарушая фазовый синхронизм. Поэтому нелинейные эффекты не накапливаются и перекачка энергии между спектральными составляющими незначительна. В этих условиях заметных искажений формы волны не происходит.

В этом параграфе будут рассмотрены явления, которые возникают при распространении волн в средах со слабой дисперсией и слабой нелинейностью. Условие малости предполагает, что за период электромагнитных колебаний при распространении возмущения на расстояния, сравнимые с длиной волны, происходят сравнительно слабые изменения формы волны. Это позволяет упростить решение исходных волновых уравнений путем использования метода медленно изменяющегося про-

филя волны. Подобные условия возникают, в частности, при распространении в оптических волноводах сигналов достаточно большой амплитуды. Поэтому особенности распространения нелинейных волн в диспергирующей среде проанализируем на примере одномодового световода.

В § 3.12 был проведен анализ влияния на огибающую импульса дисперсии первого и второго порядка. Теперь его необходимо дополнить учетом влияния зависимости показателя преломления световода от амплитуды сигнала, а также дисперсии третьего порядка.

Дисперсию волновода и нелинейность среды в следствие их малости будем считать независимыми:

$$\beta(\omega, E) = \beta_1(\omega) + \beta_2(E). \tag{7.9}$$

Каждую из этих компонент разложим в ряд Тейлора и учтем только первые наиболее существенные члены ряда. Тогда

$$\beta_1(\omega) = \beta_1(\omega_0) + \beta'(\omega - \omega_0) + \frac{1}{2!}\beta''(\omega - \omega_0)^2 + \frac{1}{3!}\beta'''(\omega - \omega_0)^3,$$

где

$$\beta' = \left(\frac{\partial\beta}{\partial\omega}\right)_{\omega_0} = \frac{n}{c}\left(1 - \frac{\lambda}{n}\frac{\partial n}{\partial\lambda}\right)_{\omega_0}; \tag{7.10а}$$

$$\beta'' = \left(\frac{\partial^2\beta}{\partial\omega^2}\right)_{\omega_0} = \frac{\lambda}{2\pi c^2}\left(\lambda^2\frac{\partial^2 n}{\partial\lambda^2}\right)_{\omega_0}; \tag{7.10б}$$

$$\beta''' = \left(\frac{\partial^3\beta}{\partial\omega^3}\right)_{\omega_0} = -\frac{\lambda^2}{4\pi^2 c^3}\frac{\partial}{\partial\lambda}\left(\lambda^3\frac{\partial^2 n}{\partial\lambda^2}\right)_{\omega_0}; \tag{7.10в}$$

$$\beta = \frac{n\omega}{c},\quad \lambda = \frac{2\pi c}{\omega},\quad \frac{\partial\lambda}{\partial\omega} = -\frac{2\pi c}{\omega^2} = -\frac{\lambda^2}{2\pi c} = -\frac{\lambda}{\omega}. \tag{7.10г}$$

Для пассивных прозрачных сред достаточную точность получают, предполагая, что нелинейность пропорциональна плотности мощности волны

$$I = \Pi = \frac{1}{2}v_0\frac{\varepsilon E_m^2}{2} = \frac{1}{2}c\varepsilon_0 n_0^3 E_m^2 = \frac{1}{2}c\varepsilon_0 n_0^3 a_0^2,$$

где $n_0 = n(\omega_0)$. Тогда

$$\beta_2(I) = \beta_2(I_н) + \left(\frac{\partial\beta}{\partial I}\right)_{\omega_0}(I - I_н).$$

Иногда нелинейность учитывают выражением

$$n(\omega, E) = n_1(\omega, 0) + n_2 E_m^2 = n_1(\omega, 0) + 2n_2 I/c\varepsilon_0 n_0^3 =$$
$$= n_1(\omega, 0) + b^2 k_0^2 a_0^2.$$

В этом случае

$$\left(\frac{\partial\beta}{\partial I}\right)_{\omega_0} = \frac{\partial}{\partial I}\left(\frac{n\omega}{c}\right)_{\omega_0} = \frac{2\omega_0 n_2(\omega_0)}{c^2\varepsilon_0 n_0^3};$$

$$\left(\frac{\partial\beta}{\partial a_0^2}\right)_{\omega_0} = \frac{\omega_0 n_2(\omega_0)}{c} = \frac{b^2\omega_0}{ck_0^2};$$

$$\beta_2(a_0^2) = \beta_2(a_н^2) + \omega_0 n_2(\omega_0)(a_0^2 - a_н^2)/c. \qquad (7.11)$$

Следовательно, согласно формулам (7.9)...(7.11),

$$\Delta\beta = \beta - \beta_0 = \beta'(\omega - \omega_0) + \frac{1}{2}\beta''(\omega - \omega_0)^2 +$$

$$+ \frac{1}{6}\beta'''(\omega - \omega_0)^3 + 2\pi\frac{n_2}{\lambda}(a_0^2 - a_н^2), \qquad (7.12)$$

где

$$\beta_0 = \beta_1(\omega_0) + \beta_2(a_н^2) = \frac{\omega_0}{c}n_1(\omega_0) = \frac{n_0\omega_0}{c} = 2\pi\frac{n_0}{\lambda},$$

если $\beta_2 = (a_н^2) = 0$.

Положив $a_н = 0$, с учетом выражения (7.12) по аналогии с (3.68) получим уравнение пространственно-временно́й зависимости огибающей импульса:

$$\left(\frac{\partial}{\partial z} + \beta'\frac{\partial}{\partial t}\right)\dot{a}_0 = \frac{1}{2}j\beta''\frac{\partial^2\dot{a}_0}{\partial t^2} - \frac{1}{6}\beta'''\frac{\partial^3\dot{a}_0}{\partial t^3} - j2\pi\frac{n_2}{\lambda}|\dot{a}_0|^2\dot{a}_0 + \alpha\dot{a}_0. \qquad (7.13)$$

Дальнейший анализ упрощает переход к системе координат $t - \beta' z$, которая движется с групповой скоростью $u = 1/\beta'$, и к нормированным переменным

$$\dot{g} = \dot{a}_0\sqrt{2\pi n_2}, \quad \xi = z/\lambda, \quad \tau = (-\lambda\beta'')^{-1/2}(t - \beta' z), \qquad (7.14)$$

предполагающим аномальную дисперсию ($\beta'' < 0$). В этом случае уравнение (7.13) для нормированной огибающей волны сводится к виду

$$\frac{1}{2}\frac{\partial^2\dot{g}}{\partial\tau^2} - j\frac{\partial\dot{g}}{\partial\xi} + |\dot{g}|^2\dot{g} = -j\alpha\lambda\dot{g} + jB\frac{\partial^3\dot{g}}{\partial\tau^3}, \qquad (7.15)$$

где $B = \frac{1}{6}\frac{\beta'''}{\beta''}(-\lambda\beta'')^{-1/2}$.

В среде без потерь ($\alpha = 0$), где отсутствует дисперсия третьего порядка ($\beta''' = 0$), правая часть (7.15) становится равной нулю и это уравнение принимает вид

$$\frac{1}{2}\frac{\partial^2 \dot{g}}{\partial \tau^2} - j\frac{\partial \dot{g}}{\partial \xi} + |\dot{g}|^2 \dot{g} = 0. \qquad (7.16)$$

Параболическое уравнение (7.16) называют *нелинейным уравнением Шредингера* (НУШ), поскольку его можно представить в виде

$$j\frac{\partial \dot{g}}{\partial \xi} = Q\dot{g},$$

где $Q = \frac{1}{2}\frac{\partial^2}{\partial \tau^2} + |\dot{g}|^2$, а $|\dot{g}|^2$ играет роль потенциала. Однако, с уравнением Шредингера из квантовой механики его связывает лишь название. В действительности оно используется в макроскопической электродинамике для описания распространения огибающих волновых пакетов в нелинейных дисперсионных системах, в которых не происходит рассеяния, в частности, для анализа стационарной самофокусировки плоской волны, одномерной модуляции монохроматической волны, термодинамических волн в твердом теле, ленгмюровских волн в плазме, а также для взаимодействия электромагнитного поля с двухуровневой квантовой системой.

Дальнейший анализ решения НУШ будет ограничен условиями распространения сигналов в световодах, в которых при достаточных мощностях светового луча дисперсия и нелинейность оптических свойств материала волокна существенно влияет на уменьшение искажений и длительности передаваемых импульсов. Такая возможность уменьшения искажений и длительности импульсов рассматривается как один из путей повышения скорости передачи информации в оптических линиях связи (вплоть до 1 Тбит/с).

В хороших одномодовых волокнах на основе кварца (SiO_2) потери не превышают 0,2 дБ/км на длине волны 1,5 мкм и групповая дисперсия

$$\beta'' \approx -\frac{0{,}053}{c\omega_0}\left(1 - \frac{\lambda_c}{\lambda}\right), \qquad (7.17)$$

Предполагается, что в таком световоде при $\lambda = \lambda_c = 1{,}27$ мкм групповая дисперсия отсутствует ($\beta'' = 0$), а при $\lambda = 1{,}3$ мкм $\alpha\lambda = 3 \cdot 10^{-2}$, $B \approx -2 \cdot 10^{-3}$. При этом оба члена правой части уравнения (7.15) действительно достаточно малы, т. е. потери и дисперсию третьего порядка можно рассматривать только как малое возмущение. Тогда анализ распространения волны в световоде с хорошим приближением можно вести на основе уравнения (7.16).

Представим нормированную амплитуду в виде

$$\dot{g} = g(\tau, \xi)e^{j\theta(\tau, \xi)}. \qquad (7.18)$$

Тогда после подстановки выражения (7.18) в (7.16) и разделения действительных и мнимых компонент получим

$$\frac{\partial^2 g}{\partial \tau^2} - g\left(\frac{\partial \theta}{\partial \tau}\right)^2 + 2g\frac{\partial \theta}{\partial \xi} + 2g^3 = 0,$$
$$g\frac{\partial^2 \theta}{\partial \tau^2} + 2\frac{\partial g}{\partial \tau}\frac{\partial \theta}{\partial \tau} - 2\frac{\partial g}{\partial \xi} = 0. \qquad (7.19)$$

Решение системы (7.19) найдем в виде выражения для бегущей волны, у которой огибающая сигнала движется с групповой скоростью $u = 1/\beta'$, а модуляция несущей частоты по мере распространения импульса вдоль оси световода z происходит со скоростью w. Это означает, что $\theta(\tau, \xi) = \theta(\tau_\omega)$, где $\tau_w = \dfrac{t - z/w}{\sqrt{-\lambda\beta''}} = \tau - \lambda\xi(1/w - 1/u)/\sqrt{-\lambda\beta''}$, причем $\dfrac{\partial \theta}{\partial \xi} = \dfrac{\partial \theta}{\partial \tau}\dfrac{\partial \tau}{\partial \tau_w}\dfrac{\partial \tau_w}{\partial \xi} = -\dfrac{\lambda}{w\sqrt{-\lambda\beta''}}\dfrac{\partial \theta}{\partial \tau}$. Учтем, что в соответствии с ранее принятым допущением

$$\frac{\partial g}{\partial \xi} = \frac{\partial g}{\partial \tau}\frac{\partial \tau}{\partial \xi} = -\frac{\lambda}{u\sqrt{-\lambda\beta''}}\frac{\partial g}{\partial \tau}.$$

Поэтому уравнения (7.19) можно переписать в виде системы обыкновенных дифференциальных уравнений

$$\frac{\partial^2 g}{\partial \tau^2} - g\left(\frac{\partial g}{\partial \tau}\right)^2 - 2\frac{\lambda}{w\sqrt{-\lambda\beta''}}g\frac{\partial \theta}{\partial \tau} + 2g^3 = 0; \qquad (7.20)$$

$$g\frac{\partial^2 \theta}{\partial \tau^2} + 2\frac{\partial g}{\partial \tau}\frac{\partial \theta}{\partial \tau} + 2\frac{\lambda}{u\sqrt{-\lambda\beta''}}\frac{\partial g}{\partial \tau} = 0. \qquad (7.21)$$

Интегрирование выражения (7.21) приводит к уравнению $\left(\dfrac{\partial \theta}{\partial \tau} + \dfrac{\lambda}{u\sqrt{-\lambda\beta''}}\right)g^2 = \text{const}$, в котором последнюю постоянную удобно принять равной нулю. Интегрирование (7.20) после подстановок $\partial\theta/\partial\tau = -\lambda/(u\sqrt{-\lambda\beta''})$ и $\partial g/\partial\tau = y$ приводит к эллиптическому интегралу

$$\int_{g(0)}^{g(\alpha)} \frac{\mathrm{d}\, g}{\sqrt{P(g)}} = \tau, \qquad (7.22)$$

где

$$P(g) = -g^4 - \frac{\lambda}{\beta''}\left(\frac{1}{u^2} - \frac{2}{wu}\right)g^2 + C. \qquad (7.23)$$

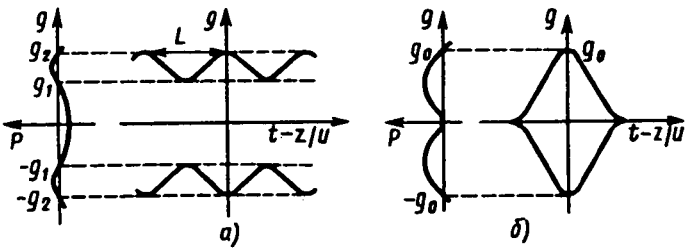


Рис. 7.14. Огибающие волн, соответствующих нелинейному уравнению Шредингера

В зависимости от значения постоянной C изменяются свойства волн, описываемые уравнением (7.16). Однако, очевидно, что стационарным волнам должны соответствовать положительные значения полинома (7.23). При

$$-\frac{\lambda^2}{2\beta''^2}\left(\frac{1}{u^2}-\frac{2}{wu}\right)^2 < C < 0,$$

согласно уравнению (7.22), огибающая волны (рис. 7.14, *а*) имеет пульсирующий характер:

$$g = g_1[1-(1-g_1^2/g_2^2)\mathrm{sn}^2\,\tau]^{-1/2}, \qquad (7.24)$$

где sn — эллиптический синус с модулем $\gamma = 1-g_1^2/g_2^2$.

При $C \to 0$ и $g_1 \to 0$ уравнение (7.24) преобразуется к виду, который можно получить при интегрировании (7.22) от 0 до g_0,

$$g = g_0\,\mathrm{sech}\left[g_0(t-z/u)/\sqrt{-\lambda\beta''}\right]. \qquad (7.25)$$

Уравнение (7.25) описывает импульс (рис. 7.14, *б*) с постоянной амплитудой

$$g_0 = \sqrt{\frac{(2u-w)\lambda}{\beta'' u^2 w}} \qquad (7.26)$$

и пространственным периодом

$$L = \frac{4}{g_0}K(\gamma),$$

где K — полный эллиптический интеграл первого рода.

При $w > 2u$ решение нелинейного уравнения Шредингера соответствует стационарной уединенной волне

$$\dot{g} = g_0\,\mathrm{sech}\left|g_0\frac{t-z/u}{\sqrt{-\lambda\beta''}}\right| e^{-j(t-z/w)\lambda/(u\sqrt{-\lambda\beta''})}, \qquad (7.27)$$

распространяющейся без затухания и искажения формы.

Уединенную волну можно рассматривать как результат взаимного уравновешивания действия нелинейности и дисперсии.

При наличии дисперсии, но в отсутствии нелинейности исключено появление уединенных волн, так как независимо от начальных условий скорость распространения различных фурье-составляющих различна. При наличии нелинейности, но в отсутствии дисперсии также исключено появление уединенных волн из-за непрерывной перекачки энергии импульса в волны более высоких частот (генерация высших гармоник).

Особенностью полученного решения является то, что два импульса вида (7.27) выходят из столкновения с неизменными огибающими и скоростями. Такую уединенную волну называют *солитоном* (solitary wave — уединенная волна).

Солитоны описываются не только нелинейным уравнением Шредингера, но и некоторыми другими нелинейными уравнениями, в частности, уравнениями Кортевега-де-Фриза, синус—Гордона, Максвелла—Блоха. К ним обратимся при рассмотрении нелинейных спиновых волн, волн в активных средах и флуксонов в распределенных туннельных переходах.

Ширину солитона (7.27) $t_c = 2t_0$ определим из условия $g_0 t_0 = \sqrt{-\lambda\beta''}$, т. е. из условия спада амплитуды до значения $\operatorname{sech}(g_0 t_0 / \sqrt{-\lambda\beta''}) = 0{,}65$. С учетом соотношений (7.10б) и (7.15) максимальное значение напряженности электрического поля

$$E_{\max} = \frac{1}{t_0}\sqrt{\frac{-\lambda\beta''}{2\pi n_2}} = \frac{1}{\omega_0 t_0}\sqrt{-\frac{\lambda^2}{n_2}\left(\frac{\partial n}{\partial \lambda^2}\right)_{\omega_0}}.$$

При этом учитывается также, что рассматриваемое решение существует только в условиях аномальной дисперсии, когда $\beta'' < 0$ и параметр t вещественен.

В одномодовом волокне площадью поперечного сечения S удобнее пользоваться пиковой мощностью

$$P_{\max} = \frac{1}{2} v n^2 \varepsilon E_{\max} S \approx -\frac{c n \varepsilon_0 \lambda^2 S}{n_2 \omega_0^2 t_0^2}\left(\frac{\partial^2 n}{\partial \lambda^2}\right)_{\omega_0}, \qquad (7.28)$$

если учесть, что $v \approx c/n$ и показатель преломления n_2 уменьшить в два раза при усреднении напряженности поля по поперечному сечению световода. В световодах, для которых справедливо соотношение (7.17),

$$P_{\max} = \frac{0{,}053 n \varepsilon_0 \lambda^2 S}{4\pi^2 c n_2 t_0^2}\left(1 - \frac{1{,}27\cdot 10^{-6}}{\lambda}\right).$$

При λ = 1,3 мкм; $n_2 = 1{,}2\cdot 10^{-22}$ м^2/В^2; n = 1,5; S = 20 мкм^2 для передачи импульса шириной t_c = 1 пс требуется мощность $P_{\max}$ = 1,6 Вт. При 1,3 > λ > 1,27 мкм возможно дальнейшее снижение мощности

$$P_{max} \approx 2St_c^{-2}\lambda(\lambda - \lambda_c), \qquad (7.29)$$

где λ_c — длина волны, соответствующая нулевой групповой дисперсии ($\beta'' = 0$). При $\lambda = \lambda_c$ солитон исчезает. Уменьшение $\Delta\lambda = \lambda - \lambda_c$ не должно быть ниже уровня, при котором не весь спектр импульса $\Delta f = t_c^{-1}$ находится в пределах аномальной дисперсии. Поэтому минимальное значение

$$\Delta\lambda_{min} = \Delta f \frac{\partial\lambda}{\partial t} = \frac{\lambda^2}{ct_c}. \qquad (7.30)$$

Минимизация мощности не должна приводить также к существенному увеличению дисперсий третьего и более высоких порядков, поскольку в уравнении (7.15) отношение $t/B \sim t_c$. Эти дисперсии должны учитываться в этом уравнении в качестве малых возмущений. Найденное из этого условия минимальное значение $\Delta\lambda_{min}$ близко к полученному ранее (в световодах на основе кварца). При $\lambda \approx 1{,}3$ мкм, с небольшим запасом $\Delta\lambda_{min} = 10^{-20}/t_c$, т. е. $\Delta\lambda_{min} = 10^{-2}$ мкм для солитона шириной $t_c = 1$ пс.

Решение нелинейного уравнения Шредингера методом обратного рассеяния показывает, что отклик среды или волновода при произвольной форме импульса $g(\tau, \xi)$ на входе ($\xi = 0$) в общем случае можно описать с помощью N дискретных солитонов и совокупности нескольких непрерывных волн (мод), которые по мере распространения асимптотически стремятся к нулю. Количество солитонов зависит от площади начального импульса:

$$A = \int_{-\infty}^{\infty} g(\tau, 0)\, d\tau. \qquad (7.31)$$

В частности, если на вход подается импульс вида $g(\tau, 0) = ag_0 \operatorname{sech}(g_0\tau)$, то количество солитонов, порожденных таким импульсом, зависит от нормированной амплитуды a и равно ближайшему целому числу, которое удовлетворяет условию

$$a + 1/2 \geqslant N.$$

Таким образом, один солитон ($N = 1$) возбуждается при $1/2 \leqslant a < 3/2$. Для рассмотренного ранее примера (λ = 1,3 мкм; t_c = 1 пс; $t_c^2 P_{max}$ = 1,6 Вт·пс^2) $a = 1$ и $A \approx \sqrt{1{,}6} \approx 1{,}3$ Вт·пс^2. Это означает, что один солитон может возбуждаться входными импульсами той же длительности мощностями, лежащими в пределах от $1{,}6(1/2)^2 = 0{,}4$ до $1{,}6(3/2)^2 = 3{,}6$ Вт. Однако для начального импульса амплитудой $a = 1 + \alpha$ ($|\alpha| < 1/2$) формирование

стационарного ($\xi \to \infty$) солитона $g(\tau, \xi) = a_\infty g_0 \operatorname{sech}(a_\infty g_0 \tau_\infty)$ происходит не сразу. Асимптотическое значение его амплитуды $a_\infty = 1 + 2\alpha$. Амплитуда и мощность выходного солитона при $\alpha > 0$ оказываются больше, чем у начального импульса, а длительность, согласно (7.28), — меньше, т. е. происходит сжатие импульса. Например, при $\alpha = 1/2$ и мощности начального импульса $P_i = 3{,}6$ Вт асимптотическое значение пиковой мощности $P_{max} = (1 + 2\alpha)^2 \cdot 1{,}6 = 6{,}4$ Вт, а длительность солитона уменьшается в $(1 + 2\alpha)$ раз и становится равной $t_\infty = t_c/(1 + 2\alpha) = 0{,}5$ пс. При этом сохраняются условия образования одного солитона ($t_\infty^2 P_{max} = 1{,}6$ Вт·пс^2). При дальнейшем увеличении мощности входного импульса ($\alpha > 1/2$) количество солитонов возрастает и форма импульса на выходе сильно искажается. В односолитонном режиме параметр $\alpha = \sqrt{P_i/P_{max}} - 1$.

Для формирования солитона начальный импульс должен пройти достаточное расстояние, на котором в полной мере должно проявиться влияние нелинейности и дисперсии. При этом необходимо учитывать, что переходный процесс формирования асимптотического солитона из начального импульса имеет осциллирующий характер с периодом

$$L_1 \approx 2c/(n_2 \omega_0 g)^2 = 5{,}3 \cdot 10^{-18} \lambda S/P.$$

В отсутствии потерь для формирования солитона импульс должен пройти расстояние не меньшее L_1. Например, при $S = 20$ мкм^2; $\lambda = 1{,}3$ мкм и $P = 10$ мВт это расстояние равно 14 км.

При больших потерях солитон теряет свой свойства. Однако и при малых потерях необходимо учитывать влияние потерь на уширение импульса $t_c(z) = t_c(0) e^{23\alpha_{дБ} z}$. В световодах на основе кремния минимум потерь располагается вблизи $\lambda = 1{,}5$ мкм. Поэтому рекомендуется использовать технологию изготовления световодов, при которой $\lambda_c \approx 1{,}5$ мкм. Если в этой области $\alpha_{дБ} = 0{,}2$ дБ/км, то удвоение ширины импульса за счет потерь происходит на расстоянии $L_2 = \ln 2/23\alpha_{дБ} = 15 \cdot 10^3$ м = 15 км.

Таким образом, для оптимизации скорости передачи информации сигналами с минимально возможной мощностью необходимо выбрать длину волны $\lambda = \lambda_c + \Delta\lambda_{min}$, а мощность сигнала — в 2,25 раза выше мощности солитона с амплитудой $a = 1$. При этом начальная ширина импульса $t_н$ за счет сжатия сначала уменьшается вдвое ($t_\infty = t_н/2$), если $\alpha L_1 < 1$, а затем на расстоянии L_2 удваивается ($t_c = 2t_\infty$), т. е. на расстоянии $L = L_1 + L_2$ образуется импульс исходной ширины $t_н$. Мощность на

входе, требуемая для такой передачи сигналов без искажения, может быть найдена из выражений (7.29) и (7.30):

$$P_i = 9\lambda^3 S/(ct_c^3).$$

При $\lambda = 1{,}5$ мкм и $S = 20$ мкм^2 $t_c^3 P_i \approx 2$ Вт·пс^3, т. е. для обеспечения скорости солитонной передачи 1 Тбит/с ($t_c = 1$ пс) на расстояние 30 км необходима мощность примерно 10 Вт. При мощности 10 мВт скорость передачи составит примерно 0,1 Тбит/с ($t_c = 10$ пс), что примерно в 10...100 раз превысит скорость передачи в лучших реальных ($\Delta\lambda/\lambda \approx 0{,}01$) линейных оптических системах, работающих с нулевой групповой дисперсией. Кроме того, солитон устойчив к возмущениям, обусловленным неоднородностью материала световода и внешними помехами (шумами) . Это обеспечивает высокую помехозащищенность и надежность высокоскоростной передачи информации.

§ 7.9. Акустоэлектромагнитные волны в пьезокристаллах

Смещение частиц y_i твердого тела в поле акустической волны описывает уравнение

$$\rho \frac{\partial^2 y_i}{\partial t^2} = \frac{\partial T_{ij}}{\partial x_j}, \qquad (7.32)$$

где ρ — плотность тела; x_j — пространственная координата; T_{ij} — компонента тензора механических напряжений.

Смещение частиц определяет тензор деформаций

$$\sigma_{ij} = \frac{1}{2}\left(\frac{\partial y_i}{\partial x_j} + \frac{\partial y_j}{\partial x_i} + \frac{\partial y_k}{\partial x_i}\frac{\partial y_k}{\partial x_j}\right).$$

Последнее слагаемое этого выражения учитывает нелинейный характер связи компонент тензора механических деформаций с производными от смещения. Эта особенность тензора не зависит от свойств деформируемого тела; она называется *геометрической нелинейностью.*

Упругая нелинейность приводит к следующему виду «эффективного» тензора нелинейной упругости

$$c_{ijklqr}^{эф} = c_{ijrl}\delta_{qk} + c_{jlqr}\delta_{ik} + c_{jklr}\delta_{iq} + c_{ijklqr},$$

где δ — символ Кронекера.

В пьезоэлектрических кристаллах необходимо учитывать также нелинейный пьезоэлектрический эффект, связанный с возбуждением волной механических деформаций волны электрического поля E и нелинейной поляризуемостью атомов.

В этом случае тензор механических напряжений и вектор индукции описываются уравнениями состояния:

$$T_{ij} = c_{ijkl}\sigma_{kl} + \frac{1}{2}c_{ijklmn}\sigma_{kl}\sigma_{mn} - p_{kij}E_k - \frac{1}{2}Q_{klij}E_kE_l - p_{kijlm}E_k\sigma_{lm}, \quad (7.33)$$

$$D_i = \varepsilon_{ij}E_j + \frac{1}{2}\varepsilon_{ijk}E_jE_k + p_{ijk}\sigma_{jk} + Q_{ijkl}E_j\sigma_{kl} + \frac{1}{2}p_{ijklm}\sigma_{jk}\sigma_{lm}, \quad (7.34)$$

где p_{ijk}, p_{ijklm} — компоненты тензоров линейного и нелинейного пьезоэффектов; ε_{ij}, ε_{ijk} — компоненты тензоров линейной и нелинейной диэлектрических проницаемостей; Q_{ijkl} — компонента тензора электрострикции. В уравнениях (7.33) и (7.34) первые слагаемые соответствуют линейным, а остальные — нелинейным составляющим механического напряжения и индукции. Из этих соотношений следует, что *решеточную нелинейность кристалла* можно подразделить на упругую, пьезоэлектрическую, электрострикционную и диэлектрическую. В магнитных материалах возникают также взаимодействия, связанные с нелинейностью магнитных свойств, в частности, магнитострикцией и нелинейной намагниченностью.

Уравнения движения сплошной среды (7.32)—(7.34) вместе с уравнениями Максвелла (5.7) образуют полную систему уравнений, описывающую упругие и электромагнитные процессы в пьезокристаллах. Поскольку в уравнение упругих колебаний (7.32) входят параметры электрических полей, а в уравнения Максвелла — компоненты тензора деформации, полная система описывает связанные акустоэлектромагнитные волны. Связь электромагнитных и упругих процессов определяется пьезомодулями кристалла.

Малая скорость распространения акустических колебаний (0,3...10 км/с) дает возможность получать задержку сигнала 1...30 мкс/см, причем длина акустической волны в веществе, примерно, в 10^5 раз меньше длины электромагнитной волны в свободном пространстве. Размеры акустоэлектрических приборов пропорциональны длине волны, поэтому трудности обеспечения верхней границы рабочих частот (3 ГГц) определяются необходимостью изготовления элементов с размерами, меньшими единиц микрометра, а также возрастанием потерь с ростом частоты.

В нелинейном акустоэлектронном устройстве свертки (рис. 7.15, *а*) используется перемножение амплитуд объемных акустических волн $\dot{Y}_1 = Y_1(t)e^{j(\omega_1 t - \mathbf{k}_1\mathbf{R})}$, $\dot{Y}_2 = Y_2(t)e^{j(\omega_2 t - \mathbf{k}_2\mathbf{R})}$, возбуждаемых двумя пьезоакустическими преобразователями *1* и *3*. Если не учитывать геометрическую нелинейность и опустить тензорные обозначения, то деформация в каждой волне $\sigma = \partial y/\partial R$ и сопровождающая волна электрического поля $E = -\frac{p}{\varepsilon}\frac{\partial y}{\partial R}$. Согласно (7.34), возникающая при взаимодействии волн нелинейная компонента индукции

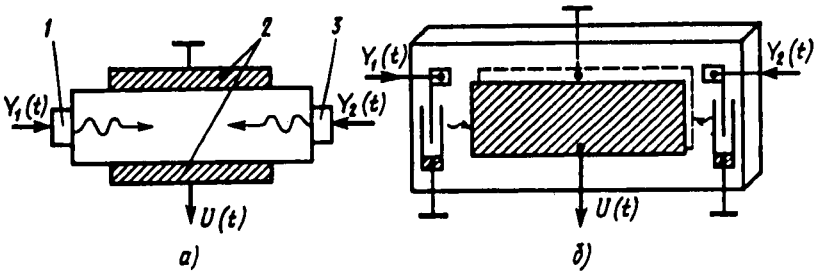


Рис. 7.15. Электроакустические устройства свертки на основе решеточной нелинейности

$$D^{нл} = K\left(\frac{\partial y}{\partial R}\right)^2,$$

где $K = \frac{\varepsilon^{нл} p^2}{2\varepsilon^2} - \frac{pQ}{\varepsilon} + \frac{1}{2} p^{нл}$ — нелинейный коэффициент.

Если волны распространяются навстречу друг другу с одинаковыми частотами ω, то нелинейные компоненты электрического поля

$$D^{нл} = Kk_1^2 y_1 y_2, \quad E^{нл} = Kk_1^2 y_1 y_2 / \varepsilon.$$

Напряжение, снимаемое с распределенных электродов *2*, находящихся в области взаимодействия,

$$u(t) \sim \int_{-L/2}^{L/2} Y_1\left(t - \frac{x}{v}\right) Y_2\left(t + \frac{x}{v}\right) \mathrm{d}x,$$

где L — длина электродов; v — скорость распространения акустической волны. Если протяженность сигнала не превышает размеров области взаимодействия, то выходной сигнал

$$u(t) \sim \int_{-\infty}^{\infty} Y_1(\tau) Y_2(2t - \tau) \mathrm{d}\tau \qquad (7.35)$$

представляет собой акустическую свертку входных сигналов. Если сигнал Y_1 подается на кристалл в виде электрического поля, а сигнал Y_2 — в виде акустической волны, то выходной сигнал

$$u(t) \sim \int_{-\infty}^{\infty} Y_1(\tau) Y_2(2\tau - t) \mathrm{d}\tau \qquad (7.36)$$

принимает вид акустической корреляции. Выражения (7.35) и (7.36) отличаются от принятых определений свертки и корреляции двух функций. Для их получения необходимы дополнительные преобразования входных

сигналов Y_1 и Y_2 (обращение во времени, сжатие или растяжение одного из сигналов) либо использование двух последовательных нелинейных взаимодействий. Например, когда в выражении (7.35) сигнал Y_1 является δ-функцией, на выходе формируется сигнал

$$u(t) \sim \int_{-\infty}^{\infty} \delta(\tau) Y_2(2t-\tau)\,d\tau = Y_2(2t),$$

сжатый во времени в два раза.

Когда в выражении (7.36) сигнал Y_2 имеет вид δ-функции, на выходе формируется сигнал

$$u(t) \sim \int_{-\infty}^{\infty} Y_1(\tau)\delta(2\tau-t)\,d\tau = Y_1(t/2),$$

растянутый во времени в два раза. Если же в выражении (7.36) вид δ-функции имеет сигнал Y_1, то происходит обращение сигнала во времени:

$$u(t) \sim \int_{-\infty}^{\infty} \delta(\tau) Y_2(2\tau-t)\,d\tau = Y_2(-t).$$

Пример реализации корреляционной функции с использованием двух устройств свертки показан на рис. 7.16. Один из входных сигналов *1* на частоте ω подается на входной преобразователь *I*. Путем подачи на параметрический электрод δ-импульса *4* на частоте 2ω этот сигнал обращается во времени и затем подается на один из входов *2* преобразователя *II*. На другой вход преобразователя *II* подается сигнал *5* на частоте ω. Выходной сигнал *3*, снимаемый с параметрического электрода, пропорционален функции корреляции сигналов *1* и *5* на частоте 2ω.

В устройствах свертки на поверхностных волнах (см. рис. 7.15, *б*) эффективность нелинейных взаимодействий акустических волн намного выше. Например, при использовании ниобата лития выигрыш в эффективности может достигать 50. Однако, на частотах 1 ГГц и выше технологические трудности заставляют применять устройства на объемных акустических волнах.

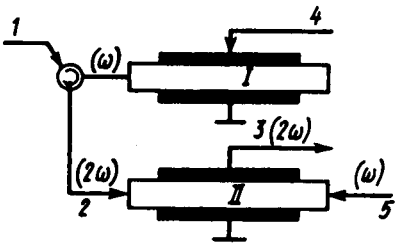


Рис. 7.16. Коррелятор на двух устройствах свертки

К перспективным материалам для звукопроводов устройств свертки и корреляции относят ниобат лития $LiNbO_3$, танталат лития $LiTaO_3$, иодат лития $LiIO_3$, германат висмута $Bi_{12}GeO_{20}$, силикат висмута $Bi_{12}SiO_{20}$ и пьезокерамики типа ЦТС (цирконат-титанат свинца).

§ 7.10. Электроакустические волны в пьезополупроводниках

В материалах, обладающих одновременно и пьезоэлектрическими, и полупроводниковыми свойствами (пьезополупроводники), а также в слоистых структурах пьезоэлектрик—полупроводник используется более эффективная, чем решеточная, *концентрационная нелинейность*. Для этого можно применить, например, ниобат лития, восстановленный в атмосфере водорода. При этом возникают носители *n*-типа, концентрация которых зависит от условий восстановления. Оптимальным считают концентрацию носителей $3 \cdot 10^{19}$ м$^{-1}$. Дальнейшее ее увеличение приводит к чрезмерному затуханию акустической волны.

Воздействие акустической волны на свободные электроны обусловлено изменением внутрикристаллических полей при смещении (колебаниях) атомов решетки. Наиболее существенный вклад в подобное взаимодействие на частотах до 10^{11} Гц вносит пьезоэлектрический эффект (в пьезоэлектрических средах).

Доминирующим типом нелинейности пьезоэлектриков-полупроводников является концентрационная нелинейность, проявляющаяся в нелинейной зависимости кинетических коэффициентов от концентрации электронов, связанная с перераспределением электронов проводимости в присутствии акустической волны и приводящая к нелинейности уравнений для электрического тока, потока энергии и уравнения баланса энергии.

Если распространение двух акустических волн сопровождается распространением электрического поля $e_1 = E_1 \sin(\omega_1 t - \mathbf{k}_1\mathbf{R}_1)$, $e_2 = E_2 \sin(\omega_2 t - \mathbf{k}_2\mathbf{R}_2)$, то в полупроводнике возникают и волны концентрации свободных электронов

$$n_1 = N_1 \sin(\omega_1 t - \mathbf{k}_1\mathbf{R}_1), \quad n_2 = N_2 \sin(\omega_2 t - \mathbf{k}_2\mathbf{R}_2).$$

Одним из проявлений нелинейных процессов является взаимодействие электрического поля одной акустической волны с носителями, сопровождающими другую волну. Это приводит к образованию нелинейной составляющей индукции $D^{нл}$ и нелинейной составляющей тока смещения $\partial D^{нл}/\partial t$. Поэтому без учета тока диффузии нелинейная составляющая плотности тока $j^{нл} = e\mu(n_1 e_2 + n_2 e_1)$.

При сравнительно небольших амплитудах акустических волн, когда еще не сказываются другие типы нелинейности, эффективность взаимодействия за счет концентрационной нелинейности оказывается на несколько порядков выше, чем при взаимодействии за счет упругой нелинейности. Эффективность концентрационной нелинейности легко поддается управлению путем изменения количества носителей или смещающего электрического поля.

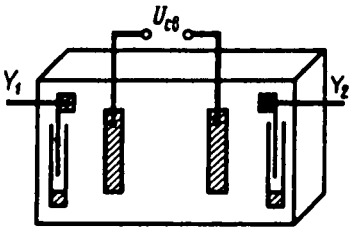


Рис. 7.17. Устройство свертки на основе концентрационной нелинейности

На основе взаимодействия акустических волн и волн объемного заряда (в постоянном электрическом поле) может быть получено усиление сигнала: при превышении дрейфовой скорости электронов скорости акустической волны носители заряда отдают энергию бегущей волне сигнала. Построенные на этом эффекте усилители при несущей частоте до 1,5 ГГц могут иметь коэффициент усиления до 100 дБ и коэффициент шума, меньший 10 дБ. Режим генератора обеспечивает включение в цепь усилительного прибора с положительной обратной связью.

Важным свойством пьезоэлектриков-полупроводников является эффект акустоэлектронной памяти. Для акустоэлектронного взаимодействия и запоминания можно также использовать матрицу диодов, расположенных над поверхностью звукопровода.

Конструкция устройств свертки на пьезополупроводниках принципиально не отличается от рассмотренной конструкции устройств свертки на пьезодиэлектриках, если не считать возможности использования продольного съема выходного сигнала свертки $U_{св}$ (рис. 7.17), при котором выделяется тангенциальная составляющая электрического поля.

В другом типе устройств свертки применяют сложные структуры пьезоэлектрик-полупроводник, в которых взаимодействие происходит при проникновении электрических полей поверхностной акустической волны (ПАВ) в полупроводник. В одних таких структурах пьезоэлектрик и полупроводник разделены зазором, в других — используется монолитная конструкция.

Типичная конструкция монолитного устройства свертки представлена на рис. 7.18 (МЦОК-*структура*: металл—оксид цинка—оксид кремния—кремний). Слой оксида кремния предохраняет кремний от загрязнений, возникающих в процессе напыления пленки пьезоэлектрика ZnO. Его толщина обычно лежит в пределах 0,05...0,2 мкм. Металлические площадки *1* можно рассматривать как элементы встречно-штыревых преобразователей *2*. Они выполняют функции экрана, уменьшающего нежелательное взаимодействие электрических полей преобразователей с носителями заряда в полупроводнике. Металлические площадки снижают также уровни ложных сигналов.

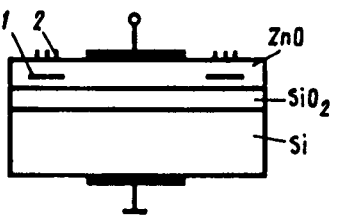


Рис. 7.18. Схема монолитного устройства свертки

В гибридном устройстве свертки (рис. 7.19) реализован новый принцип нелинейной обработки поверхностной акустической волны. Слои ZnO и SiO_2 образуют затворный диэлектрик МДП-транзистора. Квадратичная зависимость тока стока от напряжения на затворе в области насыщения обусловливает нелинейность гибридной структуры в условиях инверсии поверхности Si. Это позволяет получить сигнал свертки более, чем в два раза превышающей сигнал в МЦОК-устройстве в условиях обеднения поверхности полупроводника (за счет варакторного механизма нелинейности).

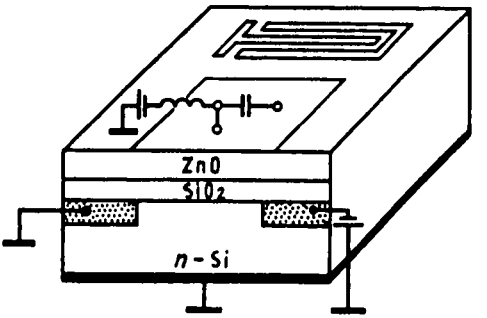


Рис. 7.19. Схема гибридного устройства свертки

Особый класс нелинейных акустоэлектронных устройств составляют устройства, использующие запоминание и хранение информации. В одной из модификаций акустоэлектронного запоминающего устройства (АЗУ) запоминание осуществляется непосредственно на поверхностных электронных состояниях (ПЭС) полупроводника, входящего в состав слоистой структуры. При записи пространственно-неоднородное электрическое поле возникает за счет нелинейного взаимодействия типа «звук—звук» или «звук—поле». Под действием электрического поля на поверхности полупроводника происходит соответствующее распределение концентрации свободных носителей заряда, захваченных на ПЭС. В кристаллах Si при комнатной температуре получено время памяти около 1 мс. Минимальное время записи по теоретическим оценкам составляет 1 мс. В кристаллах CdS и GaAs при комнатной температуре время памяти составило 10 мс и 0,5 с соответственно. При понижении температуры время памяти существенно возрастает. Быстрое стирание записанного сигнала и замену его новым можно произвести с помощью соответствующей подсветки.

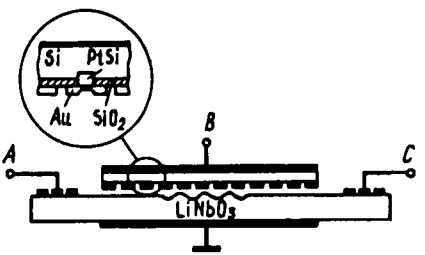


Рис. 7.20. Акустоэлектронное запоминающее устройство на диодах Шотки

В другой модификации АЗУ запоминание осуществляется на матрице диодов. Типичным примером может служить структура АЗУ на диодах Шотки (рис. 7.20). Если на акустический вход *A*

подается сигнал $S(t)\sin\omega t$, то он возбуждает поверхностную акустическую волну, сопровождаемую поперечным электрическим полем $e_{\perp} \sim S(t - x/v)\sin(\omega t - kx)$, которое приводит к локальным изменениям плотности заряда в приповерхностном слое полупроводника. Эти изменения под каждым барьером Шотки можно зафиксировать, если подать на структуру короткий импульс отрицательного напряжения, открывающий диоды. Заряд матрицы диода складывается из однородного заряда q_0, обусловленного импульсом смещения и пропорционального входному сигналу заряда памяти $q_s(x) \sim S(-x/v)\sin kx$. Время хранения записанной зарядовой решетки определяется токами утечки диодов, которые при обратном смещении малы.

Кроме рассмотренного импульсного способа записи может быть использован и параметрический способ, в основе которого лежит нелинейное взаимодействие волн. При этом учитывается то обстоятельство, что для сигналов, возникающих в АЗУ, сигнал памяти частотой $\omega = 0$ может быть получен при взаимодействии поверхностной акустической волны $S(t - x/v)$ как с однородным высокочастотным электрическим полем $W(t)$, так и со встречной акустической волной $W\left(t - \dfrac{L-x}{v}\right)$ той же частоты (L — длина матрицы). В результате возникает составляющая электрического тока i, модулированная по закону $\sin kx$ или $\sin 2kx$ и пропорциональная амплитудам записываемого сигнала $S(t)$ и сигнала записи $W(t)$. Эта составляющая тока вызывает на диодах заряд $\mathrm{d}q_s = i\,\mathrm{d}t$. После окончания процесса записи на матрице возникает распределение заряда

$$q_s(t,x) \sim \left[\int S(t - x/v)W(t)\,\mathrm{d}t\right]\sin\omega t$$

или

$$q_s(t,x) \sim \left[\int S(t - x/v)W\left(t - \frac{L-x}{v}\right)\mathrm{d}t\right]\sin 2\omega t,$$

в зависимости от того, куда подан сигнал записи — на параметрический электрод B или на противоположный акустический вход C. В обоих случаях распределение заряда представляет собой функцию корреляции сигналов S и W.

Полученное распределение заряда может быть считано в любой момент в пределах времени памяти АЗУ с помощью одного из тех воздействий, которые использовались при записи. Если, например, подать на параметрический электрод B сигнал считывания $R(t)\sin\omega t$, то на входе A будет зарегистрирован сигнал корреляции

$$u_A(t) \sim \left[\int S(\tau - t)R(\tau)\,\mathrm{d}\tau\right]\sin\omega t,$$

а на входе C — сигнал свертки

$$u_C(t) \sim \left[\int S(t - \tau)R(\tau)\,\mathrm{d}\tau\right]\sin\omega t,$$

записанного и считывающего сигналов.

Диоды Шотки способны полностью зарядиться за время, не превышающее 1 мс, что достаточно для записи сигналов с полосой в сотни мегагерц. В некоторых матрицах время памяти доходит до 400 мс, а при использовании матрицы *p-n*-диодов — до 10 с. «Очистку» памяти удобнее всего осуществлять освещением матрицы светодиодами через полупрозрачный заземленный электрод. Быстрая разрядка структуры происходит о помощью тока, обусловленного генерированными в области пространственного заряда неравновесными носителями.

Емкость цифровой памяти в АЗУ с перекрестными акустическими каналами и использованием обращения фазы поверхностной волны может быть доведена до 10^5 бит, скорость ввода и вывода информации — до 10^8 бит/с, плотность памяти — до $4 \cdot 10^5$ бит/см^2 при длительности выборки 210 мкс и времени хранения информации 0,1 с.

Созданные на базе современной технологии нелинейные акустоэлектронные устройства позволяют существенно расширить возможности акустоэлектроники. Благодаря многофункциональности, адаптивности к виду обрабатываемых сигналов, миниатюрности, низкому энергопотреблению и высокой надежности акустоэлектронные процессы оказались достаточно конкурентоспособными по сравнению с аналогичными устройствами, основанными на других физических принципах.

АЗУ существенно превосходят традиционные акустические линии задержки по времени хранения, одновременно обладая столь же высокой скоростью передачи данных и плотностью памяти. Основное достоинство АЗУ — возможность одновременно с запоминанием аналоговых сигналов выполнения ряда специфических обработок поступающих сигналов. Стремительное развитие устройств с акустической памятью было вызвано реализацией на основе запоминающего устройства программируемого согласованного фильтра. При обработке сигналов большой длительности удалось получить корреляционную функцию уже в стадии записи.

Современные цифровые фурье-процессоры обладают высокой точностью, большим динамическим диапазоном и универсальностью. Однако с увеличением полосы обрабатываемых частот их масса, габариты, потребляемая мощность и стоимость резко возрастают, что ограничивает их применение и делает конкурентноспособными с ними акустоэлектронные аналоговые фурье-процессоры. Перспективными становятся

устройства свертки, алгоритм работы которых определяется видом опорного сигнала. В акустоэлектронных вариантах быстрая смена опорных сигналов обеспечивает их использование в системах связи с адаптивным приемом. Изменение параметров опорных сигналов позволяет осуществить выбор нужной разрешающей способности. Таким образом, оказывается возможной и компенсация искажений, вызванных неточностями топологии и температурной нестабильностью.

Аналоговые устройства, осуществляющие программируемые операции свертки и согласованной фильтрации сложных сигналов в реальном масштабе времени, называют *конвольверами*. Информационная емкость (произведение полосы пропускания на время интегрирования) акустоэлектрических конвольверов может достигать нескольких тысяч. Основные их достоинства — быстродействие, широкая полоса частот, малые габаритные размеры и масса, высокая надежность работы и относительно низкая стоимость.

Глава 8
Электромагнитные явления в ферромагнетиках

Между магнитными свойствами ферромагнетиков и электрическими свойствами сегнетоэлектриков имеется аналогия. Тем и другим свойственна спонтанная поляризация (магнитная или электрическая), которая исчезает при изменении температуры в точке фазового перехода (точке Кюри).

Существенные отличия между ферромагнитными и сегнетоэлектрическими явлениями вызваны различием в характере тех микроскопических сил взаимодействия, которые приводят к спонтанной поляризации. У сегнетоэлектриков, например, вектор спонтанной поляризации относительно прочно связан с определенными направлениями в кристалле, взаимодействие молекул в кристаллической решетке анизотропно. В ферромагнетиках возникновение спонтанной намагниченности мало зависит от направления суммарного магнитного момента относительно решетки и в основном определяется обменным взаимодействием атомов.

Наибольшее быстродействие ферро- и ферритомагнитных элементов, используемых в вычислительной технике и технике СВЧ, связано с применением тонких магнитных пленок. Время их перемагничивания достигает нескольких наносекунд при относительно малых мощностях управляющих полей. Вихревые токи и поверхностный эффект в пленках практически отсутствуют. Магнитные пленки отличаются высокой технологичностью изготовления и более разнообразными анизотропными свойствами, чем у других магнитных материалов. В связи с этим дальнейшее изложение будет посвящено рассмотрению свойств и технических применений тонких пленок.

§ 8.1. Магнитные пленки с плоскостной анизотропией

Магнитной тонкой пленкой называют структуру, толщина магнитного слоя которой соизмерима с линейными размерами доменов в монокристаллах. Обычно их толщина колеблется от 30 до 1000 нм. Магнитный слой наносят на тщательно отполированную диэлектрическую или металлическую подложку.

Поведение тонких пленок отличается от поведения массивных магнитных материалов в силу двух основных причин. Во-первых, в отличие от внутренних электронных спинов поверхностные спины находятся в структуре с более низкой симметрией, так как имеют «соседей» только со стороны пленки. Во-вторых, расположение атомов в нескольких слоях, ближайших к подложке, зависит от природы подложки и условий осаждения пленки, в частности температуры, величины и направления внеш-

него магнитного поля. Свойства пленок зависят также от неровностей подложки и окисления наружных слоев.

Наиболее широкое применение в технике находят пленки с одноосной анизотропией. У таких пленок имеется одно преимущественное направление (вдоль которого стремится ориентироваться вектор намагниченности), называемое *направлением легкого намагничивания*. Направление, вдоль которого намагничивание пленки требует наибольшего внешнего поля, называют *направлением трудного намагничивания*. Первоначально использовали пленки с плоскостной анизотропией, у которых ось легкого намагничивания располагалась в плоскости пленки. Их обычно изготовляли из пермаллоевых сплавов, для которых характерно нулевое значение коэффициента магнитострикции в направлении оси анизотропии (81% Ni, 19% Fe). Это позволяло уменьшить энергию управления и зависимость свойств пленок от внешних воздействий. В настоящее время большое внимание уделяют монокристаллическим пленкам, у которых ось легкого намагничивания (ОЛН) перпендикулярна поверхности. Типичными материалами для этих пленок являются марганцово-висмутовый сплав, а также гексаортоферриты и некоторые редкоземельные ферриты со структурой граната. Свойствам пленок с перпендикулярной анизотропией посвящен § 8.2.

В пленках из материалов с вектором намагниченности, параллельным плоскости пленки, одноосная анизотропия создается в процессе изготовления пленки (например, вакуумным напылением) путем приложения постоянного магнитного поля напряженностью 0,4...2,4 кА/м параллельно плоскости подложки или перемещения пленки. Расположение вектора намагниченности **M** в плоскости пленки из магнитомягкого материала связано с большим размагничивающим фактором в направлении, перпендикулярном плоскости пленки ($N_z = 1$). Поэтому при малейшем выходе вектора намагниченности из плоскости пленки (под углом ψ) создается большое размагничивающее поле $H_{\text{размаг}} = M \sin\psi$, которое возвращает вектор намагниченности обратно в плоскость пленки. Внешнее поле, приложенное во время роста пленки, лишь обеспечивает одинаковую ориентацию оси легкого намагничивания по всей плоскости пленки. Вследствие малой толщины пленки по сравнению с другими ее линейными размерами размагничивающие поля в направлении, параллельном плоскости пленки, малы ($N_x \approx N_y \approx 0$) и разбиение на домены в отсутствие внешнего магнитного поля не приводит к заметному уменьшению энергии пленки. Пленка обычно содержит один центральный домен, ориентированный вдоль ОЛН и окруженный небольшими краевыми доменами с противоположным направлением вектора намагниченности. Такое состояние можно рассматривать как практически монодоменное.

Если направление внешнего поля в плоскости пленки составляет некоторый угол с направлением ОЛН, то новое направление вектора намагниченности пленки под углом θ к ОЛН (x) определяется минимумом энергии магнитной пленки. Для магнитной пленки в виде круглого пятна общая свободная энергия складывается в основном из энергии анизотропии ($K\sin^2\theta$) и энергии намагниченного тела во внешнем магнитном поле ($\mu_0\mathbf{MH}$):

$$W = K\sin^2\theta - \mu_0 M(H_x\cos\theta + H_y\sin\theta).$$

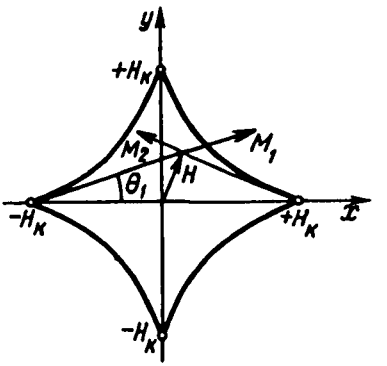


Рис. 8.1. Критическое поле необратимого когерентного вращения намагниченности пленки

Из условия минимума энергии (устойчивого равновесия)

$$\frac{\partial W}{\partial\theta} = 2K\sin\theta\cos\theta - \mu_0 MH_y\cos\theta + \mu_0 MH_x\sin\theta = 0, \qquad (8.1)$$

а в критической точке перехода из одного устойчивого состояния равновесия в другое

$$\frac{\partial^2 W}{\partial\theta^2} = 2K(\cos^2\theta - \sin^2\theta) + \mu_0 MH_y\sin\theta + \mu_0 MH_x\cos\theta = 0. \qquad (8.2)$$

Из совместного решения уравнений (8.1) и (8.2) следует уравнение астероиды (рис. 8.1)

$$H_x^{2/3} + H_y^{2/3} = H_k^{2/3}, \qquad (8.3)$$

где

$$H_k = 2K/(\mu_0 M). \qquad (8.4)$$

Касательная к астероиде является геометрическим местом точек, соответствующих равновесным положениям вектора намагниченности. Направление θ намагниченности пленки, соответствующее какому-либо вектору напряженности, можно получить, проведя из его конца касательные к астероиде на участках, примыкающих к оси легкого намагничивания. Если конец вектора **H** лежит внутри астероиды, то вектор намагниченности имеет два равновесных положения: $\mathbf{M}_1$ и $\mathbf{M}_2$ (см. рис. 8.1). Когда конец вектора напряженности лежит вне астероиды, можно провести лишь одну касательную, и вектор намагниченности будет иметь лишь одно равновесное направление.

Следовательно, астероида отделяет значения H_x и H_y, которым соответствуют два устойчивых состояния пленки, от значений, которым соот-

ветствует одно устойчивое состояние. Таким образом, уравнение (8.3) описывает кривую пороговых значений, при превышении которых происходит переключение пленки.

Поле (8.4) называют *полем анизотропии*. Оно имеет физический смысл минимального поперечного поля, под действием которого вектор намагниченности поворачивается на 90°. В этом можно убедиться, если переписать уравнение (8.1) в виде

$$\frac{H_y}{\sin\theta} - \frac{H_x}{\cos\theta} = H_k \qquad (8.5)$$

и положить продольное поле $H_x = 0$.

Модуль вектора намагниченности не зависит от его направления θ. Намагниченность в направлении оси тяжелого намагничивания

$$M_y = M\sin\theta, \qquad (8.6)$$

в направлении оси легкого намагничивания

$$M_x = M\cos\theta. \qquad (8.7)$$

При намагничивании вдоль оси тяжелого намагничивания ($H_x = 0$) согласно уравнениям (8.5) и (8.6) $M_y = MH_y/H_k$, откуда следует, что M_y является линейно меняющейся функцией H_y, если $|H_y| \leqslant H_k$ (рис. 8.2, *а*). При намагничивании вдоль оси легкого намагничивания ($H_y = 0$) вектор намагниченности под действием энергии анизотропии поворачивается в направлении ОЛН, имея при $|H_x| \leqslant H_k$ два устойчивых состояния $M_x = \pm M$, а при $|H_x| > H_k$ — одно (рис. 8.2, *б*).

Меняя соотношение и порядок изменения полей H_x и H_y и наблюдая за изменением различных составляющих M_x и M_y, можно получить большое многообразие петель гистерезиса.

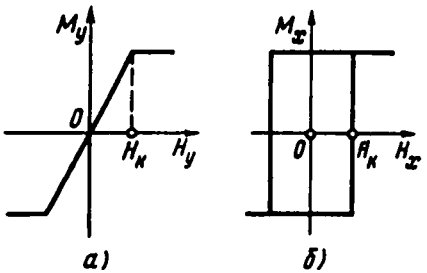


Рис. 8.2. Характеристики намагничивания в направлении осей тяжелого (*а*) и легкого (*б*) намагничивания

Проведенные расчеты справедливы только в том случае, когда перемагничивание осуществляется когерентным необратимым вращением (спины всех электронов одновременно поворачиваются в направлении поля) в отсутствие процессов смещения доменных границ и некогерентного вращения. Неоднородность магнитной пленки приводит, как правило, к возникновению некоге-

рентного (неоднородного) вращения. Оно обусловлено микронатяжениями, кристаллографической анизотропией, анизотропией формы, неоднородностями химического состава, немагнитными включениями, остаточными газами, косым напылением, а также неоднородностью толщины пленки, дефектами полировки и т. п. В этом случае уравнение астероиды (8.3) может не соответствовать результатам эксперимента.

Перемагничивание пленки путем смещения доменных границ происходит при приложении поля $H_c < H_y < H_k$. Этот процесс протекает, как правило, сравнительно медленно.

Время переключения и выходной сигнал определяются динамикой процессов перемегничивания. Высокие скорости переключения тонких пленок возникают при приложении внешнего поля выше порогового и связаны с когерентным вращением электронных спинов. Наличие у электрона механического момента приводит в процессе переключения к вращению (прецессии) вектора намагниченности вокруг направления внешнего поля. При этом вектор намагниченности выходит из плоскости пленки и на ее поверхностях возникает сильное размагничивающее поле $H_{\text{разм}} = NM_n = M\sin\psi$, перпендикулярное поверхности. Вокруг этого поля также начинает прецессировать намагниченность, пока не повернется параллельно приложенному полю и действующий момент сил не станет равен нулю. Так как скорость прецессии, как правило, пропорциональна $H_{\text{разм}}$ ($\omega = \gamma\mu_0 H_{\text{разм}}$), а размагничивающее поле велико, время переключения пленки получается очень малым (порядка наносекунды).

Время переключения монодоменной пленкой при когерентном вращении с учетом магнитной вязкости можно определить из уравнения движения магнитного момента (6.8), если его переписать в виде

$$\frac{d\mathbf{M}}{dt} = -\gamma\mathbf{T} - \frac{\lambda}{M^2}\mathbf{M}\times\mathbf{T}, \qquad (8.8)$$

где $\mathbf{T} = \mathbf{R}\times\mathbf{F} = -\mathbf{R}\times\nabla W$ — вращающий момент, дейст-

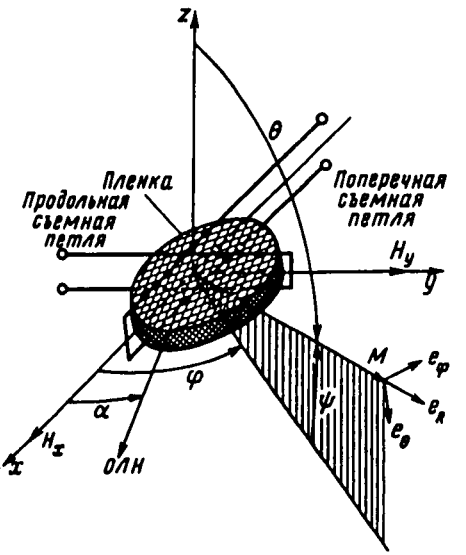


Рис. 8.3. Ориентация системы координат относительно оси анизотропии тонкой пленки

вующий на вектор **M**; его можно выразить через радиус-вектор **R** в направлении вектора $\mathbf{M} = M\mathbf{e}_R$ и обобщенную силу $\mathbf{F} = -\nabla W$. При этом свободная энергия спиновой системы должна учитывать угол наклона ψ намагниченности к поверхности пленки, поворот оси x на угол α относительно ОЛН (рис. 8.3) и энергию размагничивающего поля ($0{,}5\mu_0 H_{\text{разм}} M_n$):

$$W = K\sin^2(\varphi - \alpha) - \mu_0 H_x M\cos\psi\cos\varphi -$$

$$-\mu_0 H_y M\cos\psi\sin\varphi + \frac{1}{2}\mu_0 M^2\sin^2\psi. \qquad (8.9)$$

В сферической системе координат $\partial W/\partial R = 0$ и

$$\nabla W = \mathbf{e}_\theta \frac{1}{R}\frac{\partial W}{\partial\theta} + \mathbf{e}_\varphi \frac{1}{R\sin\theta}\frac{\partial W}{\partial\varphi}.$$

Поэтому

$$\mathbf{T} = -\mathbf{R}\times\nabla W = -\begin{vmatrix} \mathbf{e}_R & \mathbf{e}_\theta & \mathbf{e}_\varphi \\ R & 0 & 0 \\ 0 & \frac{1}{R}\frac{\partial W}{\partial\theta} & \frac{1}{R\sin\theta}\frac{\partial W}{\partial\varphi} \end{vmatrix} = \frac{\mathbf{e}_\theta}{\sin\theta}\frac{\partial W}{\partial\varphi} - \mathbf{e}_\varphi\frac{\partial W}{\partial\theta}. \qquad (8.10)$$

Подставив уравнение (8.10) в (8.8) и учтя, что здесь $\theta = 90° - \psi$, найдем

$$\frac{\partial\mathbf{M}}{\partial t} = M\frac{\partial\theta}{\partial t}\mathbf{e}_\theta + M\cos\psi\frac{\partial\varphi}{\partial t}\mathbf{e}_\varphi.$$

Приравняв одноименные проекции, получим

$$M\frac{\partial\varphi}{\partial t}\cos\psi = -\gamma\frac{\partial W}{\partial\psi} - \frac{\lambda}{M\cos\psi}\frac{\partial W}{\partial\varphi}, \qquad (8.11)$$

$$M\frac{\partial\psi}{\partial t} = -\frac{\gamma}{\cos\psi}\frac{\partial W}{\partial\varphi} + \frac{\lambda}{M}\frac{\partial W}{\partial\varphi}. \qquad (8.12)$$

Проведя дифференцирование и подстановку уравнения (8.11) в (8.12), можно свести их к приближенным дифференциальным уравнениям второго порядка:

$$\frac{\partial^2\varphi}{\partial t^2} + \lambda\mu_0\frac{\partial\varphi}{\partial t} + \gamma^2\mu_0\frac{\partial W}{\partial\varphi} = 0; \qquad (8.13)$$

$$\frac{\partial\psi}{\partial t} = \frac{1}{\mu_0\gamma M}\frac{\partial^2\varphi}{\partial t^2}, \qquad (8.14)$$

если учесть, что с достаточной точностью можно считать $M >> 2K/(\mu_0, M) = H_k$ и $\lambda \ll \gamma M$.

В общем случае решение не только точных уравнений (8.11) и (8.12), но и приближенных уравнений (8.13) и (8.14) встречает большие трудности. Приближенное решение оказывается несколько проще в случае сравнительно малых скоростей перемагничивания, когда можно пренебречь второй производной $\partial^2\varphi/\partial t^2$. Тогда

$$\lambda\frac{\partial\varphi}{\partial t} + \gamma^2\frac{\partial W}{\partial\varphi} = 0; \tag{8.15}$$

$$\partial\psi/\partial t = 0. \tag{8.16}$$

Поскольку в начальный момент $\psi = 0$, то, согласно (8.16), вектор намагниченности будет вращаться в плоскости пленки. В этом случае в соответствии с (8.15) и (8.9) при , $\alpha = 0$ время перемагничивания

$$\tau_\varphi = \frac{\lambda}{\gamma^2}\int_{\varphi_2}^{\varphi_1}\frac{\mathrm{d}\varphi}{\partial W/\partial\varphi} = \frac{\lambda}{\gamma^2\mu_0 M}\int_{\varphi_2}^{\varphi_1}\frac{\mathrm{d}\varphi}{H_k\sin\varphi\cos\varphi + H_x\sin\varphi - H_y\cos\varphi} \tag{8.17}$$

будет определяться параметром затухания λ (магнитной вязкостью) и вращающим моментом $T = \mathrm{d}W/\mathrm{d}\varphi$ (φ_1, φ_2 — равновесные значения угла в начальный и конечный моменты времени). В частности, если составляющая поля вдоль направления трудного намагничивания отсутствует ($H_y = 0$), то векторы $\mathbf{M}$ и $\mathbf{H} = H\mathbf{e}_x$ антипараллельны ($\varphi = 180°$), вращающий момент отсутствует и время перемагничивания бесконечно велико. При возрастании поперечной компоненты H_y время переключения быстро уменьшается.

Использованная аппроксимация $\partial^2\varphi/\partial t^2 = 0$, как следует из формулы (8.16), не учитывает время движения намагниченности поперек пленки, т. е. дает заниженное значение времени перемагничивания. Приближенно оценить время поворота вектора намагниченности в направлении ψ можно, положив, например, в уравнениях (8.13) и (8.14) $\partial W/\partial\varphi = \mathrm{const}$ и решив его относительно ψ. При таком подсчете угол ψ будет достигать равновесного значения (порядка 2...3°) за время $\tau_\psi \approx \lambda^{-1}$, равное в пермаллое примерно 1 нс. Таким образом, соотношение (8.17) будет достаточно точным только для времени перемагничивания, составляющего не менее 5 нс.

Когда ферромагнетик находится под одновременным воздействием постоянного и синусоидального магнитных полей, то при определенной комбинации их параметров возникает резонансное поглощение. Оно обу-

словлено резонансными переходами между энергетическими уровнями, ответственными за ферромагнетизм. Если в выражении (8.9) положить $H_y = H_\perp \sin\omega t$, $H_x = H_0 >> H_\perp$, то для малых углов φ ($\sin\varphi \approx \varphi$) при $\alpha = 0$ и 90° из (8.13)

$$\frac{\partial^2\varphi}{\partial t^2} + \lambda\mu_0 \frac{\partial\varphi}{\partial t} + \gamma^2\mu_0 M\left(H_0 \pm \frac{2K}{\mu_0 M}\right)\varphi = \gamma^2\mu_0 M H_\perp \sin\omega t.$$

Откуда в установившемся режиме комплексный угол

$$\underline{\varphi} = \frac{\gamma^2\mu_0 M H_\perp}{\omega_0^2 - \omega^2 + j\omega\lambda}, \qquad (8.18)$$

где резонансная частота

$$\omega_0 = \gamma\sqrt{\mu_0 M(H_0 \pm H_k)}, \qquad (8.19)$$

а знаки «+» и «–» относятся к $\alpha = 0°$ и $\alpha = 90°$. Ширина резонансной кривой (8.18) на половинном уровне $\Delta\omega = \lambda$.

Следовательно, снятие частотных характеристик тонкой магнитной пленки при двух ориентациях переменной составляющей магнитного поля относительно оси легкого намагничивания позволяет измерить не только поле анизотропии H_k и магнитную вязкость λ, но и гиромагнитное отношение γ. При приложении переменного поля вдоль оси легкого намагничивания резонансная частота ω_0 может монотонно увеличиваться (при увеличении H_0), начиная только с минимального значения $\gamma\sqrt{\mu_0 M H_k}$. При переменном поле, направленном вдоль оси тяжелого намагничивания, резонансная частота имеет конечное значение только при $H_0 > H_k$; при $H_0 \to H_k$ происходит насыщение пленки и $\omega_0 \to 0$, что позволяет изучить резонансное поведение тонкой пленки при экстремально низких частотах.

§ 8.2. Магнитные пленки с нормальной анизотропией

Магнитоодноосными *пленками с нормальной анизотропией* называют пленки и тонкие кристаллы, у которых ось пленки легкого намагничивания перпендикулярна поверхности, а энергия анизотропии больше энергии поля рассеяния. Для их изготовления применяют кобальт, гексаферриты, ортоферриты, а также некоторые ферриты-гранаты, сплавы (MnBi) и поликристаллы с наведенной одноосной анизотропией. Особенностью такой .среды является разомкнутая доменная структура, в которой энергетически более выгодным становится отсутствие замыкающих поперечных доменов (рис. 8.4).

Структура с полосовыми доменами (*страйп-структура*) типа изображенных на рисунке 8.4, *а* для обычных условий представляет собой некоторую идеализацию, поскольку реальные домены в большинстве случаев не имеют прямых границ (рис. 8.4, *б*). Тем не менее их рассмотрение представляет интерес по нескольким причинам. Во-первых, данную доменную структуру можно образовать, направляя ось легкого намагничивания с небольшим отклонением от нормали (при этом доменные границы располагаются в направлении наклона ОЛН). Во-вторых, параллельные границы реально существуют в некоторых кристаллах. В-третьих, образующаяся управляемая дифракционная решетка может быть использована для записи и обработки информации, в частности при регистрации распределения энергии в лазерном луче.

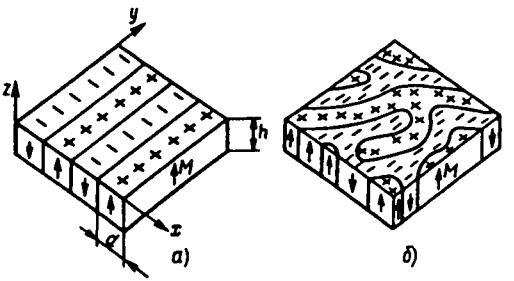


Рис. 8.4. Полосовая (*а*) и лабиринтная (*б*) доменные структуры с выходом магнитных зарядов на поверхность

Размеры полосовых доменов d в статическом режиме в отсутствие внешнего поля можно приближенно оценить из выражений для энергии полей рассеяния, создаваемых зарядами *одной* грани, и энергии доменных границ. При этом предполагают, что кристалл имеет бесконечные размеры в направлениях $\pm x$, $\pm y$ и $-z$ (см. рис 8.4, *а*) и поверхностная плотность магнитных зарядов

$$\sigma_м = \begin{cases} M_s, 2kd < x < (2k+1)d; \\ -M_s, (2k+1)d < x < 2(k+1)d, \end{cases} \tag{8.20}$$

где $k = 0, 1, 2, \ldots$.

Скалярный магнитный потенциал φ этой системы должен удовлетворять уравнению Лапласа

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial z^2} = 0 \tag{8.21}$$

и граничному условию

$$\left(\frac{\partial \varphi}{\partial z}\right)_{z=+0} - \left(\frac{\partial \varphi}{\partial z}\right)_{z=-0} = \sigma_м,$$

причем в силу симметрии поля относительно поверхности

$$\left(\frac{\partial \varphi}{\partial z}\right)_{z=+0} = -\left(\frac{\partial \varphi}{\partial z}\right)_{z=-0},$$

откуда

$$\left(\frac{\partial\varphi}{\partial z}\right)_{z=+0} = \frac{\sigma_м}{2}. \tag{8.22}$$

Подставив в формулу (8.21) выражение

$$\varphi = \sum_{n=1}^{\infty} A_n \sin\frac{n\pi x}{d} e^{n\pi z/d}, \qquad n = 1, 3, ..., \tag{8.23}$$

нетрудно убедиться, что оно является решением уравнения Лапласа. Коэффициенты A_n находят из граничных условий (8.20) и (8.22):

$$2\frac{\pi}{d}\sum_{n=1}^{\infty} nA_n \sin\frac{n\pi x}{d} = \begin{cases} M_s, 2kd < x < (2k+1)d, \\ -M_s, (2k+1)d < x < 2(k+1)d. \end{cases} \tag{8.24}$$

Если учесть, что

$$\int_{2kd}^{2(k+1)d} \sin^2\frac{n\pi x}{d}\,dx = \frac{d}{2\pi},$$

умножить обе части уравнения (8.24) на $\sin(n\pi x/d)$ и проинтегрировать, то получим

$$A_n = \frac{M_s}{2\pi n}\left[\int_{2kd}^{2(k+1)d} \sin\frac{n\pi x}{d}\,dx - \int_{(2k+1)d}^{2(k+1)d} \sin\frac{n\pi x}{d}\,dx\right] = \frac{2M_s d}{\pi^2 n^2}.$$

После подстановки этого выражения в формулу (8.23) при $z = 0$ получим

$$\varphi(x, 0) = \frac{2M_s d}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}\sin\frac{n\pi x}{d}.$$

Магнитостатическая энергия на единицу поверхности

$$W_м = -\frac{\mu_0}{2}\int_{V_0}\mathbf{MH}\,dV = \frac{\mu_0}{2}\int_{S_0}\sigma_м\varphi(x, 0)\,dx = \frac{\mu_0 M_s^2 d}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2 d}\int_0^d \sin\frac{n\pi x}{d}\,dx =$$

$$= \frac{2\mu_0 M_s^2 d}{\pi^3}\sum_{n=1}^{\infty}\frac{1}{n^3} \approx \frac{2\mu_0 M_s^2 d}{\pi^3}1{,}0518 = 0{,}0676\mu_0 M_s^2 d. \tag{8.25}$$

Если толщина h пластины (пленки) конечна, но настолько больше d, что магнитные заряды на верхней и нижней поверхностях не влияют друг на друга, то магнитостатическая энергия пластины

$$W_м \approx 4{,}2072\mu_0 M_s^2 d/\pi^3, \tag{8.26}$$

т. е. равна удвоенной энергии (8.25) полей рассеяния одной грани.

Если σ — поверхностная плотность энергии доменной стенки, то полная энергия доменных границ, отнесенная к единице поверхности пластины,

$$W_\sigma = \sigma h/d, \tag{8.27}$$

а плотность полной энергии пленки

$$W = W_м + W_\sigma = 4{,}2072\frac{\mu_0}{\pi^3}M_s^2 d + \frac{\sigma h}{d}.$$

Из условия минимума этого выражения равновесная ширина страйп-структуры

$$d = \sqrt{\frac{\pi^3 \sigma h}{4{,}2072\mu_0 M_s^2}}. \tag{8.28}$$

Например, для монокристаллической пленки с $\sigma = 2\cdot 10^{-7}$ Дж/см^2, $M_s = 10^3$ А/см и $h = 1$ мкм, $d = 10$ мкм.

Во внешнем поле **H**, направленном вдоль оси z, ширина доменов, намагниченных вдоль **H**, увеличивается до d_+, а ширина противоположно намагниченных доменов уменьшается до d_-.

Средняя намагниченность

$$M = \frac{d_+ - d_-}{d_+ + d_-}M_s,$$

энергия внешнего поля на единицу площади

$$W_н = -\mu_0 HMh,$$

магнитостатическая энергия на единицу поверхности

$$W_м = \frac{1}{2}\mu_0 M^2 h + 2\frac{\mu_0}{\pi^3}M_s d\sum_{n=1}^{\infty}\frac{1}{n^3}\sin^2\left[\frac{n\pi}{2}\left(1+\frac{M}{M_s}\right)\right](1-e^{-2\pi nh/d}).$$

В этом случае полная энергия изолированного домена

$$W = W_\sigma + W_н + W_м.$$

Устойчивому состоянию системы соответствуют намагниченность и период доменной структуры, удовлетворяющие условиям

$$\partial W/\partial M = 0 \text{ и } \partial W/\partial d = 0.$$

Результаты расчета численными методами приведены на рис. 8.5, где $\lambda = \sigma/(\mu_0 M_s^2)$ — *характеристическая длина* материала. Эти результаты хорошо совпали с результатами измерений.

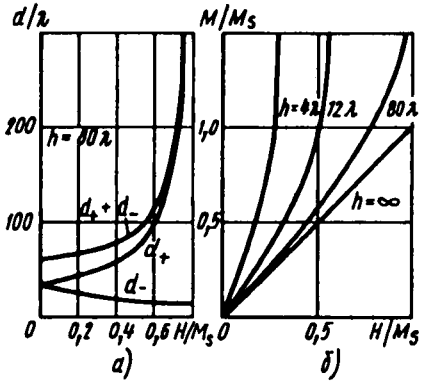


Рис. 8.5. Зависимости ширины полосовых доменов (*а*) и средней намагниченности (*б*) пленки с полосовыми доменами от напряженности внешнего магнитного поля

Силы, действующие на доменные стенки во внешнем магнитном поле, можно наглядно объяснить с помощью эквивалентных поверхностных токов.

Доменная стенка бесконечно малой толщины представляет собой поверхность разрыва вектора намагниченности, на которой он скачком меняется от значения $M_{1t} = M_s$ до $M_{2t} = -M_s$. Следовательно, для 180°-ной доменной стенки плотность эквивалентного поверхностного тока $J = M_{1t} - M_{2t} = 2M_s$, а общий ток $I = 2hM_s$, если высота стенки h. В полосовом домене токи в соседних стенках имеют противоположные направления, поэтому они отталкиваются.

Среднее магнитное поле B_σ, создаваемое одной из стенок на соседней стенке, легко оценить с помощью закона полного тока в двух предельных случаях. В первом случае, когда ширина домена $d \gg h$, ток в стенке можно считать сосредоточенным на оси, а силовые линии — окружностями. Тогда $B'_\sigma = \mu_0 2hM_s/(2\pi d)$ или $B'_\sigma/(\mu_0 M_s) = h/(\pi d)$, поскольку материал внутри домена находится в состоянии насыщения и $\mu_r = 1$. Во втором предельном случае, когда $d \ll h$, силовая линия практически прилегает к стенке и имеет длину $2h$. Поэтому $B''_\sigma = \mu_0 2hM_s/(2h)$ или $B''_\sigma/(\mu_0 M_s) = 1$. Постепенный переход от одного предельного случая к другому может быть выражен приближенной функцией

$$\frac{B_\sigma}{\mu_0 M_s} = \frac{1}{1 + \pi d/h}, \tag{8.29}$$

которая только на несколько процентов отличается от точного решения

$$\frac{B_\sigma}{\mu_0 B_s} = \frac{2}{\pi}\left[\operatorname{arctg}\frac{h}{d} - \frac{d}{2h}\ln\left(1 + \frac{h^2}{d^2}\right)\right], \tag{8.30}$$

полученного в предположении однородной намагниченности внутри и вне домена с бесконечно тонкой стенкой.

Сила, действующая на проводник с током, $\mathbf{F} = \mathbf{I} \times \mathbf{B}$. Поэтому отталкивающая сила двух стенок $F_м = IB_\sigma$, а сжимающая сила внешнего поля

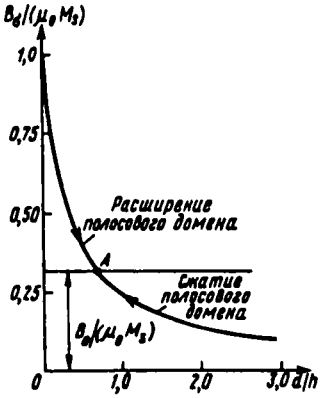


Рис. 8.6. Зависимость среднего магнитного поля стенки от ширины домена



Рис. 8.7. Лабиринтная структура доменов во внешнем магнитном поле

$F_s = IB_0$. Ширина d_- полосового домена с инверсной намагниченностью во внешнем поле определяется равновесием этих двух сил, когда среднее поле стенки B_σ равно внешнему полю B_0 (точка A на рис. 8.6).

Магнитное поле, создаваемое эквивалентным поверхностным током доменной стенки, не оказывает на нее силового воздействия, пока она прямая и бесконечно длинная, поскольку эти силы взаимно компенсируются. Однако гибкий и подвижной «проводник» с током, каковым можно считать стенку, подвержен неустойчивости на изгиб. Появление даже незначительного изгиба приводит к уплотнению силовых линий (увеличению плотности энергии) с внутренней стороны изгиба и разрежению — с внешней стороны, что приводит к дальнейшему развитию изгиба и увеличению длины полосового домена. Такая неустойчивость хорошо известна в плазме. Стабилизация такой неустойчивости в некотором диапазоне внешних полей связана с ростом энергии доменных стенок и объясняет устойчивую лабиринтную структуру полосовых доменов (рис. 8.7).

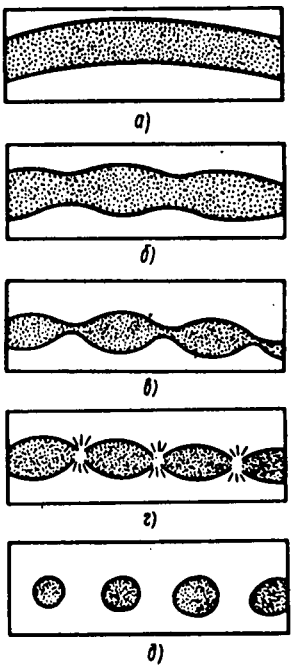


Рис. 8.8. Образование ЦМД в результате сжатия и разрыва полосового домена

Другой вид нестабильности, приводящий к полному разрушению полосовых доменов, обнаруживается при совместном действии постоянного поля и дополнительного импульсного внешнего магнитного поля. В импульсном поле сначала происходит волнообразное сжатие (рис. 8.8, *б*, *в*) полосовых доменов (рис. 8.8, *а*), потом их разрыв (рис. 8.8, *г*) и, наконец, симметричное сжатие малых участков. Если после этого выключить импульсное поле, то изолированные домены слегка расширяются, образуя в постоянном поле B_0 устойчивые объекты (рис. 8.8, *д*). Они имеют форму небольших цилиндров, пронизывающих пленку или пластину сверху донизу. Их называют *цилиндрическими магнитными доменами* (ЦМД).

§ 8.3. Цилиндрические магнитные домены

Устойчивость ЦМД определяется равновесием между сжимающей силой внешнего поля и направленного навстречу внутреннего магнитного давления, созданного собственной намагниченностью (встречной по отношению к внешнему полю). Круглая форма ЦМД определяется третьей силой — магнитным «поверхностным натяжением» стенки.

Применение ЦМД в приборах обработки информации определяется их способностью к перемещению, иначе — подвижностью. В таких приборах с записью и преобразованиями двоичной системы чисел наличию ЦМД соответствует «1», а его отсутствию — «0». Малый диаметр ЦМД (порядка микрометров) и большая скорость движения (до нескольких метров в секунду) позволяет довести скорость передачи информации до нескольких мегабит в секунду. Другим их преимуществом является реализация процессов передачи информации в двух измерениях, что позволяет сочетать параллельные и последовательные коды, логику и память системы.

Внешнюю сжимающую силу, действующую на ЦМД, можно определить, рассматривая его как цилиндр, на поверхности которого эквивалентный ток $I = 2hM_s$. Тогда действующая на доменную стенку и направленная по радиусу внутрь домена общая сила

$$F_0 = \pi DIB_0 = 2\pi DhM_sB_0, \qquad (8.31)$$

где $D = 2r$ — диаметр домена.

Расширяющую силу, обусловленную внутренним полем домена, можно найти, рассматривая ЦМД как замкнутую цепь (виток) с током I. В частном случае длинного ЦМД ($h \gg D$) поле равно полю длинного соленоида $B = \mu_0 I/h$, магнитный поток через соленоид $\Phi = \mu_0 \pi D^2 I/(4h) = \mu_0 \pi r^2 I/h$, индуктивность $L' = \Phi/I = \pi\mu_0 r/h$, энергия магнитного поля — $W = L'I^2/2$ и расширяющая сила

$$F'_{м} = \frac{\partial W}{\partial r} = \frac{1}{2} I^2 \frac{\partial L'}{\partial r} = \pi\mu_0 \frac{r}{n} I^2 = 2\pi DhM_s \mu_0 M_s = 2\pi DhM_s B'_{м}, \quad (8.32)$$

где $B'_{м} = \mu_0 M_s$ — среднее эффективное поле, вызываемое собственной намагниченностью домена и определяемое из сравнения (8.31) и (8.32).

В другом предельном случае широкого ЦМД ($D \gg h$), эквивалентного одновитковой катушке большого диаметра, по А. Зоммерфельду, индуктивность $L'' = \mu_0 r \ln(4\pi r/n)$ и сила

$$F''_{м} = \frac{1}{2} I^2 \frac{\partial L''}{\partial r} = \frac{\mu_0}{2}\left(1 + \ln\frac{4\pi r}{h}\right) I^2 = 2\pi DhM_s \frac{\mu_0 M_s}{\pi D/(hK)} = 2\pi DhM_s B''_{м},$$

где $K = 1 + \ln(4\pi r/h)$ — постоянная; $B''_{м} = \mu_0 hKM_s/(\pi D)$ — среднее эффективное поле в стенке.

В общем случае нормированное поле может быть аппроксимировано формулой

$$B_{м} = \frac{\mu_0 M_s}{1 + \pi D/(Kh)}, \quad (8.33)$$

которая в предельных случаях приводит к полученным значениям $B'_{м}$ и $B''_{м}$. При $\pi/K = 0{,}726$ выражение (8.33) дает хорошее приближение к точному решению для одиночного ЦМД:

$$\frac{B_{м}}{\mu_0 M_s} = \frac{2}{\pi}\left[\left(1 + \frac{D^2}{h^2}\right)^{1/2} E(k) - \frac{D}{h}\right],$$

где $E(k)$ — полный эллиптический интеграл второго рода для аргумента $k = (D/h)(1 + D^2/h^2)^{-1/2}$.

На рис. 8.9 пунктирной кривой изображены результаты приближенного решения, а сплошной кривой — точного решения.

Сжимающая сила доменной стенки может быть получена из выражения для ее энергии

$$W_{\sigma} = \sigma\pi Dh,$$

поскольку сила

$$F_{\sigma} = \frac{\partial W_{\sigma}}{\partial r} = 2\pi h\sigma = 2\pi DhM_s \frac{\sigma}{DM_s} = 2\pi DhB_{\sigma}, \quad (8.34)$$

а ее эффективное поле $B_{\sigma} = \sigma/(DM_s)$.

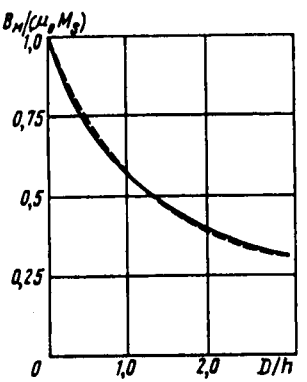


Рис. 8.9. Зависимость среднего расширяющего поля ЦМД от его диаметра

Статические свойства ЦМД. Равновесие ЦМД в статическом режиме определяет-

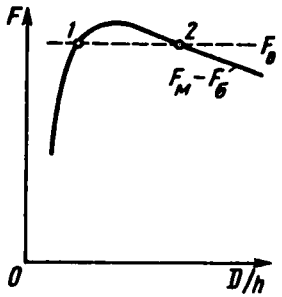


Рис. 8.10. Баланс сил в равновесном ЦМД

ся балансом сил $F_0 + F_\sigma = F_м$ или полей $B_0 + B_\sigma = B_м$, откуда с учетом (8.31), (8.33) и (8.34)

$$B_0 = \frac{\mu_0 M_s}{1 + 0{,}726\, D/h} - \frac{\sigma}{h M_s} \frac{1}{D/h}$$

или

$$\frac{B_0}{\mu_0 M_s} = \frac{1}{1 + 0{,}726\, D/h} - \frac{\lambda/h}{D/h}, \qquad (8.35)$$

где характеристическая длина $\lambda = \sigma/(\mu_0 M_s^2)$ для большинства ЦМД-материалов составляет порядка 1 мкм.

Уравнение (8.35) имеет два решения (для точек *1* и *2* на рис. 8.10). Первое решение соответствует режиму, при котором отклонение диаметра ЦМД в любую сторону от точки *1* вызывает силу, стремящуюся увеличить это отклонение. Второе решение отвечает условию устойчивого равновесия.

Устойчивость ЦМД возможна не при любом внешнем поле. Из рис. 8.10 видно, например, что по мере увеличения внешнего поля B_0, т. е. силы F_0, два решения *1* и *2* будут приближаться друг к другу, пока не сольются при достижении максимума кривой $F_м$—F_σ. В этой точке ЦМД исчезает (коллапсирует) и при больших внешних полях существовать не может. Соответствующие этой точке поле и диаметр домена определяют границу *радиальной неустойчивости* ЦМД.

Граничная точка соответствует максимуму правой части уравнения (8.35), где

$$-\frac{\partial(B_м - B_\sigma)}{\partial(D/h)} = \mu_0 M_s \left[\frac{0{,}726}{(1 + 0{,}726\, D_{рад}/h)^2} - \frac{\lambda/h}{(D_{рад}/h)^2} \right] = 0. \qquad (8.36)$$

Положительный корень этого квадратного уравнения определяет диаметр радиальной неустойчивости

$$\frac{D_{рад}}{h} = \frac{\sqrt{\lambda/0{,}726h}}{1 - \sqrt{0{,}726\lambda/h}}, \qquad (8.37)$$

а после подстановки его значения в (8.35) — поле радиальной неустойчивости

$$\frac{B_{рад}}{\mu_0 M_s} = \left(1 - \sqrt{0{,}726 \frac{\lambda}{h}}\right)^2. \qquad (8.38)$$

Другая граница устойчивости ЦМД связана с появлением *эллиптической неустойчивости*, при которой ЦМД превращается (вытягивается)

в полосовой домен. Количественную оценку этого явления можно приближенно получить, рассматривая баланс сил в полосовом домене конечной длины l и постоянной ширины d. Энергия двух боковых стенок такого домена $W_\sigma = 2\sigma hl$, и общая сжимающая (укорачивающая) сила этих стенок $F_\sigma = \dfrac{\partial W_\sigma}{\partial l} = 2\sigma h.$

На конец домена со стороны внешнего поля действует также сжимающая сила $F_0 = IB_0 d = 2dhM_s B_0$.

Тенденция к вытягиванию прямоугольного домена связана с полем собственной намагниченности и индуктивностью пары параллельных полосок высотой h, разнесенных на расстояние d. Магнитный поток, пронизывающий образуемый им виток, можно найти приближенно, интегрируя (8.29) по ширине домена: $\Phi = 2(\mu_0/\pi)lhM_s \ln(1+\pi d/h)$, откуда индуктивность на единицу длины

$$\frac{\partial L}{\partial l} = \frac{1}{I}\frac{\partial \Phi}{\partial l} = \frac{\mu_0}{\pi}\ln\left(1+\pi\frac{d}{h}\right)$$

и удлиняющая домен сила

$$F_м = \frac{1}{2}I^2\frac{\partial L}{\partial l} = \frac{1}{2}(2hM_s)^2\frac{\mu_0}{\pi}\ln\left(1+\pi\frac{d}{h}\right).$$

Внешнее поле $B_0 = B_{эл}$, соответствующее началу эллиптической неустойчивости, когда полосовой домен стягивается (в ЦМД вытягивается), найдем из условия $F_м = F_\sigma + F_0$ или

$$2\sigma h + 2dhM_s B_{эл} = \frac{1}{2}(2hM_s)^2(\mu_0/\pi)\ln(1+\pi d/h), \qquad (8.39)$$

причем d/h можно исключить из условия равновесия полосового домена с учетом (8.29):

$$B_{эл} = \frac{\mu_0 M_s}{1+\pi d/h}.$$

Тогда согласно (8.39),

$$\sigma h + (\mu_0/\pi)h^2M_s^2[1 - B_{эл}/(\mu_0 M_s)] = (\mu_0/\pi)h^2M_s^2\ln(\mu_0 M_s/B_{эл}).$$

В результате получим

$$\frac{\lambda}{h} = \frac{1}{\pi}\left(\ln\frac{\mu_0 M_s}{B_{эл}} - 1 + \frac{B_{эл}}{\mu_0 M_s}\right). \qquad (8.40)$$

Диапазон полей смещения, в котором могут существовать устойчивые ЦМД в пластине с данным значением λ/h, характеризуют кривые,

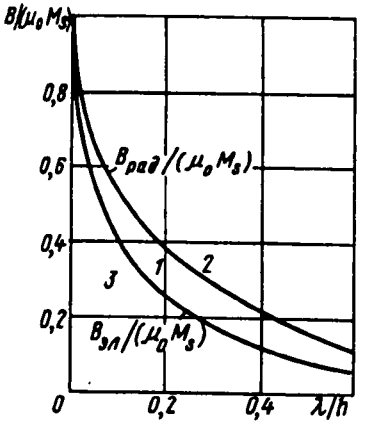


Рис. 8.11. Поле радиальной и эллиптической неустойчивости ЦМД:

1 — область устойчивых ЦМД, *2* — область радиальной неустойчивости, *3* — область эллиптической неустойчивости

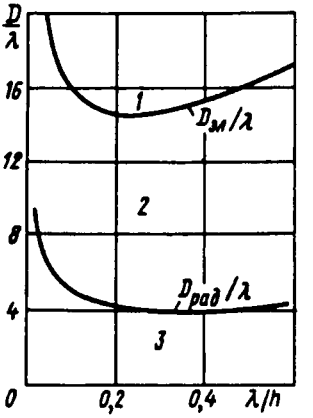


Рис. 8.12. Диаметры радиальной и эллиптической неустойчивости ЦМД:

1 — область эллиптической неустойчивости; *2* — область устойчивых ЦМД; *3* — область радиальной неустойчивости

построенные на рис. 8.11 по формулам (8.38) и (8.40). Несмотря на их приближенный характер, они отличаются всего на несколько процентов от результатов строгого анализа. Из рисунка 8.11 видно, что с уменьшением толщины пластины увеличивается отношение $B_{рад}/B_{эл}$ и уменьшается поле смещения B_0 при заданном значении $\mu_0 M_s$.

Диаметр эллиптической неустойчивости определяют из (8.35) после подстановки $B_{эл}$ из (8.40), а диаметр радиальной неустойчивости определяют из (8.37). Построенные таким образом кривые (рис 8.12) характеризуют диапазон устойчивых диаметров. Минимум кривых, казалось бы, обеспечивает минимальный диаметр ЦМД и максимальную плотность записи информации, однако этот режим не удается реализовать. Выбор диаметра осуществляют, исходя из условий надежного детектирования ЦМД и его надежного взаимодействия с магнитными аппликациями (при $D \gg h$ внешнее магнитное поле ЦМД мало, а при $D \ll h$ взаимодействие с поверхностными аппликациями слишком слабое).

В магнитостатике изолированного ЦМД, рассмотренной ранее, исходили из равновесия домена в однородном поле смещения, направленном навстречу вектору намагниченности внутри домена. Для продвижения ЦМД обычно используют *неоднородное* по поверхности пластины магнитное поле смещения. Например, в периодическом поле B_z (рис. 8.13) внешнее давление на ЦМД с двух сторон будет неодинаково, если он не расположен симметрично относительно минимума поля, который

соответствует устойчивому положению ЦМД. Одним из способов создания неоднородного поперечного магнитного поля в ЦМД-приборах является использование пермаллоевой магнитной аппликации (рис. 8.14, *а*), продольно намагниченной до насыщения M_0 внешним управляющим полем B_D (рис. 8.14, *б*). При этом на одном конце аппликации поперечная компонента поля управления B_D будет складываться с полем смещения B_0, а на другом — вычитаться из него (рис. 8.14, *в*), обеспечивая устойчивое положение ЦМД.

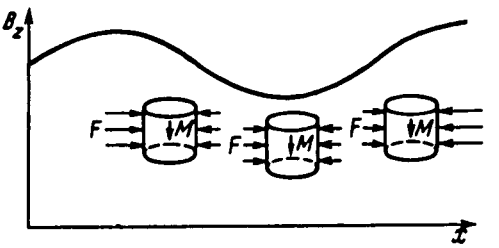


Рис. 8.13. Внешние силы F, действующие на ЦМД в периодическом поле

Если периодическое поле образовано бегущей волной, то ЦМД переносится с волной, оставаясь вблизи минимума поля. Бегущая волна в ЦМД-приборах может быть создана, например, за счет вращения управляющего продольного поля в структуре из цепочки пермаллоевых аппликаций. При этом минимумы поля B_z перемещаются вдоль цепочки аппликаций.

Скорость доменной стенки пропорциональна локальному магнитному полю, которое на нее действует. Это означает, что продвижение стенки связано с преодолением силы вязкого трения ($\beta v_n h$), пропорциональной скорости v_n (βv_n — сила трения на единицу поверхности стенки). Если локальное поле обозначить через B (оно включает и внешнее приложенное поле, и поля других стенок), то сила, действующая на единицу длины доменной стенки, $IB = 2hM_sB$. В стационарном режиме $\beta v_n h = 2hM_sB$, причем скорость, перпендикулярная поверхности доменной стенки, на линейном участке

$$v_n = \mu_\sigma B,$$

где $\mu_\sigma = 2M_s/\beta$ — *подвижность доменной стенки.*

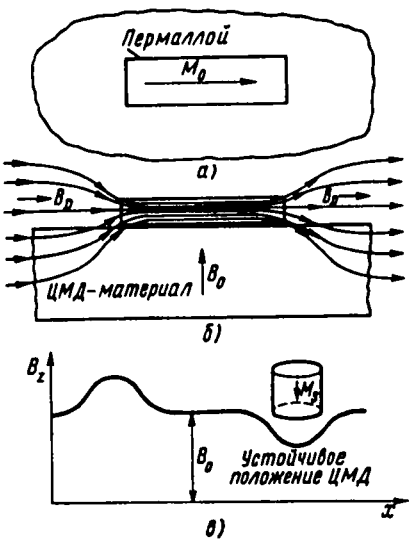


Рис. 8.14. Формирование магнитной ловушки вблизи пермаллоевой аппликации

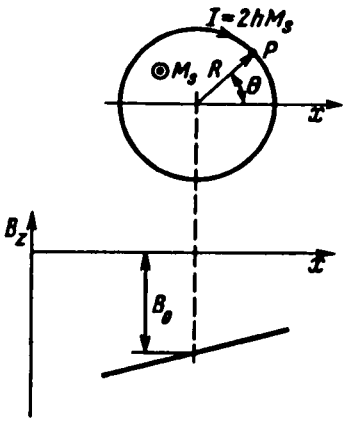


Рис. 8.15. ЦМД в поле с постоянным градиентом

Следовательно, силу вязкого трения, действующую на элемент длины стенки dl, можно представить в виде

$$dF_n = \beta v_n h\, dl = 2hM_s(v_n/\mu_\sigma)\, dl.$$

Если v_x — скорость движения ЦМД (рис. 8.15), то составляющая скорости, нормальная к стенке, $v_n = v_x \cos\theta$, а горизонтальная составляющая силы вязкого трения, действующая на элемент $dl = R\,d\theta$, $dF_x = \cos\theta\, dF_n$. Поэтому действующая на ЦМД общая сила трения

$$F_x = 2\int_{-\pi/2}^{\pi/2} 2hM_s \frac{v_x}{\mu_\sigma} R\cos^2\theta\, d\theta = 2hM_s \frac{v_x}{\mu_\sigma}\pi R. \qquad (8.41)$$

Если ЦМД расположен в поле с постоянным градиентом таким образом, что в центре поле равно полю смещения B_0 (см. рис. 8.15), то такой домен должен двигаться вправо. В точке P радиальная сила, действующая со стороны внешнего поля на элемент длины $R\,d\theta$ доменной стенки, может быть получена (если пренебречь коэрцитивной силой) аналогично тому, как это определено в (8.31):

$$dF_0 = 2hM_s\left(B_0 + \frac{\partial B_z}{\partial x} R\cos\theta\right) R\, d\theta.$$

Если предположить, что ЦМД при этом сохраняет круглую форму и радиус, соответствующие однородному полю B_0, то общая направленная вправо сила

$$F_0 = 2hM_s\int_0^{2\pi}\left(B_0 + \frac{\partial B_z}{\partial x} R\cos\theta\right) R\cos\theta\, d\theta = 2hM_s\pi R^2 \frac{\partial B_z}{\partial x}. \qquad (8.42)$$

В стационарном режиме силы (8.41) и (8.42) должны уравновешивать друг друга. Тогда скорость домена

$$v_x = \mu_\sigma \frac{\partial B_z}{\partial x} R = \mu_\sigma \Delta B_z,$$

где $\Delta B_z = R\dfrac{\partial B_z}{\partial x}$.

С учетом коэрцитивной силы

$$v_x = \mu_\sigma(\Delta B_z - B_c).$$

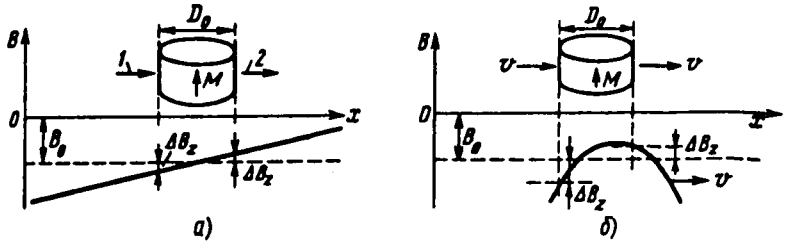


Рис. 8.16. Движение ЦМД в поле, зависящем только от x (*а*), и в движущейся параболической ловушке (*б*):

1 — часть доменной стенки, находящейся в стабилизирующем градиенте; *2* — то же, в дестабилизирующем градиенте, v — скорость движения ловушки

Динамические свойства ЦМД. Форма ЦМД существенно зависит от природы градиента магнитного поля, в котором движется доменная стенка. Например, если ДМД начинает двигаться и расширяется в поле с постоянным градиентом (рис. 8.16, *а*), причем с одной стороны поле оказывается больше начального B_0, а с другой стороны — меньше, то левая сторона начинает двигаться в сторону более слабого поля (как при коллапсе), а правая (передняя) движется с ускорением в сторону более сильных полей. Поэтому ЦМД не может оставаться круглым, расширяется и не будет устойчивым на своей передней стороне.

Если ЦМД диаметром D_0, соответствующим полю B_0, передвигается вместе с движущейся магнитной ловушкой в виде параболоида вращения (рис. 8.16, *б*), отставая от центра ловушки на некоторое расстояние, то движение происходит в «правильном» направлении, при котором градиент локального магнитного поля обеспечивает устойчивость на обеих сторонах домена. Локальное поле, уменьшающееся в направлении движения доменной стенки, представляет собой разность между полем потенциальной ямы и полем доменной стенки (последнее равно по модулю и противоположно по направлению полю B_0, необходимому для стабилизации диаметра D_0). Если скорость любой стороны ЦМД отклонится от «правильной», то возникнет стабилизирующий эффект, при котором изменившееся поле на доменной стенке восстановит «правильную» скорость (соответствующую D_0).

Перед рассмотрением динамической устойчивости ЦМД необходимо обратить внимание на то, что диаметры радиальной и эллиптической устойчивости ЦМД, расположенного в минимуме поля даже покоящейся ловушки, отличаются от их значений во внешнем равномерном поле (B_0). Так, при определении нового диаметра *радиальной неустойчивости* в уравнении (8.35) необходимо вместо B_0 подставить B_z и учесть его зависимость от координаты $r = D/2$ (рис. 8.17, *а*):

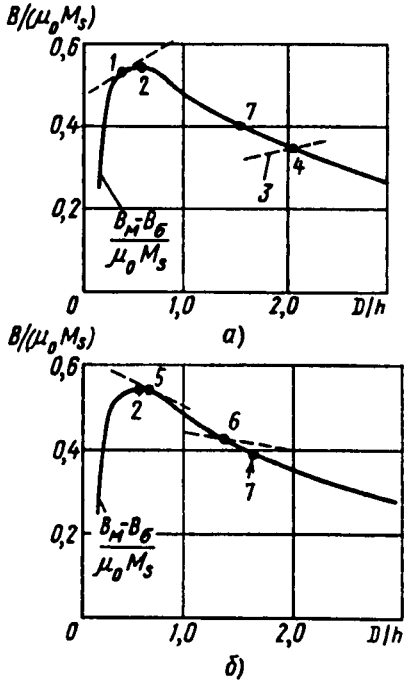


Рис. 8.17. Положение точек неустойчивости ЦМД в магнитной ловушке с положительным (*а*) и отрицательным (*б*) градиентами:

1 — радиальная неустойчивость в поле смещения с положительным градиентом; *2* — то же, в отсутствии градиента; *3* — поле смещения с положительным градиентом; *4* — стабильное решение при диаметре ЦМД, большем диаметра эллиптической неустойчивости; *5* — радиальная неустойчивость в поле смещения с отрицательным градиентом, возникающая при больших диаметрах ЦМД; *6* — эллиптическая неустойчивость в поле смещения с отрицательный градиентом, возникающая при $D < D_{эл}$; *7* — эллиптическая неустойчивость в отсутствии градиента

$$\frac{1}{1+0{,}726\,D/h} - \frac{\lambda/h}{D/h} = \frac{B_z}{\mu_0 M_s}. \qquad (8.43)$$

Поэтому условием коллапса будет теперь не равенство нулю производной (8.36) от левой части, а равенство производных от левой и правой частей уравнения (8.43). Из рис. 8.17, *а* видно, что коллапс наступает при меньших диаметрах, если происходит в поле с положительным градиентом, и при бóльших диаметрах, если осуществлен отрицательный градиент (рис. 8.17, *б*).

На максимальные значения D_{max}, определяющие возникновение *эллиптической неустойчивости* в неравномерном поле, изменение знака градиента оказывает противоположное влияние. При больших λ/h влияние отрицательных градиентов на радиальную неустойчивость становится очень сильным. Благодаря этому уже при небольшом отрицательном градиенте поля смещения (точка *A* на рис. 8.18) само существование ЦМД становится невозможным.

Здесь «положительный градиент» предполагается в ловушке, изображенной на рис. 8.16, *б*, а «отрицательный» — в ловушке, представленной на рис. 8.19. Из последнего рисунка видно, что движение ЦМД влево вызывает уменьшение поля смещения на левой части доменной стенки и расширение домена; правая же часть доменной стенки будет продвигаться в увеличивающееся поле и вызывать сжатие домена. Таким образом, положение ЦМД оказывается неустойчивым, причем если ЦМД движется влево, то ловушка ускоряет его движение. Следовательно,

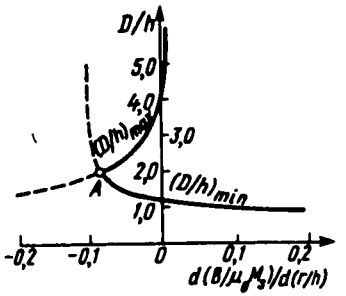


Рис. 8.18. Диаметры радиальной $(D/h)_{min}$ и эллиптической $(D/h)_{max}$ неустойчивости в неравномерном поле при $\lambda/h = 0{,}3$

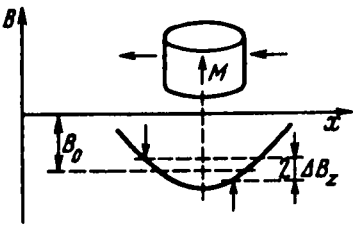


Рис. 8.19. Магнитная ловушка с отрицательным градиентом во всех точках стенки ЦМД

зависимости на рис. 8.18 в области отрицательных градиентов могли бы быть справедливы лишь в том случае, когда положение ЦМД стабилизировалось бы некоторой внешней силой.

Рассмотренные явления объясняют ситуацию в поле с постоянным градиентом, изображенным на рис. 8.16, *а*, когда левая сторона ЦМД стабилизируется положительным градиентом поля, а правая сторона, находясь в отрицательном (по отношению к центру домена) градиенте поля, оказывается дестабилизированной. Вместе с тем стабилизирующее действие положительного градиента обеспечивает устойчивое перемещение ЦМД вместе с движущейся ловушкой. Основные черты этого динамического процесса помогает рассмотреть рис. 8.20.

В состоянии равновесия на доменную стенку неподвижного ЦМД действует локальное поле смещения B_0, его радиус $R_0 = D/2$ определяется уравнением (8.35). При движении ЦМД будет отставать от ловушки на постоянное расстояние x', причем доменная стенка будет находиться в ускоряющем локальном поле. Следовательно, во-первых, различие локального поля в разных точках границы ЦМД ($\Delta B_L \neq \Delta B_R$) несовместимо с условием сохранения доменом при движении круглой формы; во-вторых, движущийся ЦМД уменьшается в размере. Это следует из того, что боковые доменные стенки в точках P_1 и P_2 не имеют нормальной составляющей ско-

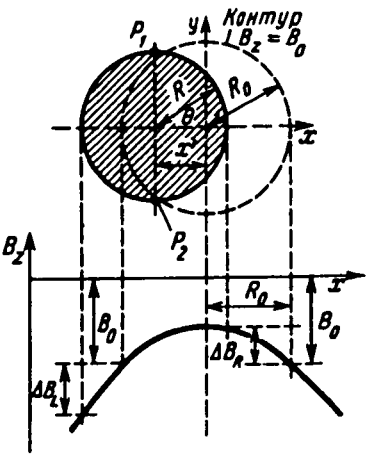


Рис. 8.20. Движение ЦМД вместе с магнитной ловушкой

рости и, следовательно, находятся в равновесии со своими локальными полями, приблизительно равными B_0. Если бы радиус движущегося домена оставался равным R_0, а поле в точках P_1 и P_2 стало бы больше B_0, то это вызвало бы сжатие домена.

Уменьшение радиуса при движении увеличивает магнитную энергию ЦМД. Оценим этот эффект при малых изменениях радиуса, исходя из аналогичного (8.31) выражения для действующей на стенку ЦМД силы

$$F = 2hM_s B_0 2\pi R_0 = -\frac{\partial W}{\partial R},$$

где W — энергия домена.

Следовательно, небольшое изменение радиуса ΔR вызывает изменение энергии домена

$$\Delta W = -2hM_s B_0 2\pi R_0 \Delta R.$$

При малом $(x')^2 \approx R_0^2 - R^2 = (R_0 - R)(R_0 + R)$

$$\Delta R \approx -(x')^2/2R_0 \quad \text{и} \quad \upsilon_x \approx \mu_\sigma(\partial B_z/\partial x')x'.$$

Поэтому прирост энергии домена

$$\Delta W \approx \frac{\pi 2hM_s B_0 \upsilon_x^2}{\mu_\sigma^2(\partial B_z/\partial x')^2} = \frac{m_д \upsilon_x^2}{2}$$

оказывается пропорциональным квадрату скорости, где

$$m_д = \frac{4\pi hM_s B_0}{\mu_\sigma^2(\partial B_z/\partial x')^2}$$

— эффективная «магнитная масса» движущегося домена.

Наличие массы у движущегося домена характеризует его инерцию, т. е. для изменения скорости ЦМД, движущегося вместе с магнитной ловушкой, требуется конечное время. Это время можно оценить, если записать полное уравнение движения ЦМД в виде суммы силы инерции $F_i = -\frac{\partial}{\partial x}\Delta W$, силы вязкого трения (8.41) и силы F_B, создаваемой локальным полем в ловушке:

$$F_B - \frac{4\pi hM_s B_0}{\mu_\sigma^2(\partial B_z/\partial x')^2}\upsilon_x \frac{\partial \upsilon_x}{\partial x} - 2\pi hM_s R\frac{\upsilon_x}{\mu_\sigma} = 0.$$

Учитывая, что

$$\upsilon_x \frac{\partial \upsilon_x}{\partial x} = \frac{\partial x}{\partial t}\frac{\partial \upsilon_x}{\partial x} = \frac{\partial \upsilon_x}{\partial t},$$

полученное уравнение можно переписать в виде

$$\tau \frac{\partial v_x}{\partial t} + v_x = F_B(t),$$

где

$$\tau \approx 2B_0/[\mu_0 R(\partial B_z/\partial x)^2]$$

— *постоянная времени инерции*, обусловленная изменением радиуса ЦМД и зависящая не только от его свойств, но и от формы ловушки.

В некоторых материалах ЦМД коллапсируют не при ожидаемом значении $B_{рад}$, а при других и притом разных полях смещения. Такие домены с повышенной радиальной устойчивостью называют *твердыми*. Их поведение связывают с более сложной, неоднородной структурой доменной стенки. Если в обычной 180°-ной стенке (*простой блоховской*) поворот намагниченности происходит вокруг радиуса, то в стенках твердых доменов имеются участки (*блоховские линии*), где намагниченность поворачивается вокруг направления, касательного к стенке и лежащего в плоскости пленки. Эти участки обладают более высокой плотностью энергии и имеют более сложную зависимость от радиуса ЦМД и взаимного расстояния друг от друга. Динамические свойства твердых и нормальных доменов сильно отличаются. Твердые ЦМД коллапсируют при существенно более высоких полях смещения. Низкая подвижность твердых ЦМД может сильно уменьшить быстродействие ЦМД-устройств, если не принять меры к предупреждению их возникновения.

§ 8.4. Доменная стенка как нелинейная волна намагниченности

Перемещение доменной границы практически не сопровождается смещением атомов кристалла со своих мест, происходит лишь поворот атомных спинов. Сопровождающий движение стенки процесс носит волновой характер, при котором распределение намагниченности зависит не только от координаты, но и от времени. Например, для 180°-ной стенки, движущейся со скоростью v вдоль оси x, намагниченность в поперечном направлении $M_z = M_s(x - vt)$.

Феноменологический подход к описанию динамики намагниченности исходит из уравнения Ландау—Лифшица (6.5). Рассмотрение спиновых волн, исходя из этого уравнения, было проведено в § 6.6 в предположении их малой амплитуды. Это позволило упростить расчет за счет линеаризации уравнения (6.5). Однако, в случае движения доменной границы, изменение намагниченности нельзя считать малым. Например, для 180°-ной стенки вектор намагниченности поворачивается на 180° и его изменение становится порядка намагниченности насыщения. Поэтому задача оказывается существенно нелинейной.

Рассмотрим нелинейную динамику намагниченности в ферромагнетике с наиболее простой — одноосной — анизотропией, предполагая, что в нем имеется однородное внешнее магнитное поле H, направленное параллельно оси анизотропии (ось z) и намагниченность M зависит от единственной координаты ξ. При этом плотность магнитной энергии

$$W = \frac{\alpha}{2}\left(\frac{\partial M}{\partial \xi}\right)^2 - \frac{\beta}{2}M_z^2 - \mu_0 H M_z, \qquad (8.44)$$

если ее отсчитывать от энергии основного состояния магнетика в магнитном поле и не учитывать магнитодипольные взаимодействия (α и β — постоянные обменного взаимодействия и анизотропии).

Напомним, что в общем случае эффективное магнитное поле $\mathbf{H}_{эф} = \frac{1}{\mu_0}\frac{\delta W}{\delta \mathbf{M}}$ представляет собой функциональную производную от энергии по магнитному моменту. Поэтому с учетом выражения (8.44) уравнение (6.5) преобразуем к виду

$$\frac{1}{\gamma}\frac{\partial \mathbf{M}}{\partial t} + \alpha\left[\mathbf{M}\frac{\partial^2 \mathbf{M}}{\partial \xi^2}\right] + \beta[\mathbf{M}\mathbf{e}_z](\mathbf{M}\mathbf{e}_z) + \mu_0[\mathbf{M}\mathbf{e}_z]H = 0. \qquad (8.45)$$

Зададим направление вектора намагниченности угловыми переменными θ и φ сферической системы координат (рис. 8.21). Тогда

$$M_z = M_s\cos\theta, \quad \underline{\psi} = M_x + jM_y = M_s\sin\theta e^{j\varphi}, \qquad (8.46)$$

и уравнение (6.5) примет вид

$$\sin\theta\frac{\partial\theta}{\partial t} = -\frac{\gamma}{M_s}\frac{\partial W}{\partial\varphi}, \quad \sin\theta\frac{\partial\varphi}{\partial t} = \frac{\gamma}{M_s}\frac{\partial W}{\partial\theta}, \qquad (8.47)$$

а уравнение (8.44) —

$$W = \frac{1}{2}\alpha M_s^2[(\nabla\theta)^2 + \sin\theta(\Delta\varphi)^2] + \\ + \frac{1}{2}\beta M_s^2\sin^2\theta + \mu_0 M_s H(1-\cos\theta). \qquad (8.48)$$

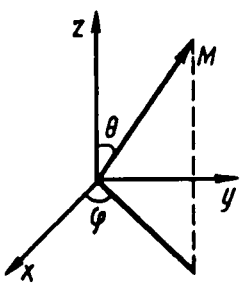


Рис. 8.21. Полярные координаты вектора намагниченности

Подставив уравнение (8.48) в (8.47), получим

$$l_0^2\Delta\theta - [1 + l_0^2(\nabla\varphi)^2]\sin\theta\cos\theta + \frac{1}{\omega_0}\left(\frac{\partial\varphi}{\partial t} - \omega_н\right)\sin\theta = 0,$$

$$l_0^2\,\mathrm{div}(\sin^2\theta\nabla\varphi) - \frac{1}{\omega_0}\frac{\partial\theta}{\partial t}\sin\theta = 0\ ^*.$$

---

* Здесь $l_0 = \sqrt{\alpha/\beta}$, $\omega_0 = \gamma m_0\beta M_s$.

С учетом зависимости намагниченности только от одной пространственной координаты ξ

$$\alpha\frac{\partial^2\theta}{\partial\xi^2}-\left[\beta+\alpha\left(\frac{\partial\varphi}{\partial\xi}\right)^2\right]\sin\theta\cos\theta+\frac{1}{\gamma M_s}\sin\theta\frac{\partial\varphi}{\partial t}=\mu_0\frac{H}{M_s}\sin\theta; \qquad (8.49)$$

$$\alpha\frac{\partial}{\partial\xi}\left(\sin^2\theta\frac{\partial\varphi}{\partial\xi}\right)-\frac{1}{\gamma M_s}\sin\theta\frac{\partial\theta}{\partial t}=0. \qquad (8.50)$$

При рассмотрении нелинейной волны намагниченности, движущейся с постоянными скоростью v и частотой прецессии ω, будем искать общее решение полученных уравнений в виде

$$\theta=\theta(\xi-v\,t),\quad \varphi=\varphi(\xi-v\,t)+\omega t.$$

Тогда уравнения (8.49) и (8.50) можно свести к системе обыкновенных дифференциальных уравнений:

$$l_0^2\frac{\mathrm{d}^2\theta}{\mathrm{d}\xi^2}-\left[1+l_0^2\left(\frac{\mathrm{d}\varphi}{\mathrm{d}\xi}\right)^2\right]\sin\theta\cos\theta-\frac{v}{\omega_0}\sin\theta\frac{\mathrm{d}\varphi}{\mathrm{d}\xi}-\frac{\omega}{\omega_0}\sin\theta=0; \qquad (8.51)$$

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\left(l_0^2\sin^2\theta\frac{\mathrm{d}\varphi}{\mathrm{d}\xi}-\frac{v}{\omega_0}\cos\theta\right)=0. \qquad (8.52)$$

Для граничных условий при $\xi=\pm\infty$ в виде

$$\theta=0,\quad \frac{\mathrm{d}\varphi}{\mathrm{d}\xi}<\infty \qquad (8.53)$$

интегрирование (8.52) позволяет определить связь между угловыми координатами:

$$l_0\frac{\mathrm{d}\varphi}{\mathrm{d}\xi}=-\frac{1}{\cos^2(\theta/2)}\frac{v}{v_m}, \qquad (8.54)$$

где $v_m=2l_0\omega_0$ — минимальная фазовая скорость спиновых волн.

Подставив выражение (8.54) в (8.51), получим обыкновенное дифференциальное уравнение

$$l_0^2\frac{\mathrm{d}^2}{\mathrm{d}\xi^2}\left(\frac{\theta}{2}\right)+\sin\left(\frac{\theta}{2}\right)\left(\cos\frac{\theta}{2}+2\cos^3\frac{\theta}{2}+\frac{v}{v_m^2}\cos^{-3}\frac{\theta}{2}-\frac{\omega}{\omega_0}\right)=0.$$

Первое интегрирование этого уравнения с учетом граничных условий (8.53) приводит к уравнению

$$l_0\frac{\mathrm{d}\theta}{\mathrm{d}\xi}=2\,\mathrm{tg}\frac{\theta}{2}\sqrt{\cos^4\frac{\theta}{2}-\frac{v^2}{v_m^2}-\frac{\omega}{\omega_0}\cos^2\frac{\theta}{2}}.$$

После второго интегрирования получим

$$\operatorname{tg}^2\frac{\theta}{2}=\frac{(\varkappa l_0)^2}{\Omega\operatorname{ch}^2\varkappa(\xi-v t)-(\Omega-\Omega_1)/2}, \tag{8.55}$$

где $(\varkappa l_0^2)=1-\omega/\omega_0-(v/v_m)^2$; $\Omega=\omega/\omega_0=\sqrt{\Omega_1^2+4(\varkappa l_0)^2 v^2/v_m^2}$ — нормированная частота прецессии магнитного момента в нелинейной спиновой волне; $\Omega_1=\omega/\omega_0+2v^2/v_m^2$ — нормированная частота прецессии магнитного момента в солитоне. Выражение для фазы этой прецессии можно получить после подстановки решения (8.55) в (8.54) и его интегрирования:

$$\varphi=\omega t-\frac{v^2}{v_m^2}\frac{\xi-v t}{l_0}+\operatorname{arctg}\left[\sqrt{\frac{\Omega-\Omega_1}{\Omega+\Omega_1}}\operatorname{th}\varkappa(\xi-v t)\right]. \tag{8.56}$$

Полученное решение уравнения Ландау—Лифшица в виде локализованной в пространстве функции (8.55) описывает движущееся со скоростью v возмущение на фоне однородного состояния намагниченности. Вторым параметром процесса является частота ω прецессии вектора **M** в системе отсчета, движущейся вместе с центром тяжести возбуждения (со скоростью v). Параметры v и ω являются независимыми и полностью характеризуют решение (8.55), дающее математическое описание локализованной в пространстве волны перемагничивания, обладающей всеми свойствами солитона — *магнитного солитона* в одноосном ферромагнетике.

Для неподвижного солитона (v = 0) в диапазоне частот $0<\omega<\omega_0$ выражение (8.55) упрощается и приобретает стандартный вид

$$\operatorname{tg}\frac{\theta}{2}=\sqrt{\frac{\omega_0}{\omega}-1}\,\operatorname{ch}^{-1}\left(\frac{\xi}{l_0}\sqrt{1-\frac{\omega}{\omega_0}}\right). \tag{8.57}$$

В частности, при малых частотах прецессии $\omega \ll \omega_0$ магнитный солитон представляет собой широкую область практически перемагниченного магнетика ($\theta\to\pi$), отделенную от остальной части ферромагнетика двумя доменными границами разных знаков (рис. 8.22). Перемагничивание среды в неподвижном солитоне связано с дополнительной прецессией вектора намагниченности, происходящей синхронно во всем объеме. Связанное состояние двух доменных границ является проявлением динамики намагниченности.

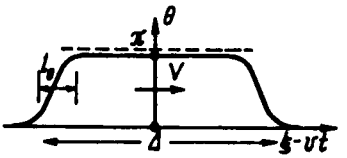


Рис. 8.22. Распределение намагниченности в неподвижном солитоне при $\omega \ll \omega_0$

Из выражения (8.57) следует, что при малых частотах длина области перемагничивания $\Delta=l_0\ln(\omega_0/\omega)$. Поэтому

при $\omega = 0 = v$ решение описывает две бесконечно удаленные доменные границы ($\Delta \to \infty$), где

$$\theta = 2\operatorname{arctg} e^{(x-x_0)/l_0}, \qquad (8.58)$$

если x отсчитывать от начала переходной области (стенки), а через x_0 обозначить центр переходной области. Поскольку направление намагниченности θ в доменной границе (8.58) изменяется на 180°, то это не что иное, как 180°-ная доменная стенка, толщина которой порядка магнитной длины l_0 (в области относительно низких частот рассматриваемое описание предполагает, что она много больше межатомных расстояний в кристаллической решетке). Поперечная составляющая вектора намагниченности в такой стенке

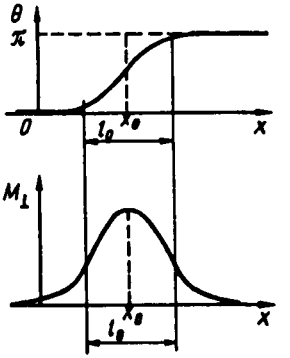


Рис. 8.23. Структура неподвижной доменной стенки

$$M_\perp = \sqrt{M_x^2 + M_y^2} = M_s \sin\theta = \frac{M_s}{\operatorname{ch}[(x-x_0)/l_0]}. \qquad (8.59)$$

Функции (8.58) и (8.59) совпадают с подобными же зависимостями и в отсутствии внешнего магнитного поля. Они и их графики (рис. 8.23) также описывают локальное возмущение, которое называется солитоном. Таким образом, неподвижная плоская 180°-ная доменная стенка представляет собой простейший одномерный солитон поля намагниченности.

В общем случае при отличных от нуля скорости перемещения солитона v и частоте ω прецессии магнитного момента возникающая в этой области локальная неоднородность намагниченности характеризуется поворотом $\Delta\varphi$ вектора **M**, дополнительным к его повороту в обыкновенной спиновой волне. В формуле (8.56) первые два члена описывают прецессию в спиновой волне, а последний — дополнительный поворот

$$\Delta\varphi = -2\operatorname{arctg}\sqrt{\frac{\Omega - \Omega_1}{\Omega + \Omega_1}},$$

который может изменяться в пределах от 0 при $v = 0$, $\omega > 0$ до π при $v = 0$, $\omega < 0$.

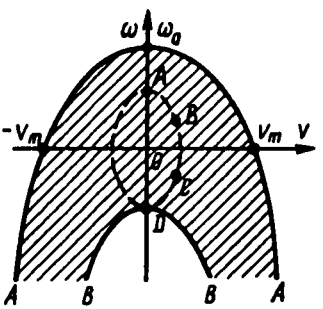


Рис. 8.24. Область существования солитона

По мере увеличения скорости перемещения и прецессии от начальных нулевых значений область Δ локализации солитона сужается и достигает минимального размера порядка l_0 при $\varkappa l_0 = \Omega$. Этому условию соответствует эллипс (рис. 8.24), охватывающий начало координат на плос-

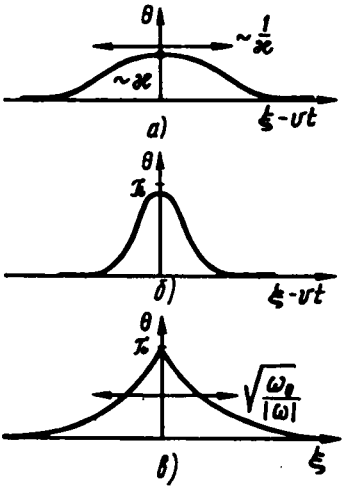


Рис. 8.25. Магнитные солитоны

кости v–ω и имеющий оси v_m и ω_0. Внутри него лежат значения v и ω, для которых солитон ограничен двумя четко выраженными разделенными, но связанными друг с другом доменными границами. Движущийся солитон здесь имеет сплющенную форму, мало отличающуюся от формы неподвижного солитона (см. рис. 8.22).

Вне этого эллипса при дальнейшем увеличении v и $\omega > 0$ характер магнитного солитона существенно изменяется в связи с полным исчезновением индивидуальности двух доменных границ. Парабола *AA* на рис. 8.24 характеризует закон дисперсии спиновых волн и является границей области существования солитонов. При приближении к этой границе, где $\varkappa l_0 = 0$, амплитуда солитона убывает пропорционально $\varkappa l_0$, область его локализации становится согласно уравнению (8.55) приблизительно равной $1/\varkappa$ и безгранично возрастает с уменьшением $\varkappa$. Примерная форма такого малоамплитудного солитона ($\varkappa l_0 \ll 1$) приведена на рис. 8.25, *а*.

Совершенно иной вид имеет солитон для противоположного направления прецессии ($\omega < 0$). Здесь для медленных солитонов большой амплитуды при увеличении $|\omega|$ область локализации $\Delta \to l_0\sqrt{\omega_0/\omega}$ уменьшается до значений, меньших l_0 (рис. 8.25, *б*). Неподвижный солитон ($v = 0$) имеет вид, приведенный на рисунке 8.25, *в*, если l_0 намного превосходит межатомное расстояние a. При $|\omega|/\omega_0 < (l_0/a)^2 + (v/v_m)^2$ изменение намагниченности становится в промежутке между атомами столь заметным ($\partial\theta/\partial\xi > a^{-1}$), что низкочастотное приближение не пригодно. На рисунке 8.24 эта область отделяется параболой *BB*.

Допустимые значения скорости перемещения магнитного солитона лежат в конечном интервале — $v_m < v < v_m$. Скорости вне этого интервала соответствуют нелинейным спиновым волнам (θ = const), однако они в рассматриваемых условиях оказываются неустойчивыми. При значениях v, ω, лежащих вблизи вершины параболы *AA* (см. рис. 8.24), исходные уравнения Ландау—Лившица (8.49) и (8.50) могут быть преобразованы в нелинейное уравнение Шредингера, т. е. форма магнитного солитона в этой области совпадает с представленной на рис. 7.13, *б*.

Когда фаза φ меняется достаточно медленно и первым слагаемым в уравнении (8.49) можно пренебречь, система уравнений (8.49) и (8.50) преобразуется к стандартной форме уравнения синус-Гордона (10.45). Соответствующие им солитоны рассмотрены в § 10.9.

§ 8.5. Магнитооптические явления в ферромагнитных материалах

Оптические методы записи и считывания информации рассматриваются в настоящее время как более перспективные по сравнению с традиционными магнитными методами, используемыми для внешних запоминающих устройств ЭВМ, поскольку они могут обеспечить гораздо бóльшую емкость (5000 вместо 150 Мбайт на одной рабочей поверхности) при одинаковой скорости записи. Повышение плотности записи обусловлено возможностью сфокусировать лазерный луч в пятно, соизмеримое с длиной световой волны. Оптические дисковые запоминающие устройства находят широкую сферу применения — от банков данных до персональных ЭВМ. В большинстве разработанных дисковых устройств не предусмотрена перезапись информации. Однако уже существуют коммерческие оптические диски с перезаписью, которые обладают значительно бóльшими функциональными возможностями. Одно из направлений их разработки основано на использовании магнитооптических явлений.

Запись информации в магнитооптической регистрирующей среде может происходить в результате локального нагрева ферромагнитного материала оптическим излучением выше точки Кюри. После остывания материала направление намагниченности оказывается противоположным первоначальному под действием размагничивающего поля соседних участков либо внешнего магнитного поля. При считывании записанной информации на просвет применяется эффект Фарадея, а при считывании отраженным лучом — эффект Керра. Стирание записанной информации может производиться внешним магнитным полем. Значительно упрощена система модуляции, считывания и записи в оптических дисковых запоминающих устройствах. Применение полупроводниковых лазеров уменьшает потребляемую мощность и габариты. Скорость обмена данных в магнитооптических дисках доходит до 5 МБит/с. Ее ограничивает время перемагничивания магнитной катушки.

Магнитооптическими называют явления, в которых проявляется влияние намагничивания на излучение и распространение света. К ним, в частности, относят эффекты Фарадея и Коттона—Мутона, а также магнитооптический эффект Керра, который обнаружен только в ферромагнитных материалах. Он заключается во влиянии намагниченности ферро-

магнитного зеркала на поляризацию отраженного от его поверхности света. Например, при некоторых условиях угол поворота плоскости поляризации отраженного света $\alpha_K = K_K M$, оказывается пропорциональным намагниченности образца и *постоянной Керра* K_K, зависящей от частоты света и температуры.

В оптической области спектра, уже начиная с длинноволновых инфракрасных лучей, магнитная проницаемость ферромагнетиков мало отличается от единицы. Их основные оптические свойства определяются через диэлектрическую проницаемость и электропроводность, причем основную роль играет не ускорение электронов проводимости в поле, а квантовые переходы. Намагничивание ферро- и ферримагнетика полем световой волны имеет при этом несравнимо меньшее значение, хотя оно и не исчезает: их магнитная проницаемость на оптических частотах соответствует магнитной проницаемости парамагнетиков в статических магнитных полях. В некоторых частных (редких) случаях в магнитооптические эффекты недиагональные компоненты тензоров не только диэлектрической, но и магнитной проницаемости вносят примерно одинаковый вклад.

Феноменологическое описание магнитооптических эффектов в ферромагнетиках проводят в рамках макроскопической электромагнитной теории, исходя из того, что в гиротропной среде тензоры диэлектрической и магнитной проницаемости в общем случае представляют собой несимметричные тензоры второго ранга с одной комплексной недиагональной компонентой. Для намагниченной вдоль оси z среды

$$\varepsilon = \begin{vmatrix} \underline{\varepsilon}_r & -j\underline{\varepsilon}_r Q & 0 \\ j\underline{\varepsilon}_r Q & \underline{\varepsilon}_r & 0 \\ 0 & 0 & \varepsilon_z \end{vmatrix}, \quad \mu = \begin{vmatrix} \underline{\mu}_r & -j\underline{\mu}_r Q' & 0 \\ j\underline{\mu}_r Q' & \underline{\mu}_r & 0 \\ 0 & 0 & \mu_z \end{vmatrix}, \tag{8.60}$$

где Q и Q' — комплексные магнитооптические параметры, которые в изотропной среде пропорциональны намагниченности среды ($\underline{\varepsilon}_r Q$ и $\underline{\mu}_r Q'$ — составляющие тензоров гирации).

Если $Q' = 0$ и $Q \neq 0$, то среду называют *гироэлектрической*; если $Q = 0$ и $Q' \neq 0$ — *гиромагнитной*; если $Q \neq 0$ и $Q' \neq 0$ — *бигиротропной*.

Особенности распространения электромагнитных волн в гиромагнитной среде были подробно рассмотрены в главе 6 на примере намагниченного феррита, где $\mu_r Q' = \mu_a$. Там же были изложены эффекты Фарадея и Коттона—Мутона. Магнитооптический эффект Керра определяется в основном гироэлектрическими свойствами ферромагнетиков. Поэтому дальнейшее изложение будет ограничено рассмотрением среды, свойства

которой связаны с тензорным характером диэлектрической проницаемости (8.60), т. е. в предположении, что $Q' = 0$.

Из выражения диэлектрической проницаемости (8.60) следует, что изотропная среда в магнитном поле приобретает свойства одноосного кристалла, а тензор диэлектрической проницаемости в магнитном поле обладает цилиндрической симметрией:

$$\begin{gathered}\varepsilon_{xx} \doteq \varepsilon_{yy} = \varepsilon_r; \\ \varepsilon_{xy} = -\varepsilon_{yx} = -j\varepsilon_r Q \doteq M; \\ \varepsilon_{xz} = \varepsilon_{yz} = \varepsilon_{zx} = \varepsilon_{zy} \doteq 0; \\ \varepsilon_{zz} = \varepsilon_z.\end{gathered} \tag{8.61}$$

В отсутствие магнитного поля недиагональные составляющие обращаются в нуль, а диагональные становятся одинаковыми.

Диагональные компоненты тензора диэлектрической проницаемости также являются функциями магнитооптического параметра Q. Поскольку этот параметр достаточно мал ($Q \ll 1$), то

$$\varepsilon_r(Q) \approx \varepsilon_z + \left(\frac{\partial \varepsilon_r}{\partial Q}\right)_{Q=0} Q + \frac{1}{2}\left(\frac{\partial^2 \varepsilon_r}{\partial Q^2}\right)_{Q=0} Q^2.$$

Из условия, что физические свойства тела не должны меняться при изменении направления намагниченности, имеем

$$\varepsilon_r(Q) = \varepsilon_r(-Q) \approx \varepsilon_z - \left(\frac{\partial \varepsilon_r}{\partial Q}\right)_{Q=0} Q + \frac{1}{2}\left(\frac{\partial^2 \varepsilon_r}{\partial Q^2}\right)_{Q=0} Q^2.$$

Следовательно, с достаточной точностью

$$\varepsilon_r = \varepsilon_z(1 + fQ^2), \tag{8.62}$$

где $f = (2\varepsilon_z)^{-1}(\partial^2 \varepsilon_r / \partial Q^2)_{Q=0}$.

Представив вектор рефракции (5.9) в виде

$$\mathbf{n} = n(\mathbf{e}_x \cos\alpha + \mathbf{e}_y \cos\beta + \mathbf{e}_z \cos\theta)$$

и подставив его вместе с (8.61) в уравнение Френеля (5.11б), получим

$$\begin{vmatrix} n^2 - n^2\cos\alpha - \varepsilon_r & -n^2\cos\alpha\cos\beta + j\varepsilon_r Q & -n^2\cos\alpha\cos\theta \\ -n^2\cos\alpha\cos\beta - j\varepsilon_r Q & n^2 - n^2\cos^2\beta - \varepsilon_r & -n^2\cos\beta\cos\theta \\ -n^2\cos\alpha\cos\theta & -n^2\cos\beta\cos\theta & n^2 - n^2\cos^2\theta - \varepsilon_z \end{vmatrix} = 0,$$

откуда после несложных, но громоздких преобразований

$$\begin{aligned}&(\varepsilon_r \sin^2\theta + \varepsilon_z \cos^2\theta)n^4 - [\varepsilon_r^2(1-Q^2)\sin^2\theta + \\ &+ \varepsilon_r\varepsilon_z(1+\cos^2\theta)]n^2 + \varepsilon_r\varepsilon_z(1-Q^2) = 0.\end{aligned} \tag{8.63}$$

Из данного биквадратного уравнения можно получить в общем случае две пары значений показателя преломления, которые определяют две волны, распространяющиеся в одном и том же направлении с разными скоростями, и две обратные волны с такими же скоростями. Например, при распространении света вдоль направления намагниченности ($\theta = 0$) две пары показателей преломления $n_{\pm}^2 = \varepsilon_r(1 \pm Q)$ соответствуют двум циркулярно поляризованным волнам с правым (n_+) и левым (n_-) вращением*. Эти два показателя преломления определяют угол поворота большой оси эллипса

$$\alpha_\Phi = \mathrm{Re}\left(\pi \frac{n_+ - n_-}{\lambda} h\right) = \pi \frac{n'_+ - n'_-}{\lambda} h$$

и эллиптичность

$$\left(\frac{B}{A}\right)_\Phi = \mathrm{th}\,\mathrm{Im}\left(\pi \frac{n_+ - n_-}{\lambda} h\right) = \mathrm{th}\left(\pi \frac{n''_+ - n''_-}{\lambda} h\right)$$

проходящей волны при *эффекте Фарадея* в магнитной пленке толщиной h.

В общем случае (при произвольном θ) решение уравнения (8.63) можно упростить, если подставить в него (8.62)

$$n^2 \approx \varepsilon_r(1 + fQ^2)\left[1 - \frac{1}{2}(1 + f)Q^2 \sin^2\theta \pm Q\sqrt{\cos^2\theta - f\sin^2\theta}\right]. \quad (8.64)$$

В большинстве случаев в первом приближении можно пренебречь членами второго порядка относительно Q и членом $f\sin^2\theta$. Тогда согласно формуле (8.63),

$$n^2 \approx n_0^2(1 \pm Q\cos\theta)$$

или

$$n_\pm \approx n_0\left(1 \pm \frac{1}{2}Q\cos\theta\right), \quad (8.65)$$

где $n_0 = 0{,}5(n_+ + n_-)$ — показатель преломления в отсутствие намагниченности, знаки «+» и «–» относятся к циркулярно поляризованным волнам с правым и левым вращением. В бигиротропной среде с теми же допущениями

$$n_\pm \approx n_0\left[1 \pm \frac{1}{2}(Q + Q')\cos\theta\right].$$

---

* С учетом бигиротропных свойств $n_\pm^2 = \varepsilon_r[1 \pm (Q + Q')]$. В данном параграфе в целях упрощения опущен символ комплексной величины в обозначениях показателя преломления (1.8) и коэффициента отражения.

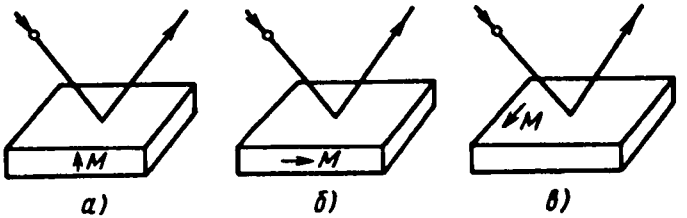


Рис. 8.26. Направление намагниченности в полярном (*а*), меридианальном (*б*) и экваториальном (*в*) эффектах Керра

В зависимости от взаимной ориентации плоскости падения и намагниченности отражающей среды различают три разновидности магнитооптического эффекта Керра. *Полярный эффект* возникает, когда вектор намагниченности перпендикулярен отражающей поверхности (рис. 8.26, *а*); при *меридианальном* (продольном) *эффекте* вектор намагниченности параллелен как отражающей поверхности, так и плоскости падения (рис. 8.26, *б*); при *экваториальном* (поперечном) *эффекте* вектор намагниченности параллелен отражающей поверхности, но перпендикулярен плоскости падения (рис. 8.26, *в*). При считывании информации с магнитных пленок обычно используют полярный эффект Керра.

Магнитооптические явления используют не только как метод исследования физических свойств ферромагнетиков, в частности, их доменной структуры, но и как один из эффективных методов считывания в устройствах хранения и обработки информации.

§ 8.6. Дифракционная эффективность магнитооптической бинарной голограммы

К одному из применений тонких магнитных пленок при хранении и обработке информации можно отнести их использование в качестве пространственных оптических модуляторов, когда запись и считывание информации можно проводить как в дискретной (поточечно), так и в голографической форме. Голографические методы позволяют осуществить запись и хранение информации с плотностью, близкой к предельной для оптического диапазона длин волн, максимально использовать разрешающую способность запоминающей среды, обеспечить высокую надежность хранения информации и снизить требования к ряду технологических параметров запоминающей среды.

Хорошие результаты были получены для двух видов тонкой магнитной пленки: марганцово-висмутовой и оксида европия. Применение пленки EuO предполагает использование криогенной техники (рабочая

температура ниже 60 К). Пленки на основе MnBi работают при комнатной температуре.

Запись информации (перемагничивания) в большинстве случаев основана на зависимости магнитных свойств от температуры, т. е. термомагнитный способ записи. Нагрев пленки осуществляют либо электронным, либо лазерным лучом. Лазер позволяет более полно использовать поверхность пленки, так как его луч можно сфокусировать до пятна меньшего размера — порядка длины волны света (1 мкм). Кроме того, для отклонения лучей записи и луча считывания можно применить одну и ту же оптическую систему, обеспечивающую попадание этих лучей в одну точку.

Наиболее распространены четыре разновидности термомагнитной записи информации: 1) в точке компенсации; 2) в точке Кюри; 3) термострикционная; 4) на пленках с полосовой доменной структурой. Все эти разновидности связаны с поглощением фотонов магнитной пленкой и последующим электрон-решеточным взаимодействием, вызывающим локальный подъем температуры. Лазеры позволяют получить подъем температуры пленки за достаточно короткое время (10...30 нс).

Энергетическая эффективность запоминающей среды тем выше, чем бóльшая часть энергии считывающего луча достигает фотоприемного устройства. *Оптический КПД* запоминающей системы

$$\eta = \frac{I_{«1»} - I_{«0»}}{I_0},$$

где $I_{«1»}$, $I_{«0»}$ — интенсивности света, поступающего на вход фотоприемника при считывании единицы и нуля соответственно; I_0 — интенсивность считывающего света, падающего на запоминающую среду.

Характеристика записи ферромагнитной пленки с прямоугольной петлей гистерезиса может быть аппроксимирована ступенчатой функцией, причем области пленки с противоположным направлением намагниченности будут отличаться знаком угла поворота плоскости поляризации. Если после пленки установлен анализатор, то разность интенсивностей световых пучков, прошедших через противоположно намагниченные участки,

$$\Delta I = I_{«1»} - I_{«0»} = I_0 \, e^{-Kh}[\cos^2(\Phi + Fh) - \cos^2(\Phi - Fh)],$$

где K — коэффициент поглощения материала; Φ — угол между осями поляризатора и анализатора; $F = \psi/h$ — удельное фарадеевское вращение плоскости поляризации.

При оптимальной толщине пленки* $h \approx K^{-1}$ и считывании с максимальным контрастом, когда $\Phi = \pi/2 - Fh$, оптический КПД

---

* По условию максимума чувствительности к управляющему полю.

$$\eta = e^{-1}\sin^2(2F/K) = e^{-1}\sin^2 Q_м,$$

где $Q_м = 2F/K$ — магнитооптическая добротность, являющаяся мерой качества магнитных материалов, используемых в устройствах оптической обработки информации.

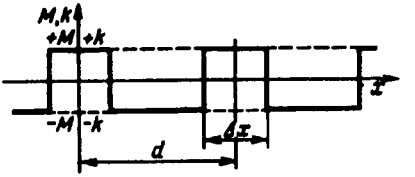


Рис. 8.27. Распределение намагниченности в бинарной голограмме

При интерференционном методе записи информации оптическую эффективность голограммы оценивают по ее *дифракционной эффективности*, т. е. по отношению интенсивности первого дифракционного максимума (несущей восстановленную информацию) к интенсивности падающего на голограмму света. Голограмму, записанную на нелинейную среду, оптические параметры которой имеют вид ступенчатой функции, называют *бинарной*. Простейшая голограмма образуется при интерференции двух плоских волн. При этом будет записана дифракционная решетка с периодом d, равным периоду интерференционной картины. Участки пленки, температура которых превысила точку Кюри, перемагнитятся. Направление их намагниченности станет противоположным ее направлению в исходном состоянии (до нагрева). Распределение намагниченности изображено на рис. 8.27.

Для считывания голограмм на магнитной пленке применяют эффекты Фарадея и Керра. Хотя считывание с использованием эффекта Керра в металлической пленке более эффективно, отношение сигнала к шуму при этом методе хуже. Кроме того, считывание на просвет приводит к упрощению конструкции оперативного голографического запоминающего устройства. Поэтому дальнейшее изложение ограничено рассмотрением восстановления бинарной решетки проходящим циркулярно поляризованным светом с той же длиной волны, как и при записи.

Распределение по поверхности пленки волнового числа соответствует распределению намагниченности (см. рис. 8.27)

$$k(x) = \begin{cases} k_+ = \omega n_+/c & \text{при } M = M_s > 0, \\ k_- = \omega n_-/c & \text{при } M = -M_s \end{cases}$$

и может быть представлено в виде гармонического ряда

$$k(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos nu_0 x + b_n \sin nu_0 x) = \frac{\Delta x}{d}(k_+ + k_-) +$$

$$+ 2\frac{k_+ - k_-}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin \pi n \frac{\Delta x}{d}\cos nu_0 x,$$

где $u_0 = 2\pi/d$.

Если не учитывать отражения от границ пленки и поглощения, то коэффициент пропускания решетки

$$T(x)=\mathrm{e}^{jkh}=\exp\left[j(\varphi_0+\varphi_1\sum_{n=1}^{\infty}\frac{1}{n}\sin\pi n\frac{\Delta x}{d}\cos nu_0x)\right]=$$

$$=\mathrm{e}^{j\varphi_0}\left[J_0(\varphi_1)+2\sum_{n=1}^{\infty}(-1)^nJ_{2n}\left(\frac{\varphi_1}{2n}\right)\sin\pi n\frac{\Delta x}{d}\cos nu_0x+\right.$$

$$\left.+2j\sum_{n=1}^{\infty}(-1)^nJ_{2n+1}\left(\frac{\varphi_1}{2n+1}\right)\sin\pi n\frac{\Delta x}{d}\cos(2n+1)u_0x\right]=$$

$$=\mathrm{e}^{j\varphi_0}\left\{J_0(\varphi_1)+\sum_{n=1}^{\infty}(-1)^nJ_{2n}\left(\frac{\varphi_1}{2n}\right)\sin\pi n\frac{\Delta x}{d}\left[\mathrm{e}^{j2nu_0x}+\mathrm{e}^{-j2nu_0x}\right]+\right.$$

$$\left.+j\sum_{n=0}^{\infty}(-1)^nJ_{n+1}\left(\frac{\varphi_1}{2n+1}\right)\sin\pi n\frac{\Delta x}{d}\left[\mathrm{e}^{j(2n+1)u_0x}+\mathrm{e}^{-j(2n+1)u_0x}\right]\right\},$$

где $\varphi_0=(k_++k_-)\frac{\Delta x}{d}h$ и $\varphi_1=2(k_+-k_-)h/\pi=4Fh/\pi$.

Следовательно, нормированная амплитуда дифрагированного решеткой поля

$$\mathcal{E}(u)=\int_{-\infty}^{\infty}T(x)\mathrm{e}^{-jux}\mathrm{d}x=\mathrm{e}^{j\varphi_0}\int_{-\infty}^{\infty}J_0(\varphi_1)\mathrm{e}^{-jux}\,\mathrm{d}x+$$

$$+\mathrm{e}^{j\varphi_0}\int_{-\infty}^{\infty}\sum_{n=1}^{\infty}(-1)^nJ_{2n}\left(\frac{\varphi_1}{2n}\right)\sin\pi n\frac{\Delta x}{d}\left[\mathrm{e}^{j2nu_0x}+\mathrm{e}^{-j2nu_0x}\right]\mathrm{e}^{-jux}\,\mathrm{d}x+$$

$$+j\mathrm{e}^{j\varphi_0}\int_{-\infty}^{\infty}\sum_{n=1}^{\infty}(-1)^nJ_{2n+1}\left(\frac{\varphi_1}{2n+1}\right)\sin\pi n\frac{\Delta x}{d}\left[\mathrm{e}^{j(2n+1)u_0x}+\mathrm{e}^{-j(2n+1)u_0x}\right]\mathrm{e}^{-jux}\,\mathrm{d}x=$$

$$=\mathrm{e}^{j\varphi_0}J_0(\varphi_1)\delta(u)-j\mathrm{e}^{j\varphi_0}J_1(\varphi_1)\sin\pi\frac{\Delta x}{d}\delta(u-u_0)-$$

$$-j\mathrm{e}^{j\varphi_0}J_1(\varphi_1)\sin\pi\frac{\Delta x}{d}\delta(u+u_0)+\mathrm{e}^{j\varphi_0}J_2\left(\frac{\varphi_1}{2}\right)\sin2\pi\frac{\Delta x}{d}\delta(u-2u_0)+$$

$$+\mathrm{e}^{j\varphi_0}J_2\left(\frac{\varphi_1}{2}\right)\sin2\pi\frac{\Delta x}{d}\delta(u+2u_0)+\ldots\doteq J_0(\varphi_1)\delta(u)-$$

$$-jJ_1(\varphi_1)\delta(u-u_0)\sin\left(\pi\frac{\Delta x}{d}\right)-jJ_1(\varphi_1)\sin\pi\frac{\Delta x}{d}\delta(u+u_0)+$$

$$+J_2\left(\frac{\varphi_1}{2}\right)\delta(u-2u_0)\sin2\pi\frac{\Delta x}{d}+J_2\left(\frac{\varphi_1}{2}\right)\sin2\pi\frac{\Delta x}{d}\delta(u+2u_0)+\ldots.$$

Первый член этого выражения соответствует центральной полосе и нулевому порядку восстановленного изображения. Поскольку его фаза с

точностью до $\varphi_0(u) = \text{const}$ совпадает с фазой падающей волны, то этот порядок поляризован так же, как и падающая волна. Второй и третий члены характеризуют ближайшие боковые дифракционные максимумы, лежащие симметрично по обе стороны от нулевого. Поляризация этих первых дифракционных максимумов повернута на 90°. Четвертый и пятый члены определяют максимум второго порядка. Поляризация этого и последующих порядков повернута на 90° по отношению к предыдущему.

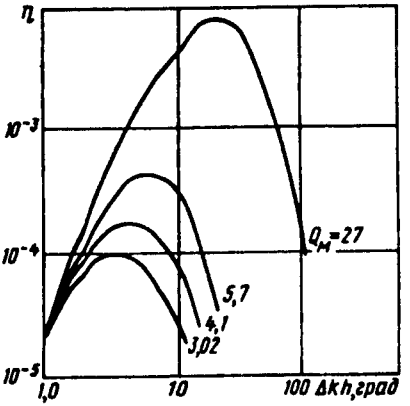


Рис. 8.28. Зависимость дифракционной эффективности от глубины фазового рельефа

Интенсивность первого порядка пропорциональна $J_1^2(\varphi_1)$, поэтому дифракционная эффективность восстановленного изображения с учетом поглощения в материале

$$\eta = e^{-Kh} J_1^2\left(\frac{4}{\pi} Fh\right) \sin^2\left(\pi \frac{\Delta x}{d}\right).$$

На рис. 8.28 показана зависимость эффективности голограммы от разности фаз $\varphi_1 = \Delta kh = 4Fh/\pi$ для магнитных пленок с различной добротностью при оптимальной скважности решетки $d/\Delta x = 2$.

Марганцово-висмутовые пленки на подложках из слюды имеют магнитооптическую добротность, обычно не превышающую 3°, и максимально достижимую эффективность 0,01% (при толщине около 55 нм). В висмутсодержащих гранатах она доходит до 5%.

Низкая дифракционная эффективность является серьезным препятствием для практического использования имеющихся магнитных материалов в реверсивных голографических запоминающих устройствах.

Магнитооптические голографические оперативные запоминающие устройства (ОГЗУ) могут быть конкурентоспособными с другими видами быстродействующих ОГЗУ только после получения менее энергоемких магнитных пленок с удовлетворительными оптическими свойствами.

Поиск и исследование новых материалов стимулируется потенциальными возможностями тонких магнитных пленок. Они не требуют энергии для хранения информации, следствием чего являются малые потребляемые мощности и сохранение информации при отключении питания. Надежность хранения обусловлена стабильностью состояний с оста-

точным намагничиванием, связанной со сравнительно высокой температурой Кюри. Физические свойства тонких магнитных пленок очень мало подвержены изменению со временем (старению) даже в сложных климатических условиях. Простота конструкции и отсутствие контактов также способствуют высокой надежности и низкой стоимости магнитопленочных элементов. Эти материалы выдерживают более 10^8 циклов перезаписи, обладают возможностью записи информации за единицы наносекунд и считывания ее без разрушения, а также высокими разрешающей способностью (более 1000 лин/мм) и магнитооптической добротностью. Последнее свойство определяет высокую эффективность оптического считывания информации.

Глава 9
Электромагнитные явления в полупроводнкоах

Полупроводниками называют материалы, у которых электропроводность при абсолютном нуле отсутствует, но заметно возрастает с повышением температуры, Чтобы сделать более определенной условную границу между изоляторами, полупроводниками и переходными металлами, принято относить к полупроводникам те вещества, у которых ширина запрещенной зоны (между зоной проводимости и валентной зоной) лежит в пределах от 0,1 до 3 эВ.

Электромагнитные свойства полупроводников обусловлены в большинстве случаев взаимодействием свободных носителей зарядов с электромагнитным полем.

§ 9.1. Дрейф свободных носителей заряда в электрическом поле

Во внешнем электрическом поле движение электрона в кристалле складывается из беспорядочного теплового и упорядоченного движения вдоль направления поля. Направленное движение совокупности свободных носителей заряда называют *дрейфом*, а скорость этого движения — *дрейфовой скоростью*.

Дрейфовую скорость электрона в элементарной электронной теории находят из уравнения

$$m_n \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = m_n \frac{\mathrm{d}^2 \upsilon_n}{\mathrm{d}t} = eE,$$

где m_n — усредненная эффективная масса электрона.

Если считать, что поле постоянно и после каждого очередного столкновения направленное движение прекращается ($\upsilon_n = 0$), то перед новым столкновением электрон приобретет дрейфовую скорость $\upsilon_n = eE\tau/m_n$ и пройдет путь $x = eE\tau^2/(2m_n)$, где τ — время свободного пробега.

Время свободного пробега в не очень сильных полях определяется в основном скоростью теплового движения и носит статистический характер. К моменту t число электронов, не испытавших соударения, $n(t) = n_o e^{-t/\tau_n}$, где $\tau_n = \nu_n^{-1}$ — среднее время свободного пробега (ν_n — частота соударений). Поэтому среднее расстояние свободного пробега n_0 электронов

$$\bar{x} = \frac{1}{n_0}(x_1 + x_2 + \ldots) = \frac{1}{n_0}\int_0^{\infty} \frac{x_n}{\tau_n}\,\mathrm{d}t = \frac{e\tau_n^2}{m_n}E$$

и средняя скорость дрейфа

$$\bar{v}_n = \frac{\bar{x}}{\tau_n} = \frac{e\tau_n}{m_n}E = \mu_n E, \tag{9.1}$$

где подвижность электронов

$$\mu_n = \frac{e\tau_n}{m_n} = \frac{e}{m_n \nu_n}. \tag{9.2}$$

Аналогично подвижность дырок:

$$\mu_p = \frac{e\tau_p}{m_p} = \frac{e}{m_p \nu_p}.$$

Дрейф свободных носителей заряда определяет плотность тока проводимости

$$J = e(n\bar{v}_n + p\bar{v}_p) = \sigma E,$$

где проводимость

$$\sigma = en\mu_n + ep\mu_p. \tag{9.3}$$

Подвижности свободных носителей заряда прямо пропорциональны их коэффициентам диффузии. В этом нетрудно убедиться, рассматривая неоднородный полупроводник в условиях термодинамического равновесия, когда внешнее электрическое поле и суммарный ток равны нулю. Здесь возникает только внутреннее поле $\mathbf{E} = -\mathrm{grad}\varphi$, определяющее ток проводимости, который должен уравновешиваться током диффузии (в равновесных условиях). Например, для электронов в одномерном случае

$$n\mu_n E = -D_n \frac{\mathrm{d}n}{\mathrm{d}x}, \tag{9.4}$$

причем в отсутствие вырождения

$$n = N_c\, \mathrm{e}^{-(U+\mathcal{E}_c-\mathcal{E}_F)/(k_Б T)} = n_0\, \mathrm{e}^{e\varphi/(k_Б T)},$$

где $U = -e\varphi$ — потенциальная энергия электрона; n_0 — равновесная концентрация электронов (при $\varphi = 0$).

Поэтому (9.4) можно переписать в виде

$$-\mu_n n_0\, \mathrm{e}^{e\varphi/(k_Б T)} \frac{\mathrm{d}\varphi}{\mathrm{d}x} = -D_n \frac{en_0}{k_Б T}\mathrm{e}^{e\varphi/(k_Б T)} \frac{\mathrm{d}\varphi}{\mathrm{d}x},$$

откуда

$$\mu_n = \frac{e}{k_Б T} D_p. \qquad (9.5а)$$

Аналогично, для дырок

$$\mu_p = \frac{e}{k_Б T} D_p. \qquad (9.5б)$$

Выражение (9.5) называют соотношением Эйнштейна.

§ 9.2. Дрейф свободных носителей заряда в магнитном поле

Магнитное поле искривляет траектории движения носителей заряда, перемещающихся со скоростью v. Если время свободного пробега электрона в твердом теле и магнитное поле достаточно велики, то в промежутке между двумя столкновениями носитель заряда будет постоянно двигаться под воздействием силы Лоренца по спиральной орбите, осью которой является направление вектора **B**. Магнитное поле не будет оказывать влияния на движение вдоль оси спирали, которое происходит со скоростью, равной продольной составляющей скорости **v**.

Угловую скорость движения ω_c в классической физике находят из условия стабильности круговой орбиты. При этом центробежная сила $m_n v^2 : r = m_n r \omega_c^2$, действующая на носитель заряда, равна центростремительной $ev_n B = e\omega_c^2 rB$, действующей со стороны магнитного поля (v_n — поперечная компонента скорости **v**, т. е. нормальная к **B**). Отсюда

$$\omega_c = \frac{e}{m_n} B. \qquad (9.6)$$

Способность носителей зарядов закручиваться в спираль вокруг оси в направлении приложенного магнитного поля с угловой скоростью, определяемой этим полем и отношением заряда к эффективной массе носителя заряда, называют *циклотронным* или *диамагнитным эффектом*. Частоту (9.6) называют *циклотронной*.

При квантовомеханическом анализе циклотронного эффекта исходят из рассмотрения распределения энергетических уровней кристалла. В отсутствие внешнего магнитного поля электроны в зоне проводимости могут находиться в любом состоянии, относящемся к квазиконтинуму разрешенных энергий, и никакие из этих электронов не обладают преимущественным поглощением. При появлении магнитного поля в зоне проводимости возникают дискретные энергетические уровни

$$\mathscr{E}_n = \frac{\hbar^2 k_z^2}{2m} + \hbar\omega_c\left(n + \frac{1}{2}\right),$$

где $\hbar k_z$ — составляющая импульса электрона вдоль магнитного поля; n — магнитное квантовое число.

Каждый из этих уровней (уровней Ландау) характеризуется большой степенью вырождения. Связанное с циклотронным эффектом селективное поглощение света определяется переходами с нижних на более высокие незанятые уровни. Переходы ограничены правилами отбора: $\Delta k_z = 0$; $n = \pm 1$. Классическая и квантовая теории приходят к сходным результатам, только при очень малых квантовых числах в отдельных случаях получают несколько различные значения резонансных частот.

§ 9.3. Дрейф свободных носителей заряда в скрещенных электрическом и магнитном полях

В однородном, бесконечно простирающемся полупроводнике во внешнем электрическом поле, направленном вдоль оси x ($E = E_x$), и слабом магнитном поле ($\mu B \ll 2\pi$), направленном вдоль оси z ($B = B_z$), на дырки действует кулоновская сила $\mathbf{F}_E^+ = e\mathbf{E}$, которая вызывает дрейф вдоль поля со скоростью $\mathbf{v}_p = \mu_p \mathbf{E}$, образуя «продольный» ток $\mathbf{J}_E^+ = ep\mu_p \mathbf{E}$. Одновременно на дырки со стороны магнитного поля действует поперечная сила $F_B^+ = e\mathbf{v}_p \times \mathbf{B}$, эквивалентная действию стороннего поля напряженностью $E_{\text{стор}}^+ = F_B^+/e = v_p B = \mu_p BE$, отклоняющая дырки от направления дрейфа в электрическом поле на небольшой угол

$$\theta_p \approx \operatorname{tg}\theta_p = F_B^+/F_E^+ = F_{\text{стор}}^+/E = \mu_p B \qquad (9.7а)$$

и определяющая «поперечный» ток $J_B^+ = ep\mu_p^2 BE \approx \theta_p J_E^+$.

На электроны также действует кулоновская сила $\mathbf{F}_E^- = -e\mathbf{E}$, определяющая дрейф со скоростью $\mathbf{v}_n = -\mu_n \mathbf{E}$ и ток $\mathbf{J}_E^- = en\mu_n \mathbf{E}$. Со стороны магнитного поля на них действует сила $\mathbf{F}_B^- = -e\mathbf{v}_n \times \mathbf{B}$, отклоняющая электроны на небольшой угол

$$\theta_n \approx \operatorname{tg}\theta_n = F_B^-/F_E^- = \mu_n B \qquad (9.7б)$$

в ту же сторону, что и дырки. Плотность соответствующего поперечного тока $J_B^- = en\mu_n^2 BE \approx \theta_n J_E^-$.

В образце полупроводника конечных размеров, например, в пластинке, параллельной плоскости xz, поперечный ток прекращается после образования на боковых изолированных поверхностях такого заряда, который полностью уравновешивает ток от стороннего поля. Возникновение в образце с током под воздействием магнитного поля поперечного электрического поля называют *эффектом Холла*. Электрическое поле зарядов на боковых поверхностях пластинки называют *полем Холла*, а углы (9.7а) и (9.7б) — *углами Холла*.

Поле Холла E_X может быть найдено из условия равенства поперечных токов, вызванных полем Холла и сторонним полем. Первый из этих токов $J_X = J_X^+ + J_X^- = e(p\mu_p + n\mu_n)E_x$, а второй $J_B = J_B^+ + J_B^- = e(p\mu_p^2 - n\mu_n^2)BE$.

В установившемся режиме $J_B = J_X$, откуда

$$E_X = E\frac{p\mu_p^2 - n\mu_n^2}{p\mu_p + n\mu_n}B. \qquad (9.8)$$

Поле Холла ослабляет эффект Холла. Поле зависит от отношения концентраций носителей заряда разных полярностей к разности их подвижностей. Например, если бы подвижности электронов и дырок в собственном полупроводнике ($n = p$) были одинаковы, то поле Холла отсутствовало бы (заряды электронов и дырок на боковых гранях полностью бы нейтрализовали друг друга). В сильно легированном полупроводнике, когда $p \gg n$ или $p \ll n$, стационарное поле Холла в первом приближении полностью компенсирует отклоняющее действие магнитного поля ($E_x = E_{\text{стор}}$) для основных носителей заряда, а для неосновных носителей заряда, наоборот, усиливает его. Поле Холла, таким образом, изменяет степень взаимодействия («связь») между дрейфом носителей заряда и магнитным полем. Поэтому отклонение траекторий носителей заряда (от направления их дрейфа в одном электрическом поле) будет характеризоваться углами, отличными от холловских.

В поле Холла для дырок угол отклонения

$$\theta^+ \approx \frac{E_{\text{стор}}^+ - E_X}{E} = \left(\mu_p - \frac{p\mu_p^2 - n\mu_n^2}{p\mu_p + n\mu_n}\right)B = \mu_p B_{p\,\text{эф}} = \alpha_p\theta_p,$$

а для электронов

$$\theta^- \approx \frac{E_{\text{стор}}^- + E_X}{E} = \left(\mu_n + \frac{p\mu_p^2 - n\mu_n^2}{p\mu_p + n\mu_n}\right)B = \mu_n B_{n\,\text{эф}} = \alpha_n\theta_n,$$

где действие поля Холла на свободный носитель заряда учитывается введением некоторого эффективного магнитного поля

$$B_{p\,эф} = n\frac{\mu_n}{\mu_p}\frac{\mu_p+\mu_n}{p\mu_p+n\mu_n}B; \quad B_{n\,эф} = p\frac{\mu_p}{\mu_n}\frac{\mu_n+\mu_p}{p\mu_p+n\mu_n}B$$

либо коэффициентом связи свободного носителя заряда с магнитным полем

$$\alpha_p = 1-\frac{p\mu_p^2-n\mu_n^2}{p\mu_p^2+n\mu_n\mu_p}, \quad \alpha_n = 1+\frac{p\mu_p^2-n\mu_n^2}{p\mu_p\mu_n+n\mu_n^2}. \tag{9.9}$$

В частности, в собственном полупроводнике $\alpha_p = \mu_n/\mu_p$, $\alpha_n = \mu_p/\mu_n$.

В полупроводнике с дырочной электропроводностью ($p \gg n$), согласно (9.9), $\alpha_p = 0$, $\alpha_n = 1+\mu_p/\mu_n$; при электронной электропроводности ($n \gg p$) — $\alpha_p = 1+\mu_n/\mu_p$, $\alpha_n = 0$. Следовательно, в сильно легированном образце ($p \gg n$ или $n \gg p$) магнитное поле не будет отклонять и искривлять прямолинейные траектории основных носителей заряда ($\theta = 0$), движущихся со средней скоростью дрейфа, а траектория неосновных носителей заряда отклонится на угол

$$\theta = \theta_n + \theta_p \approx (\mu_n + \mu_p)B. \tag{9.10}$$

Это явление лежит в основе *поперечного магнитоконцентрационного эффекта* (эффекта Суля). Он заключается в изменении распределения концентрации дрейфующих в электрическом поле неосновных носителей заряда при включении магнитного поля. Если в полупроводниковом параллелепипеде имеется ток и в отсутствии магнитного поля он распределен равномерно по поперечному сечению, то после включения магнитного поля неосновные носители будут прижиматься к одной из сторон пластинки, где скорость рекомбинации обычно выше, чем в объеме. Благодаря этому общее сопротивление образца уменьшается. Это явление называют *гальваномагниторекомбинационным эффектом.*

Продольный магнитоконцентрационный эффект наблюдается в полубесконечном образце, когда расположенный на его поверхности точечный эмиттер (или световой зонд) генерирует неравновесные носители заряда. В отсутствие магнитного поля они будут диффундировать в образце, создавая сферически симметричное распределение с центром в точке инжекции ($x = y = z = 0$). При включении магнитного поля, перпендикулярного поверхности образца ($B = B_z$), на составляющие диффузионных токов, перпендикулярных этому полю, будет действовать сила Лоренца, отклоняющая их от первоначального направления. В результате искривления траекторий в направлении, перпендикулярном B, носители заряда за время жизни продиффундируют на меньшее расстояние от точки инжекции, чем при $B = 0$, и концентрация носителей заряда вблизи оси z (вдоль светового луча) возрастет. Происходит как бы «фокусировка» диффундирующих носителей заряда вдоль направления магнитного поля. Изменение концентрации может быть обнаружено экспериментально с помощью зонда — коллектора, расположенного на поверхности образца на некотором расстоянии от эмиттера.

Эффект Суля лежит в основе работы гальваномагниторекомбинационных преобразователей. Наиболее подходящим для их изготовления является

германий. При проводимости, близкой к собственной, он сочетает достаточно высокую концентрацию электронно-дырочной плазмы с большим значением диффузионной длины в искусственно создаваемой области с высокой скоростью рекомбинации на одной из боковых сторон полупроводникового резистора. Эти условия обеспечивают наибольшую магниточувствительность — около 80 В/Тл.

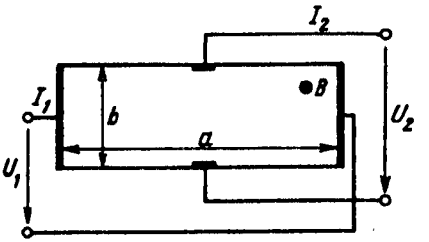


Рис. 9.1. Преобразователь Холла

Подобные гальваномагниторекомбинационные преобразователи чувствительны не только к величине магнитного поля, но и его направлению. Функциональные возможности могут быть значительно расширены созданием полевого электрода (МДП-типа) на боковой поверхности. Изменение напряжения на полевом электроде меняет скорость рекомбинации на прилегающей поверхности полупроводника, обеспечивая возможность управления магниточувствительностью.

Эффект Холла используют для измерения магнитного поля с помощью приборов, называемых *датчиками Холла*. Эти датчики изготовляют обычно либо из тонких четырехугольных пластин монокристаллов, либо из напыленных тонких пленок примесных полупроводников с высокой подвижностью свободных носителей заряда. Пластина (рис. 9.1) имеет два сплошных электрода на коротких поперечных гранях (их называют токовыми, управляющими или входными электродами) и два электрода, расположенных друг против друга посередине продольных граней пластинки (их называют выходными или холловскими электродами).

При пропускании тока I_1 через первую пару электродов в магнитном поле с индукцией B, перпендикулярной плоскости датчика, между (разомкнутыми) выходными электродами возникает напряжение

$$U_2 = \frac{\mu}{\sigma d} B I_1 = \gamma_{20} B I_1, \qquad (9.11)$$

пропорциональное I_1 и B (d — толщина пластины). Показателем качества датчика является удельная чувствительность $\gamma_{20} = \mu/(\sigma d)$.

Если под воздействием внешнего напряжения U_2 возникает ток I_2, а управляющие электроды разомкнуты, то напряжение на этих электродах $U_1 = -\gamma_{20} I_2 B$ будет иметь направление, обратное обозначенному на рис. 9.1.

Для датчика Холла, как и для всякого четырехполюсника

$$U_1 = R_{11} I_1 + R_{12} I_2, \quad U_2 = R_{21} I_1 + R_{22} I_2,$$

причем

$$R_{21} = (U_2/I_1)_{I_2=0} = \gamma_{20} B = -(U_1/I_2)_{I_1=0} = -R_{12}.$$

Следовательно, датчик представляет собой невзаимный четырехполюсник, обладающий свойствами *гиратора.*

Датчики Холла используются для измерения магнитных полей напряженностью 10...10^4 А/см. Из других многочисленных применений можно отметить бесконтактное измерение токов по имеющемуся вокруг них магнитному полю.

У преобразователя ХАГ-П из арсенида галлия удельная чувствительность достигает 45 кВ/(А·Тл).

Технология изготовления датчиков Холла близка к технологии изготовления обычных интегральных схем. Это позволяет создать микроминиатюрные датчики весьма малых размеров как самостоятельные элементы, а также как непосредственно сопряженные с электронным устройством и образующие единую функциональную схему на одной общей подложке. Интегральные схемы, выполненные с датчиком Холла на одном кристалле, могут непосредственно сопрягаться со стандартными элементами других интегральных схем. Одним из примеров использования интегральных схем с датчиками Холла, реализующими с малой погрешностью линейную выходную характеристику, является множительно-делительное устройство.

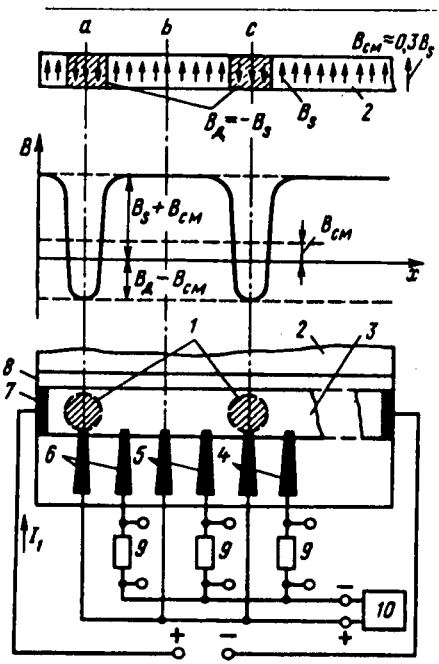


Рис. 9.2. Гальвано-магнитное устройство для индикации ЦМД

Датчики Холла применяются также для считывания информации в магнитных устройствах с ЦМД. Можно применить обычные четырехэлектродные микродатчики, активная часть которых имеет площадь, приблизительно равную площади индуцируемого домена. Однако более простую конструкцию и лучшую технологичность имеет дифференциальный вариант считывания информации, при котором система датчиков Холла реагирует не на нормальную составляющую магнитной индукции ЦМД, а на перепад индукции у его стенок (рис. 9.2). Здесь датчики, служащие для детектирования доменов *1* в точках *a*, *b* и *c* монокристаллического ферромагнетика *2*, образованы в полосе *3* из полупроводниково-

го материала с парами холловских электродов *4*, *5* и *6*, расположенных по одну сторону полосы. Ширина полосы *3* несколько больше диаметра доменов, а расстояние между электродами каждой пары *4*, *5*, *6* таково, что если один из них расположен на продолжении диаметра домена, то другой находится за краем домена. С помощью электродов *7* подводят управляющий ток I_1. Датчики располагают на изолирующей подложке *8*, плотно прижатой к ферромагнитному кристаллу *2*. Каждую пару холловских электродов через резистор *9* подключают к источнику *10*, который должен скомпенсировать напряжение между электродами от управляющего тока в отсутствие магнитного поля. На резисторе *9* разность потенциалов возникает только при наличии домена в соответствующих точках (*a*, *b*, *c*)

Явления переноса, возникающие при одновременном воздействии на проводник электрического и магнитных полей, называют *гальваномагнитными*.

Эффект Холла представляет собой одно из гальваномагнитных явлений. К ним относится и магниторезистивный эффект, обусловленный, как и эффект Холла, действием магнитного поля на равновесные носители заряда. Магниторезистивный эффект возникает в результате изменения подвижности свободных носителей заряда при появлении магнитного поля.

§ 9.4. Подвижность свободных носителей заряда в поперечном магнитном поле

Электропроводность полупроводника в магнитном поле уменьшается за счет уменьшения эффективной длины пробега носителей заряда, т. е. сокращения расстояния, проходимого носителем заряда между соударениями (актами рассеяния) вдоль направления внешнего электрического поля. Если поле Холла отсутствует, то под воздействием магнитного поля первоначальная (при $B = 0$) средняя длина свободного пробега $\bar{x}$ в направлении электрического поля уменьшается (рис. 9.3) на $\Delta\bar{x} = DK = \bar{x} - AD$, причем $\bar{x}_B = AD = CB = r\sin\theta = (\bar{x}/\theta)\sin\theta \approx \approx (\bar{x}/\theta)(\theta - \theta^3/3) = \bar{x}(1 - \theta^2/3)$, где угол $\theta = \omega_c\tau = e\tau B/m = \mu B$ и μ — подвижность при $B = 0$. В магнитном поле скорость дрейфа $\bar{v}_B = \bar{x}_B/\tau$ в $\bar{v}_B/\bar{v} = \bar{x}_B/\bar{x} \approx 1 - \theta^2/3$ раз меньше, чем при $B = 0$. Следовательно, подвиж-

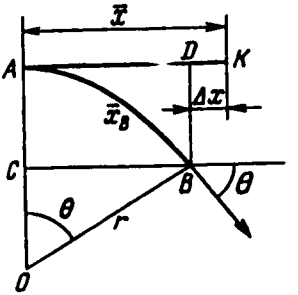


Рис. 9.3. Сокращение длины свободного пробега в магнитном поле

ность при этом уменьшается в $\mu_B/\mu \approx 1-\theta^2/3$ раз и ее относительное изменение

$$\frac{\Delta\mu}{\mu} = \frac{\mu_B - \mu}{\mu} \approx -\frac{\theta^2}{3} = -\frac{\mu^2}{3}B^2 = -\beta_0 B^2,$$

где $\beta_0 = \mu^2/3$.

После появления стационарного поля Холла искривление траектории электрона определяется эффективным магнитным полем $B_{эф} = \alpha B$ и относительное изменение подвижности

$$\frac{\Delta\mu}{\mu} = -\frac{\mu^2 B_{эф}^2}{3} = -\frac{\alpha^2\mu^2}{3}B^2 = -\beta B^2, \qquad (9.12)$$

где $\beta = \alpha^2\beta_0 = \alpha^2\mu^2/3$. Следовательно, в магнитном поле уменьшается и проводимость. Например, в материале с электронной электропроводностью $\sigma_B = en\mu_B$ относительное изменение проводимости

$$\frac{\Delta\sigma}{\sigma} = \frac{\sigma_B - \sigma}{\sigma} = \frac{\mu_B - \mu}{\mu} = \frac{\Delta\mu}{\mu} = -\beta B^2. \qquad (9.13)$$

Коэффициенты α и β не равны нулю даже в том случае, когда поле Холла уравновешивает силу Лоренца у носителей заряда, движущихся со средней скоростью дрейфа. Это связано с тем, что тепловое движение электронов вызывает значительные отклонения скоростей от среднего значения. Поэтому электроны и дырки, имеющие скорости, отличные от средней, испытывают отклонение и искривление своих траекторий, определяя изменение сопротивления полупроводникового материала.

Изменение электрического сопротивления под воздействием магнитного поля называют *магниторезистивным эффектом* или *эффектом Гаусса*. Как следует из приведенного качественного рассмотрения в сравнительно слабых магнитных полях, изменение сопротивления пропорционально квадрату магнитной индукции и квадрату подвижности носителей заряда. В связи с этим в магниторезисторных приборах используются материалы с большей подвижностью носителей заряда (InSb, InAs, GaAs, Ge и др.).

Магниторезистивный эффект зависит от угла φ между направлениями магнитного поля и тока. Максимум магнитного сопротивления наблюдается, когда поле перпендикулярно направлению тока, а минимум — при совпадении их направлений. Для полупроводников с простой зонной структурой, согласно более строгому расчету, относительное магнитосопротивление

$$\frac{\Delta\rho}{\rho}=\frac{3}{2}A\frac{e}{\sigma}\left[n\mu_n^3+p\mu_p^3-\frac{\pi}{4}\frac{(n\mu_n^2-p\mu_p^2)^2}{n\mu_n+p\mu_p}\right]B^2\sin^2\varphi,$$

где коэффициент A зависит от механизма рассеяния носителей заряда (при рассеянии на акустических колебаниях решетки $A = 3/8$, при рассеянии на ионах примеси $A = 1{,}93$; если время релаксации не зависит от энергии носителя заряда, то $A = 1$).

Зависимость магниторезистивного эффекта от направления магнитного поля и тока свидетельствует о том, что в общем случае магнитосопротивление имеет тензорный характер. В простейшем случае ($A = 1$) движение свободных электронов в первом приближении можно описать с помощью уравнения Друде-Лоренца

$$m_n\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}+m_n\nu_n\mathbf{v}=e\mathbf{E}+e\mathbf{v}\times\mathbf{B}.$$

Если дрейфовую скорость электрона $\mathbf{v}$ выразить через плотность тока $\mathbf{J}=en\mathbf{v}$, считая, что она меняется по гармоническому закону, и перейти к ее комплексному значению $\dot{\mathbf{J}}(t)=\dot{\mathbf{J}}e^{j\omega t}$, то уравнение приобретает более простой вид

$$j\omega\dot{\mathbf{J}}+\nu_n\dot{\mathbf{J}}=\frac{e^2n}{m_n}\dot{\mathbf{E}}+\frac{e}{m_n}\dot{\mathbf{J}}\times\mathbf{B}.$$

Для случая, когда вектор магнитной индукции направлен вдоль оси z и $\mathbf{B}(t)=B\mathbf{e}_z=\mathrm{const}$, получим три уравнения для компонент:

$$(\nu_n+j\omega)\dot{J}_x=\varepsilon_0\omega_p^2\dot{E}_x+\omega_c\dot{J}_y;$$

$$(\nu_n+j\omega)\dot{J}_y=\varepsilon_0\omega_p^2\dot{E}_y-\omega_c\dot{J}_x;$$

$$(\nu_n+j\omega)\dot{J}_z=\varepsilon_0\omega_p^2\dot{E}_z,$$

где $\omega_p=\sqrt{e^2n/\varepsilon_0m_n}$ — плазменная частота. Следовательно, тензор сопротивления

$$\rho=\begin{vmatrix}\rho_0 & -\rho_м & 0\\ \rho_м & \rho_0 & 0\\ 0 & 0 & \rho_0\end{vmatrix},$$

диагонален, причем от магнитного поля зависят только недиагональные компоненты $\pm\rho_м=B/(en)=\rho_n\theta_n$ (*магнитосопротивления*); диагональные компоненты $\rho_0=(\nu_n+j\omega)/(\varepsilon_0\omega_p^2)$ можно назвать *динамическими сопротивлениями* при $B = 0$. При постоянном токе выражения для динамиче-

ских сопротивлений соответствуют обычному выражению для проводимости (9.3), поскольку при $\omega = 0$

$$\rho_0 = (en\mu_n)^{-1} = \sigma_n^{-1} = \rho_n,$$

если учесть выражение для подвижности (9.2).

Более сложная магнитоэлектрическая связь имеется в тех соединениях переходных и редкоземельных элементов, которые принято называть *магнитными полупроводниками*. Характерная особенность ферромагнитных полупроводников — их огромное магнитное сопротивление, на много порядков превосходящее магнитное сопротивление немагнитных полупроводников.

Кристаллы *n*-типа обнаруживают обычно очень большое изотропное отрицательное магнитное сопротивление, зависящее от степени легирования. Отрицательное магнитное сопротивление вызывается в этом случае как увеличением концентрации носителей заряда из-за приближения дна зоны проводимости к донорному уровню, так и ростом подвижности носителей заряда из-за ослабления их рассеяния флуктуациями намагниченности (подавление полем флуктуации спинов не зависит от направления магнитного поля).

В кристаллах *p*-типа магнитное сопротивление значительно меньше и оно анизотропно. Взаимодействие дырок со спинами магнитных атомов значительно слабее, чем электронов. При температурах много бóльших точки Кюри (130 К) электрические свойства ферромагнитных полупроводников (например, Eu, EuS, EuSe, EuTe) сходны со свойствами немагнитных проводников.

Резистор, основное свойство которого заключается в способности изменять свое сопротивление при изменении воздействующего на него внешнего магнитного поля, называют *магниторезистором*. Квадратичная зависимость от магнитной индукции ограничивает применение этих приборов в слабых полях. Однако в более сильных магнитных полях получаемые сигналы значительно больше, чем сигналы, поступающие от датчиков Холла. Магниторезисторы применяют в датчиках перемещений, бесконтактных электромагнитных реле и клавишах.

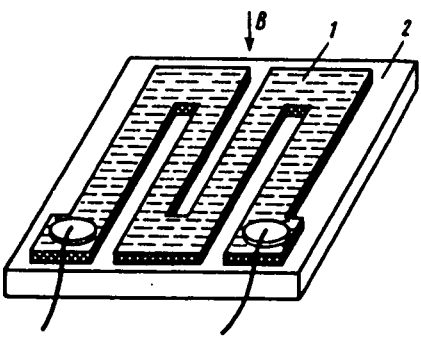


Рис. 9.4. Магниторезистор из эвтектического сплава

Конструкция магниторезистора из эвтектического сплава InSb—NiSb приведена на рис. 9.4. Толщина меандровой полоски *1* из эвтектического сплава составляет приблизительно 25 мкм, а ширина

не превышает 80 мкм. Меандр наклеивается на изолирующую подложку *2*. Для повышения чувствительности магниторезистивного преобразователя используют ферромагнитный материал подложки с высокой магнитной проницаемостью (обычно пермаллоевые сплавы). Тонкопленочная конструкция покрывается защитным изолирующим лаком.

В запоминающих устройствах с подвижным магнитным носителем применение магниторезистивного преобразователя значительно упрощает процесс воспроизведения информации. Становится возможным изготавливать тонкопленочный преобразователь методами интегральной технологии. При использовании экранированных магниторезистивных преобразователей их разрешающая способность заметно увеличивается. Практическая реализация магниторезистивного метода считывания проще оптического, обладающего такой же разрешающей способностью. Преимуществом данного метода является также независимость сигнала от скорости движения магнитного носителя.

Для считывания информации обычно используются тонкопленочные магниторезисторы из поликристаллического ферромагнетика, сопротивление которого анизотропно:

$$R = R_0 + \Delta R \cos^2 \varphi,$$

где ΔR — анизотропия магнитосопротивления; φ — угол между векторами плотности тока и намагниченности. Для пермаллоевых сплавов или сплавов железо-кобальт при комнатной температуре $\Delta R/R_0 = 0{,}02...0{,}06$. Электрическое сопротивление преобразователя с током изменяется в результате отклонения вектора **M** полем рассеяния магнитного носителя на угол φ.

Магниторезистивный метод оказался наиболее удачным и нашел практическое применение для воспроизведения информации в устройствах ЦМД. Магниторезистивный преобразователь реагирует на магнитное поле рассеяния проходящего ЦМД. Обычно он напыляется в виде прямоугольных пермаллоевых аппликаций *1* (рис. 9.5, *а*). Чувствительный элемент располагается около конца доменопродвигающей аппликации *2*, где он под воздействием управляющего поля H_D оказывается в состоянии, близком к насыщению, в направлении, перпендикулярном току, если ЦМД отсутствует. В этом состоянии преобразователь имеет малое сопротивление ($\varphi = 90°$). При наличии ЦМД *3* (рис. 9.5, *б*) магниторезистор оказывается в области, где управляющее

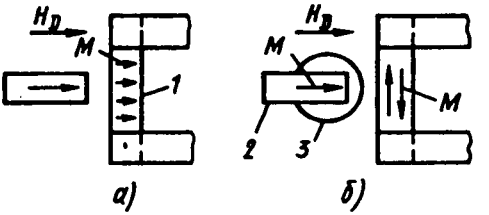


Рис. 9.5. Магниторезистивный детектор ЦМД

поле и поле рассеяния ЦМД компенсируют друг друга. При этом в пермаллоевой пленке возникают два домена с антипараллельным намагничиванием вдоль оси легкого размагничивания, обусловленной анизотропией формы и параллельной току, т. е. вдоль длинной стороны аппликации. Сопротивление преобразователя в этом случае оказывается больше ($\varphi = 0$), чем при отсутствии домена.

Обычно используется дифференциальное включение двух магниторезисторов. Один из них, выполняющий функции опорного, располагается в нерабочем канале и выдает магниторезистивный сигнал от потока «нулей». Другой магниторезистор выдает магниторезистивный сигнал от «нулей» и «единиц», индуцируемых ЦМД и продвигаемых по рабочему каналу. Выходной усилитель, воспринимая лишь разность этих сигналов, формирует двоичный сигнал, проходящий по рабочему каналу.

§ 9.5. Магнитодиодный эффект

В отличие от эффекта Холла и магниторезистивного эффекта магнитодиодный эффект относится к гальваномагнитным явлениям на неравновесных носителях заряда. Он заключается в относительно резком возрастании сопротивления полупроводникового диода в прямом направлении под воздействием поперечного магнитного поля. Наиболее отчетливо этот эффект проявляется при несимметричном *p-n*-переходе и удлиненной базе, т. е. когда концентрация равновесных носителей заряда, например в *p*-области, много больше, чем в *n*-области, и расстояние между *p-n*-переходом и вторым контактом значительно превышает диффузионную длину. Прямой ток сильно зависит от отношения толщины базовой области к длине диффузионного смещения и практически все напряжение приходится на базу; напряжение на *p-n*-переходе составляет малую долю напряжения на диоде.

Возрастание сопротивления такого диода после появления поперечного магнитного поля обусловлено совокупностью трех явлений. Во-первых, уменьшается подвижность носителей заряда. Влияние магнитного поля особенно значительно на неравновесные неосновные носители заряда, определяющие сопротивление базы, поскольку поле Холла, созданное основными носителями заряда, усиливает искривление траекторий неосновных носителей заряда. Во-вторых, уменьшается диффузионная длина $L \approx \sqrt{D_p \tau_p} = \sqrt{k_Б T \mu_p \tau_p / e}$ вследствие уменьшения подвижности μ_p (τ_p — время жизни носителя заряда). В слабых магнитных полях

$$\frac{\Delta L}{L} = \frac{L_B - L}{L} \approx \frac{1}{2}\frac{\Delta\mu}{\mu} \approx -\frac{\beta}{2}B^2.$$

Уменьшение глубины проникновения неравновесных носителей заряда существенно снижает их концентрацию на расстояниях, больших L_p, и приводит, таким образом, к дальнейшему росту сопротивления n-области. В-третьих, уменьшается инжекция дырок из p-n-перехода вследствие увеличения сопротивления базы и уменьшения напряжения на p-n-переходе.

Полупроводниковые приборы, в которых используется магнитодиодный эффект, предназначенные для применения в качестве датчиков магнитного поля, называют *магнитодиодами*. Они обладают чувствительностью к магнитному полю, в сотни и тысячи раз большей чувствительности датчиков Холла и магниторезисторов. Это существенно упрощает схемное решение многих задач, так как исключает необходимость в промежуточных усилителях: сигнал можно подавать непосредственно на регистрирующий или исполнительный орган того или иного устройства.

Магнитодиоды получили распространение в качестве преобразователей малых механических перемещений в электрические сигналы достаточно большой амплитуды — бесконтактных механических переключателей (коммутаторов) электрических сигналов. Одно из применений коммутаторов — бесколлекторные двигатели постоянного тока. Кроме того, их используют для измерения магнитных полей, магнитодефектоскопии изделий из ферромагнитных материалов (стальные трубы, рельсы, канаты, прокат и т. д.), снятия информации в магнитных устройствах памяти, измерения СВЧ-мощности и ее пространственного распределения, ввода информации в ЭВМ.

Магнитной чувствительностью магнитодиодов называют отношение $(U_B - U_0)/B$, где U_B — напряжение при номинальном (прямом) токе и B = 0,3 Тл; U_0 — то же напряжение, но без поля ($B = 0$). В кремниевых сплавных бескорпусных магнитодиодах КД301Ж при токе 1 мА магнитная чувствительность 20 В/Тл, а при токе 3 мА — 60 В/Тл. Основные недостатки современных магнитодиодов — сильная температурная зависимость параметров и разброс характеристик отдельных образцов.

§ 9.6. Резонансные свойства полупроводникового резистора

Модуляцию сопротивления полупроводниковой пластины за счет распространения в ней неосновных неравновесных носителей заряда под действием электрического поля можно применить в кристалле, выполняющем функции фильтра.

Структура простейшего полупроводникового фильтра представлена на рис. 9.6, *а*. Переменное входное напряжение, приложенное к переходу *1*, вызывает инжекцию неосновных носителей заряда в базу *2*. Неосновные носители заряда движутся в базе под действием электрического поля,

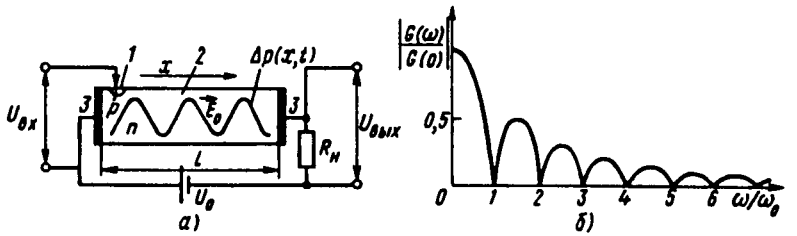


Рис. 9.6. Полупроводниковый фильтр:

а — структура и схема включения; *б* — амплитудно-частотная характеристика

создаваемого внешним источником напряжения U_0. Переменный выходной сигнал снимается с нагрузки $R_н$.

Если к инжектирующему *p-n*-переходу, смещенному в прямом направлении, приложить малое переменное напряжение сигнала $u(t)$, то переменная составляющая плотности неосновных носителей заряда в базе около *p-n*-перехода $\Delta p(t) = \gamma u(t)$, где γ — постоянный коэффициент, определяемый типом *p-n*-перехода и постоянным напряжением смещения.

Под действием электрического поля $E_0 = U_0/l$ изменение плотности неосновных носителей заряда во времени преобразуется в пропорциональное изменение вдоль оси x . Координата x отсчитывается от *p-n*-перехода вдоль направления поля E_0. Если в однородной базе не учитывать диффузию и рекомбинацию, то плотность неравновесных носителей заряда в точке x

$$\Delta p(x, t) = \gamma u[t - x/(\mu_p E_0)],$$

где μ_p — подвижность инжектированных носителей заряда.

При синусоидальном входном сигнале $u = U_m \sin \omega t$ плотность неравновесных носителей $\Delta p(x, t) = \gamma U_m \sin[\omega t - \omega x/(\mu_p E_0)]$; в комплексной форме $\Delta \dot{p} = \gamma U_m e^{-j\omega x/(\mu_p E_0)}$.

Проводимость базы $\sigma = en_0\mu_n + ep_0\mu_p + e\Delta p\mu_p = \sigma_0 + \Delta\sigma$, где n_0 и p_0 — плотности электронов и дырок при тепловом равновесии. Плотность неравновесных носителей заряда для малых сигналов много меньше равновесной плотности основных носителей. Поэтому

$$\Delta\sigma = e\Delta p\mu_p \ll e(n_0\mu_n + p_0\mu_0) \approx \sigma_0.$$

Если база имеет форму прямоугольного параллелепипеда длиной l и сечением S, то сопротивление между омическими контактами

$$Z(u)=\int_0^l \frac{dx}{\sigma S}=\frac{1}{S}\int_0^l \frac{dx}{\sigma_0+\Delta\sigma}=\frac{1}{S}\int_0^l \frac{\sigma_0-\Delta\sigma}{\sigma_0^2-\Delta\sigma^2}\,dx=R_0-\Delta Z(u)$$

и вызванное входным сигналом изменение сопротивления

$$\Delta Z(u)=R_0-Z(u)\approx\frac{1}{\sigma_0^2 S}\int_0^l \Delta\sigma\,dx=\frac{e\mu_p}{\sigma_0^2 S}\int_0^l \Delta p\,dx,$$

или в комплексной форме

$$\Delta\dot z=\frac{e\mu_p}{\sigma_0^2 S}\int_0^l \Delta\dot p\,dx=\frac{\gamma_e\mu_p U_m}{\sigma_0^2 S}\int_0^l e^{-j\omega x/(\mu_p E_0)}dx=\frac{\gamma_e\mu_p^2 E_0 U_m}{j\omega\sigma_0^2 S}(1-e^{-j\omega l/(\mu_p E_0)})=$$

$$=C\frac{1-e^{-j2\theta}}{2j\theta}=Ce^{-j\theta}\frac{e^{j\theta}-e^{-j\theta}}{2j\theta}=Ce^{-j\theta}\frac{\sin\theta}{\theta}=\Delta z\,e^{-j\theta},$$

где $\Delta z=C\sin\theta/\theta$, $2\theta=\omega l/(\mu_p E_0)=2\pi\omega/\omega_0$; резонансная частота $\omega_0=2\pi\mu_p E_0/l=2\pi/\Delta T$ определяется пролетным временем $\Delta T=l/(\mu_p E_0)$. Ток в нагрузке $R_н$

$$\dot I=\frac{U_0}{\dot z+R_н}=\frac{U_0}{R(0)-\Delta\dot z},$$

где $R(0)=R_н+R_0$. Выходное напряжение

$$\dot U_{вых}=R_н\dot I=\frac{R_н U_0}{R(0)-\Delta\dot z}=U_{вых}(0)+\Delta\dot U,$$

где $\Delta\dot U\approx R_н\Delta\dot z U_0/R^2(0)$ — переменная составляющая выходного напряжения. Его амплитудное значение $\Delta u_{max}=R_н\Delta z U_0/R^2(0)$.

Следовательно, коэффициент передачи

$$G(\omega)=\frac{\Delta u_{max}}{u_{max}}=\frac{CR_н}{R^2(0)}\frac{U_0}{U_m}\frac{\sin\theta}{\theta}=G(0)\frac{\sin\pi\omega/\omega_0}{\pi\omega/\omega_0},\qquad(9.14)$$

где $G(0)=\lim\limits_{\omega\to0}G(\omega)=CR_н U_0:[R^2(0)U_m]$ — максимальное значение коэффициента.

Нормированная амплитудно-частотная характеристика фильтра (9.14) показана на рис. 9.6, *б*. Структура, представленная на рис. 9.6, *а*, при несколько меньших напряженностях электрического поля может применяться и в качестве линии задержки. При этом задержку можно регулировать напряжением U_0.

Рассмотренный фильтр относится к такому виду полупроводниковых приборов, которые называют интегральными схемами или устройствами *молекулярной электроники*.

Интегральные схемы представляют собой микроминиатюрные электронные устройства, в которых используются явления и эффекты, возникающие в твердом теле. В настоящее время наибольшее развитие получили схемы, создаваемые в полупроводниковых монокристаллах. Интегральные схемы делятся на два больших класса: с элементной и с функциональной интеграцией.

Схемы с элементной интеграцией представляют собой полупроводниковые монолитные схемы, в которых дискретным элементам с определенными функциями (например, транзисторам, резисторам или конденсаторам) соответствуют локальные области с четкими границами. В одной полупроводниковой пластине, образующей интегральную микросхему, сосредоточены все входящие в нее элементы.

В схемах с функциональной интеграцией (функциональных схемах) невозможно отождествить отдельные структурные области с элементами обычных схем автоматики или радиотехники. Такие схемы можно характеризовать лишь в целом по выполняемой ими функции. Только в этом смысле их можно сопоставлять со схемами, состоящими из радиодеталей (генераторами, усилителями, выпрямителями, преобразователями частоты, модуляторами, фильтрами, дискриминаторами, сумматорами, коммутаторами, счетчиками и т. д.).

Потенциальные возможности функциональных схем значительно шире, поскольку в них используется большее число физических явлений, чем это возможно в схемах с элементной интеграцией. Сюда относятся накопление и взаимодействие объемных зарядов, термоэлектрические эффекты, эффекты Холла, Ганна и т. д.

К схемам с функциональной интеграцией можно отнести и рассмотренный полупроводниковый фильтр. Следующие параграфы посвящены изучению функциональных схем, в которых используется эффект Ганна.

§ 9.7. Электрические домены в полупроводниках

Структура энергетических уровней некоторых полупроводников обусловливает увеличение объемного сопротивления с ростом напряженности электрического поля в кристалле, т. е. наличие на вольт-амперной характеристике участка с отрицательным дифференциальным сопротивлением. Когда характеристика к тому же имеет N-образную форму (одному значению тока соответствует несколько значений напряжения), в системе возникает тенденция к расслоению электрически однородного при малых полях образца на области с разным удельным сопротив-

лением, что приводит к образованию областей с разной напряженностью электрического поля. Образующаяся область сильного поля, содержащая два слоя объемного заряда, получила название *домена*. Возникновение, перемещение и исчезновение домена приводит к появлению электрических колебаний. Явления, связанные с движением доменов в сильном электрическом поле в образцах арсенида галлия и фосфида индия, экспериментально обнаружены Д. Ганном в 1963 г. и были названы *эффектом Ганна*.

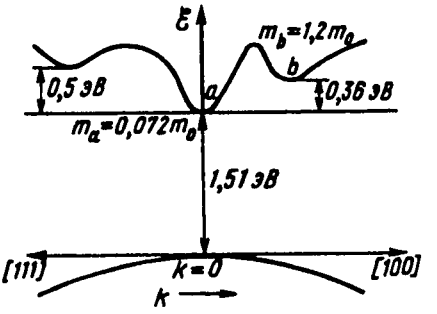


Рис. 9.7. Структура энергетических зон арсенида галлия

Отрицательный наклон вольт-амперной характеристики может возникать в результате изменения средней эффективной массы носителей заряда при электрическом поле, достаточном для их переноса из одной подзоны проводимости в другую подзону с меньшей проводимостью.

Естественной зоной структуры, необходимой для возникновения отрицательного сопротивления, обладают соединения типа $A^{III}B^{V}$. Зона проводимости этих материалов имеет самый низкий минимум при $k = 0$ и отстающие от него на несколько десятых электрон-вольта высокие минимумы в направлениях [100] и [111]. Эффективная масса электрона в этих верхних минимумах значительно превышает эффективную массу в нижнем минимуме (в соответствии с кривизной границ энергетических уровней). В качестве примера на рис. 9.7 показана структура зоны проводимости арсенида галлия.

Средняя эффективная масса и проводимость кристалла зависят от относительной заселенности каждой подзоны, которые будут функциями энергии носителей заряда. С увеличением приложенного напряжения последняя может изменяться. Следовательно, при определенных условиях можно использовать этот эффект для получения отрицательного сопротивления.

Изменение сопротивления полупроводникового кристалла обычно определяется переходами между двумя нижними подзонами a и b с энергиями минимумов $\mathcal{E}_a$ и $\mathcal{E}_b$. Если направление **E** совпадает с осью [100], то проводимость $\sigma = e(\mu_a n_a + \mu_b n_b)$, где n_a, n_b — концентрации электронов в подзонах; μ_a, μ_b — их подвижности.

Когда электроны ускоряются достаточно большим электрическим полем E, их эффективная температура поднимается. В результате изменяется не только средняя подвижность, но и концентрация электронов в подзонах.

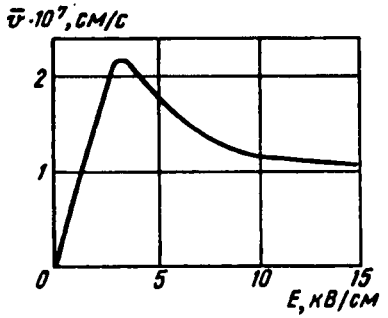


Рис. 9.8. Зависимость средней скорости дрейфа свободных электронов в арсениде галлия от напряженности однородного поля

Зависимость средней скорости дрейфа свободных электронов $\bar{v} = \dfrac{n_a\mu_a + n_b\mu_b}{n_a + n_b}E = \bar{\mu}E$, пропорциональной в однородном полупроводнике с постоянным равномерным полем плотности тока $J = en\bar{v}$, приведена на рис 9.8.

Кристалл с отрицательным сопротивлением является электрически неустойчивым. Флуктуация концентрации носителей заряда в любой точке кристалла приводит к образованию пространственного заряда, который уменьшает электрическое поле с одной стороны и увеличивает — с другой. В этом случае динамика процесса определяется дифференциальной проводимостью $\sigma_d < 0$ и временем релаксации $\tau_d = \varepsilon/\sigma_d$, которое также отрицательно.

Если бы сопротивление было положительным, то ток, текущий в область пространственного заряда электронов был бы меньше, чем из этой области, и, следовательно, пространственный заряд постепенно рассасывался бы и снова устанавливалось однородное распределение поля. Длительность этого процесса определяется обычным временем релаксации $\tau = \varepsilon/\sigma > 0$.

Когда же сопротивление отрицательно, происходит обратное явление и пространственный заряд увеличивается до тех пор, пока в обогащенной области он не будет достаточно большим, а в обедненной области — достаточно малым, чтобы дифференциальное сопротивление вновь стало положительным, а токи с обеих сторон — равными. В результате устанавливается стационарное состояние, но с сильно неоднородным распределением поля, характерным для доменной структуры. Очевидно, стационарное состояние возможно только в том случае, когда скорости электронов вне домена и в нем станут одинаковыми.

Для образования домена необходимо создать в образце с электродами определенную напряженность электрического поля, которая не может быть меньше $E_т$, соответствующей максимуму скорости электронов $v_т$. При напряженностях поля, превышающих $E_т$, средняя скорость электронов ниже, чем при напряженностях, не намного меньших $E_т$. Поэтому если напряженность локального поля проходит через значение $E_т$ при $\mathrm{d}|E|/\mathrm{d}x > 0$, то будет образовываться обогащенный электронами слой *1*,

а при $d|E|/dx < 0$, — обедненный слой *2* (рис.9.9). Обогащенный и обедненный слои, расположенные на малом расстоянии друг от друга, образуют дипольный слой, внутри которого поле объемного заряда складывается с внешним полем, что вызывает нарастание домена и приводит к неустойчивому состоянию кристалла.

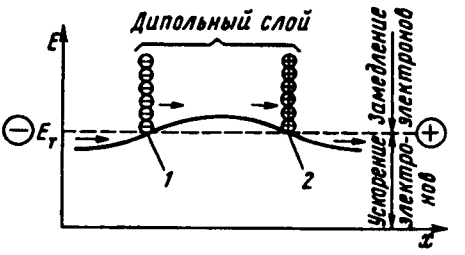


Рис. 9.9. Структура объемного заряда домена

Перераспределение напряжения внутри кристалла после появления первого домена препятствует возникновению второго домена. Это можно видеть из схематических кривых распределения потенциала в случае, когда источник энергии работает в режиме генератора напряжения (рис. 9.10). В первый момент после включения устанавливается линейное распределение потенциала (прямая *1*). Затем неустойчивость приводит к появлению узкой области очень сильного электрического поля, которая обычно возникает сначала у катода (кривая *2*). С течением времени стационарный домен перемещается вдоль образца (кривые *3*, *4* и *5*), тогда как потенциалы катода ($x = 0$) и анода ($x = l$) остаются приблизительно постоянными. Появление и распространение области сильного поля приводит к уменьшению градиента потенциала в остальных частях образца. В частности, для кривых *2*—*5* напряженность электрического поля вне домена меньше, чем для прямой *1*, и меньше значения, необходимого для образования нового домена. Это подавление прекращается после того, как первый домен достигает анода и распределение потенциала вновь соответствует прямой *1*.

Если процесс формирования поля в двойном слое быстро заканчивается, то слой начинает распространяться с постоянной скоростью v_D, формой и амплитудой. Его можно рассматривать как ударную волну, которая быстро достигает установившейся формы, характерной для стабильного домена. В этом случае распределение потенциала ударной волны можно записать в виде $\varphi = \varphi(x - v_D t) = \varphi(w)$, где

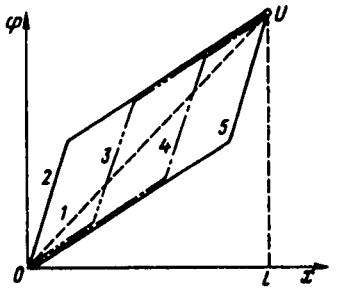


Рис. 9.10. Распределение потенциала в полупроводнике с доменом

$$w = x - v_D t. \qquad (9.15)$$

Анализ движения стабильных доменов можно провести, если известна ста-

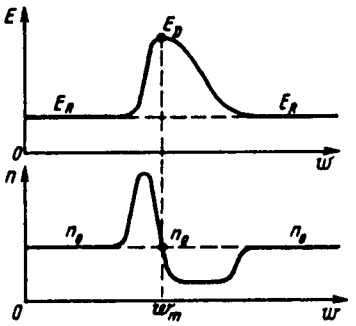


Рис. 9.11. Напряженность электрического поля и плотность электронов внутри стационарного домена

тическая характеристика дрейфовой скорости электрона $\overline{v}(E)$, исходя из выражения для плотности тока

$$\mathbf{J} = en\overline{\mathbf{v}} - e\,\mathrm{grad}\,(Dn) + \dot{\varepsilon}\frac{\partial \mathbf{E}}{\partial t} \quad (9.16)$$

и теоремы Гаусса

$$\mathrm{div}\,\mathbf{E} = e(n - n_0)/\varepsilon. \quad (9.17)$$

При этом предполагаем, что домен сильного поля движется без изменения формы со скоростью v_D, окружен нейтральным материалом, в котором плотность электронов $n_0 = n_a + n_b$, их скорость v_R и напряженность поля E_R. Тогда внутри домена E и n являются функциями переменной (9.15). Примерный вид этих функций представлен на рис. 9.11.

Если учесть, что $\partial w/\partial x = 1$, то из уравнений (9.17)

$$\frac{\partial E}{\partial x} = \frac{e}{\varepsilon}(n - n_0) = \frac{\partial E}{\partial w}. \quad (9.18)$$

Если учесть, что $\partial w/\partial t = -v_D$ и вне домена плотность тока равна en_0v_R, то согласно (9.16)

$$J = en\overline{v} - e\frac{\mathrm{d}(nD)}{\mathrm{d}w} - v_D\varepsilon\frac{\mathrm{d}E}{\mathrm{d}w} = en_0v_R,$$

или с учетом (9.18)

$$\frac{\mathrm{d}(nD)}{\mathrm{d}w} = n\overline{v} - v_D(n - n_0) - n_0v_R = n(\overline{v} - v_D) + n_0(v_D - v_R). \quad (9.19)$$

Переменную w можно исключить, разделив уравнение (9.19) на уравнение (9.18)

$$\frac{\mathrm{d}(nD)}{\mathrm{d}E} = \frac{\varepsilon}{e}\frac{n(\overline{v} - v_D) + n_0(v_D - v_R)}{n - n_0}. \quad (9.20)$$

Решение этого уравнения в общем случае может быть произведено только численными методами. С целью выявления основных черт рассматриваемых явлений и упрощения расчета при дальнейшем рассмотрении предполагаем, что коэффициент диффузии не зависит от напряженности электрического поля. Тогда из последнего уравнения

$$\frac{n}{n_0} - \ln\frac{n}{n_0} - 1 = \frac{\varepsilon}{en_0 D}\int_{E_R}^{E}\left[\overline{v} - v_D - \frac{n_0}{n}(v_R - v_D)\right] dE, \qquad (9.21)$$

в чем можно убедиться дифференцированием по E.

На границе обедненного и обогащенного слоев (при $E = E_D$) концентрации электронов $n = n_0$ (см. рис. 9.11); левая часть уравнения (9.21) обращается в нуль. Таким образом, и интеграл в правой части (9.21) при $E = E_D$ должен обращаться в нуль, причем независимо от того, с какой стороны вычисляется интеграл: со стороны обедненного ($n < n_0$) или со стороны обогащенного ($n > n_0$) участка. При интегрировании будем получать одинаковый результат. И следовательно, интеграл не будет зависеть от n только в том случае, когда

$$v_D = v_R. \qquad (9.22)$$

Отсюда следует вывод: *скорость домена равна дрейфовой скорости электронов вне домена*, как это и должно быть для стационарного домена.

Другой фундаментальный вывод следует из того, что, согласно (9.21) и (9.22), при $E = E_D$

$$\int_{E_R}^{E_D}(\overline{v} - v_D)\, dE = 0. \qquad (9.23)$$

Следовательно, напряженность поля в домене E_D может быть найдена из простого геометрического построения, при котором площади двух заштрихованных областей должны быть одинаковы (рис. 9.12). Это условие получило название *правила равных площадей*.

На рис. 9.12 с помощью правила равных площадей построена так называемая *динамическая характеристика* (пунктирная линия), которая представляет собой зависимость максимальной напряженности электрического поля в домене E_D от скорости дрейфа электронов вне домена v_R (и скорости домена v_D). Когда E_R уменьшается, скорость $v_D = v_R$ тоже уменьшается, а напряженность E_D возрастает. Существует минимальное значение напряженности поля вне домена E_{RM}, определяемое равенством $\overline{v}(E_{RM}) = v_{RM}$, ниже которого правило равных площадей не выполняется и распространение стабильных доменов невозможно.

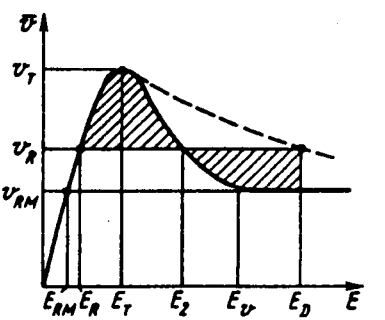


Рис. 9.12. Статическая (сплошная линия) и динамическая (пунктирная линия) характеристики домена

Форма домена, т. е. зависимость напряженности электрического поля от координаты (9.15), определяется из уравнения (9.18)

$$w = w_m + \frac{\varepsilon}{e}\int_{E_D}^{E} \frac{\mathrm{d}E}{n - n_0}, \qquad (9.24)$$

где w_n — координата максимума доменного поля в движущейся системе координат (9.15).

Точный расчет распределения поля по уравнению (9.24) в случае конечного значения D требует численного интегрирования и здесь не рассматривается.

Форма домена оказывается очень простой в том предельном случае, когда диффузией можно пренебречь. При этом следует исходить из того, что при $D \to 0$ и всех промежуточных значениях E между E_R и E_D правая часть уравнения (9.21) должна стремиться к бесконечности. В обедненном слое ($n < n_0$) левая его часть стремится к этому же пределу только при $n \to 0$, а в обогащенном слое — при $n \to \infty$. Следовательно, если не учитывать диффузию, на переднем фронте происходит полное обеднение слоя. При этом по уравнению (9.18) $\mathrm{d}E/\mathrm{d}x = -en_0/\varepsilon < 0$, т. е. напряженность E возрастает линейно и ширина обедненного слоя определяется расстоянием, на котором она возрастает от E_R до E_D:

$$d = \frac{\varepsilon}{en_0}(E_D - E_R). \qquad (9.25)$$

Поскольку на заднем фронте в обогащенном слое концентрация электронов оказывается в рассматриваемых условиях бесконечно большой и согласно (9.18)

$$\frac{\partial E}{\partial x} = \frac{e}{\varepsilon}(n - n_0) \to \infty,$$

то возрастание E от E_R до E_D происходит на бесконечно малом расстоянии, ширина обогащенного слоя равна нулю, полная ширина домена определяется только шириной обедненного слоя (9.25). При этом домен имеет треугольную форму (рис. 9.13). При уменьшении E_R ($E_R > E_{RM}$) максимальное поле домена E_D возрастает и треугольный обедненный слой увеличивается как по высоте, так и по ширине.

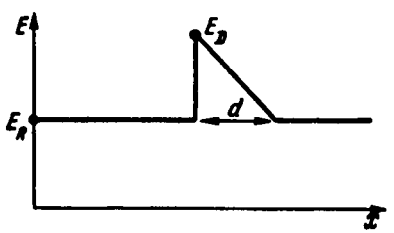


Рис. 9.13. Форма домена в предельном случае нулевой диффузии

Если длина образца равна l, то общее напряжение на нем

$$U = \int_0^l E\,\mathrm{d}w = E_R l + \int_0^l (E - E_R)\,\mathrm{d}w = E_R l + U_D, \qquad (9.26)$$

где U_D — избыточное напряжение на домене.

Для треугольного домена

$$U_D = \int_0^l (E - E_R)\,\mathrm{d}w = \frac{E_D - E_R}{2} d = \frac{\varepsilon}{en_0}\frac{(E_D - E_R)^2}{2}. \qquad (9.27)$$

Отметим, что напряжение на домене (9.27) обратно пропорционально n_0, т. е. прямо пропорционально удельному сопротивлению в слабых полях $\rho_0 = (en_0\mu_0)^{-1}$.

Для определения избыточного напряжения на домене (9.27), максимальной напряженности поля E_D и скорости домена по известному напряжению (9.26) необходимо знать форму статической характеристики $\overline{v}(E)$.

Приведенная приближенная оценка параметров домена не учитывала диффузию и зависимость коэффициента диффузии электронов от напряженности электрического поля. При таком подходе была достигнута простота анализа при малых неточностях в его физическом содержании, поскольку основным условием образования домена является наличие на статической характеристике участка с отрицательной проводимостью, а диффузия имеет второстепенное значение. Расчеты на ЭВМ с учетом диффузии показывают, что домен имеет сглаженную треугольную форму, в сравнительно слабых полях (близких к $E_т$) он становится симметричным.

Для образования стационарного домена не всегда достаточно, чтобы средняя напряженность электрического поля была выше значения $E_т$, соответствующего появлению отрицательной электропроводности. Необходимо еще, чтобы за время распространения волны объемного заряда от катода к аноду $T_0 \approx l/v_D$ успел сформироваться домен сильного поля. Время формирования домена имеет величину порядка времени релаксации в области с отрицательной дифференциальной проводимостью $\sigma_d = en_0\mu_{dm}$ ($\mu_{dm} = |\mu_{d\max}|$ — максимальное значение дифференциальной подвижности), т. е. $\tau_{dm} = \varepsilon/\sigma_d = \varepsilon/(en_0\mu_{dm})$. Чтобы в образце успел сформироваться стабильный домен, необходимо выполнение условия $\tau_{dm} < T_0$ или

$$n_0 l > \varepsilon v_D/(e\mu_{dm}) = (n_0 l)_1. \qquad (9.28)$$

Условие (9.28) иногда называют *критерием Кремера.*

Произведение равновесной концентрации носителей заряда n_0 на длину образца l (параметр $n_0 l$) определяет соотношение между временем формирования домена и временем пролета. Если параметр велик, то уже при малом превышении напряженности поля напряженности $E_т$, т. е. при малом значении $|\mu_d|$, условие (9.28) оказывается выполненным и в результате возникает домен. При малых значениях параметра, близких к $(n_0 l)_1$, для выполнения условия (9.28) нужно большое значение $|\mu_d|$ и, следовательно, заметное превышение порогового значения напряженности поля возникновения стационарного домена E_p над напряженностью $E_т$.

Если в образце, в котором распространяется домен, снизить среднее значение поля до значения, меньшего порогового $E_т$, то домен, не разрушаясь, будет продолжать двигаться к аноду. Домен начнет рассасываться только в том случае, когда поле станет меньше порогового поля исчезновения домена E_a. С увеличением $n_0 l$ напряженность этого поля монотонно убывает и при $n_0 l \gg (n_0 l)_1$ достигает значения E_{RM}, составляющего в арсениде галлия 1,25 кВ/см.

Различие в порогах возникновения и исчезновения домена определяет возможность работы *диода Ганна* (однородный кристалл с двумя контактами) в режиме, называемом *триггерным*. В этом режиме постоянное поле (поле смещения) ниже порога возникновения домена, но выше порога его исчезновения. Дополнительно к напряжению смещения подается импульс напряжения длительностью, большей времени формирования домена, но меньшей времени пролета домена. Суммарное напряжение смещения и импульса больше порога возникновения домена. В таких условиях в диоде возникает один домен, достигающий анода. Таким образом, в триггерном режиме формируется одиночный импульс тока длительностью, примерно равной времени пролета домена в образце.

Плотность тока в образце с доменом может быть выражена через параметры кристалла вне домена, т. е. подвижность μ_a в области положительной дифференциальной проводимости и напряженность поля $E_R < E_т$ вне домена: $J = e\mu_a E_R$. При малых значениях параметра $n_0 l \approx (n_0 l)_1$ напряженность поля E_R монотонно убывает с ростом напряжения смещения в интервале от $E_т$ до E_{RM}, а затем остается постоянной, равной E_{RM}. Для $n_0 l \gg (n_0 l)_1$ напряженность поля $E_R = E_{RM}$ перестает зависеть от приложенного напряжения и средней напряженности поля $\overline{E} = U/l$. Следовательно, при большом напряжении смещения и (или) при достаточно большом параметре $n_0 l$ (в арсениде галлия при $n_0 l > 2 \cdot 10^{12}$ см$^{-2}$) напряженность поля вне домена постоянна и примерно равна E_{RM}. Это означает, что плотность тока практически не зависит от площади поперечного сечения, а полный ток в образце

$I = e\mu_a E_{RM} n_0(x) S(x)$ в каждом цикле прохождения домена меняется по времени $I(t)$ по тому же закону, по какому в функции координаты x изменяются уровень легирования $n_0(x)$ и поперечное сечение $S(x)$ образца. Таким образом может быть воспроизведена (смоделирована) любая функция $f(x) = n_0(x)S(x)$. Это свойство профилированных или неоднородно легированных диодов Ганна позволяет использовать их в качестве генераторов колебаний тока сложной формы (рис. 9.14). При однородном легировании форма колебаний тока в течение пролетного времени воспроизводит форму профиля поперечного сечения диода.

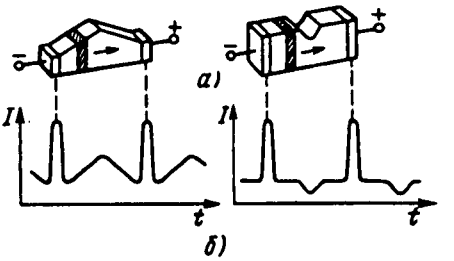


Рис. 9.14. Функциональные схемы генератора Ганна с заданной формой колебаний:

а — форма образцов; *б* — зависимости тока от времени

§ 9.8. Логические функциональные схемы на основе эффекта Ганна

Возможность использования диода Ганна как элемента логических схем основана на двух его состояниях — омическом и доменном. Такие логические схемы обладают высоким быстродействием, достигающим нескольких десятков наносекунд на каскад. Их параметры проще всего оценить на примере схемы рис. 9.15. Диод в этой схеме работает в триггерном режиме. Напряжение $U_0 = E_0 - IR_н$ меньше порога возникновения, но больше порога исчезновения домена. После подачи на вход запускающего импульса и возникновения домена ток в диоде и нагрузке падает, образуя выходной импульс с полярностью, противоположной полярности входного импульса, и длительностью, равной пролетному времени домена. Схема может выполнять логическую операцию сравнения амплитуды импульса $U_и$ с заданной величиной $U_т - U_0$, функцию дискриминатора входных импульсов по их ширине и амплитуде, а также элемента ИЛИ при наличии дополнительного входа, показанного пунк-

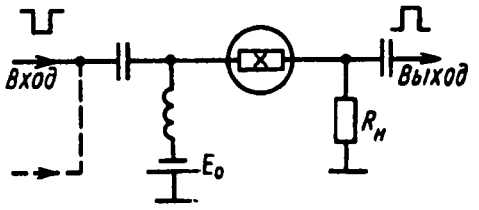


Рис. 9.15. Схема импульсного усилителя

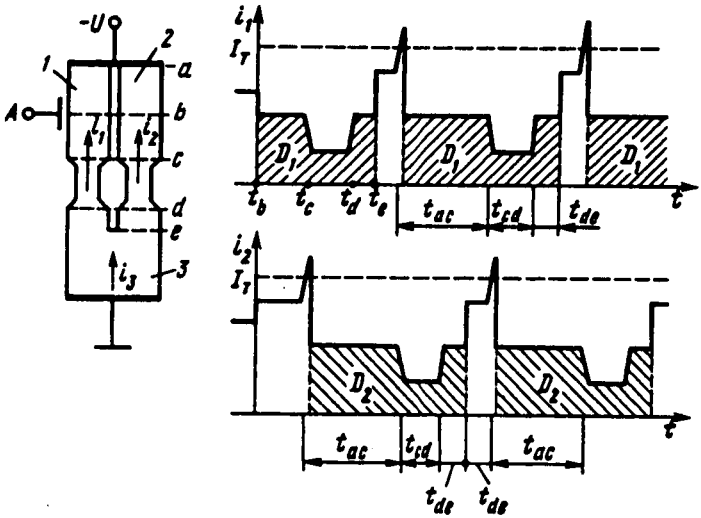


Рис. 9.16. Динамический триггер и его характеристики

тиром (если диод переключается одним импульсом, поданным на любой из входов). В сильно легированном диоде из арсенида галлия ($n_0 = 5 \cdot 10^{15}$ см$^{-3}$) длиной $3 \cdot 10^{-3}$ см время переключения $6 \cdot 10^{-11}$ с. В другом диоде ($n_0 = 2 \cdot 10^{15}$ см$^{-3}$, $l = 5 \cdot 10^{-3}$ см) в импульсном режиме был получен коэффициент усиления $10 \lg \Delta I R_н / U_0 = 32$ дБ. Минимальная мощность переключения диода с поперечным размером 3 мкм составляет $5 \cdot 10^{-4}$ Вт, произведение потребляемой мощности на время переключения — $2{,}5 \cdot 10^{-12}$ Дж.

Функции *ячейки памяти* выполняют два каскада усилителей типа показанных на рис. 9.15 с положительной обратной связью между входом и выходом (входной и выходной импульсы имеют в этом случае одинаковую полярность). Такие схемы обладали временем включения 1 нс, обеспечивая самоподдерживающуюся генерацию двухнаносекундных импульсов с интервалом 10 нс.

Аналогичную функцию может выполнять интегральная схема *динамического триггера*. Она реализуется в профилированном кристалле, имеющем две ветви *1* и *2* (рис. 9.16), в каждой из которых возможно существование домена сильного поля. В исходном состоянии («0») на катод подано напряжение несколько ниже порогового. Поэтому ни в одной из ветвей домен не образуется. Запускающий импульс, подаваемый на вход *A* в момент t_b, увеличивает в ветви *1* напряженность поля выше порогового уровня, образуя домен D_1. Возникновение домена сопровождается уменьшением тока i_1 и увеличением тока i_2, однако не таким, чтобы во

второй ветви возник домен. В момент t_c домен доходит до сечения c и, так как ширина последующего участка меньше предыдущего, ток i_1 снова уменьшается, а ток i_2 увеличивается до значения выше порогового, в результате чего в ветви *2* образуется домен D_2.

Домен D_1 исчезает в момент t_e, дойдя до сечения e, так как ширина ветви *3* значительно больше, чем ветвей *1* и *2*. Новый домен в ветви *1* образуется в тот момент, когда домен доходит до выреза в ветви *2*. Далее процесс повторяется и прибор может быть выведен из состояния генерации («1») только снижением напряжения на катоде. Для правильной работы триггера необходимо, чтобы длина участка *a*—*c* превышала длину участка *c*—*e*.

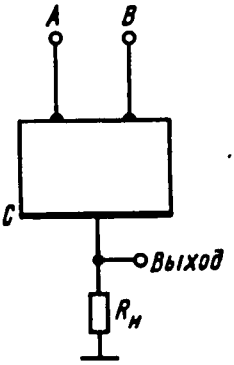


Рис. 9.17. Компаратор

Подобный принцип может быть положен в основу работы нейристорной линии и для создания регистра сдвига.

Другим примером того, как использование эффекта Ганна расширяет возможности применения многоэлектродных функциональных интегральных схем, может служить *компаратор* (рис. 9.17). Его действие основано на неравномерности электрического поля вблизи двух катодов (*A* и *B*) малых размеров. Поле вблизи электрода *C*, подсоединяемого к положительному полюсу источника смещения, практически однородно.

Если к отрицательному полюсу источника смещения подсоединен только один из малых электродов (например, *A*), то напряженность электрического поля около него выше порогового значения, необходимого для образования и продвижения домена. Импульс тока на выходе длится все время, пока домен распространяется вдоль всего образца.

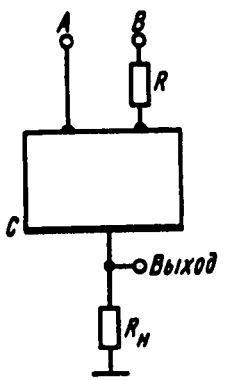


Рис. 9.18. Схема запрета

Если источник смещения присоединяется одновременно к двум электродам *A* и *B*, то за счет выравнивания поле вблизи катодов уменьшается (увеличивается эффективная площадь катода) до уровня ниже порогового, домен не возникает и сигнал на выходе схемы отсутствует. Поэтому прибор осуществляет функции схемы сравнения.

После включения между катодом *B* и источником напряжения резистора с соответствующим сопротивлением (или использования катодов разных размеров) рассмотренный компаратор сможет осуществить функции *схемы*

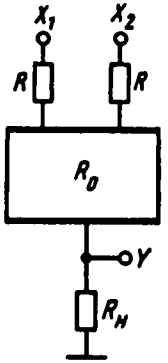


Рис. 9.19. Схема совпадений

запрета (рис. 9.18). Если на вход *A* подать напряжение достаточной амплитуды, то образуется домен и на выходе прибора появляется импульс. Однако если тот же потенциал подан на другой вход, то вследствие падения напряжения на резисторе *R* максимальное поле вблизи электрода *B* недостаточно для возникновения домена и на выходе сигнал отсутствует. Если же напряжение подано на оба входа, то в соответствии со свойствами компаратора сигнала на выходе также не будет.

Любой диод Ганна может выполнять функции *схемы совпадений* (И), если его катод соединить с двумя входными зажимами через одинаковые резисторы *R* нужного сопротивления (рис. 9.19). Если в этой схеме напряжение *U* источника подано только на один из входов, то напряжение на диоде

$$U_1 = \frac{R_0}{R + R_0 + R_н} U < U_т$$

меньше порогового ($U_т$), при котором образуется домен, и выходной сигнал отсутствует. Если же напряжение подано одновременно на оба входа, то напряжение на диоде

$$U_2 = \frac{R_0}{R/2 + R_0 + R_н} U > U_т$$

и на выходе прибора появляется сигнал.

Одним из функциональных приборов, которые могут быть выполнены на диоде Ганна планарной конструкции, является *аналого-цифровой преобразователь* (рис. 9.20). Его активный слой в виде клина с кодирующими прорезями выращен с помощью эпитаксии на полуизолирующей подложке. В приборе используется два явления: 1) зависимость длины пробега домена от напряжения между электродами (за счет снижения напряженности электрического поля по мере расширения поперечного сечения активного слоя до уровня ниже порога исчезновения домена); 2) зависимость тока домена от площади поперечного сечения (при прохождении доменом кодирующей прорези ток уменьшается). В результате с увеличением напряжения смещения растет путь, проходимый доменом, и соответственно увеличивается число всплесков тока на один пробег домена. Существенным преимуществом

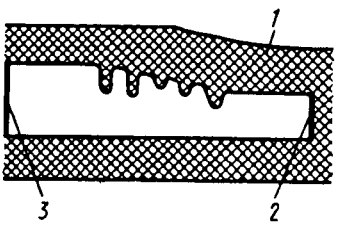


Рис. 9.20. Аналого-цифровой преобразователь:

1 — полуизолирующая подложка; *2* — анод; *3* — катод

планарных функциональных приборов Ганна является возможность их совмещения с обычными интегральными схемами. Подобные интегральные схемы иногда называют DOFIC (domain oriented functional integrated circuits).

Как видно из рассмотренных примеров, особенностью интегральных схем на основе эффекта Ганна является возможность разработки простых элементов с совокупностью свойств, которые в других случаях могут быть получены за счет усложнения обычных схем. Это преимущество обусловлено многочисленными возможностями управления образованием и перемещением домена. Важным свойством интегральных схем на основе эффекта Ганна является наличие механизма усиления при образовании домена, в результате чего между каскадами с логическими схемами не нужно вводить усилители.

§ 9.9. Нарастающие волны объемного заряда

В полупроводниковом кристалле с отрицательной дифференциальной проводимостью возможно существование нескольких видов (мод) поведения твердотельной плазмы, образуемой свободными носителями заряда. Рассмотренный ранее домен сильного поля является одним из них. Другой вид — нарастающая волна объемного заряда — возникает при напряженностях поля выше порогового значения ($E > E_т$) и малых параметрах $n_0 l < (n_0 l)_1$, когда плотность объемного заряда доноров недостаточна для формирования домена за время пролета. Именно эта мода в большинстве случаев определяет наблюдаемое усиление электромагнитных колебаний в объеме рассматриваемых кристаллов.

Для объяснения процесса усиления можно воспользоваться тем упрощенным анализом, который дает линейная теория при рассмотрении малых отклонений от равновесных решений (методом малых возмущений).

«Малосигнальное решение» получают путем разделения каждой из искомых функций на установившуюся (однородную) и нестационарную (неоднородную) части

$$\mathbf{E}(z,t) = \mathbf{E}_0 + \mathbf{e}_1(z,t), \quad \rho(z,t) = \rho_0 + \rho_1(z,t);$$
$$\mathbf{u}(z,t) = \mathbf{u}_0 + \mathbf{u}_1(z,t), \quad \mathbf{J}(z,t) = \mathbf{J}_0 + \mathbf{j}_1(z,t)$$

в предположении, что малые возмущения, например, $|e_1| << E_0$ и $|j_1| << J_0$, изменяются по гармоническому закону.

Рассмотрим особенности распространения волн пространственного заряда ρ_1 в безграничной плазме полупроводников с учетом дрейфа носителей заряда и их нагрева статическим электрическим полем напряженностью $\mathbf{E}_0 = E_0 \mathbf{e}_z$.

Исходными для анализа электромагнитных процессов в полупроводнике с плотностями заряда ρ_1 и тока j_1 служат уравнения Максвелла. Решив их относительно напряженности электрического поля, можно получить сначала неоднородное волновое уравнение

$$\nabla^2\mathbf{E}_1 + \frac{1}{v^2}\frac{\partial^2\mathbf{E}_1}{\partial t^2} = \mu\frac{\partial \mathbf{j}_1}{\partial t} + \frac{\nabla\rho_1}{\varepsilon},$$

а затем — уравнение Гельмгольца

$$\nabla^2\dot{\mathbf{E}}_1 + k_1^2\dot{\mathbf{E}}_1 = j\omega\mu\dot{\mathbf{J}}_1 + \nabla\dot{\rho}_1/\varepsilon, \tag{9.29}$$

где плотности заряда и тока связаны уравнением непрерывности

$$j\omega\dot{\rho}_1 + \nabla\dot{\mathbf{J}}_1 = 0, \tag{9.30}$$

а плотность тока

$$\mathbf{j}_1 = \rho_0\mathbf{u}_1 + \rho_1\mathbf{u}_0 = \sigma\mathbf{E}_1 + D_e\mathbf{1}\nabla\rho_1 + \rho_1\mathbf{u}_0, \tag{9.31}$$

причем

$$\mathbf{1} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}; \qquad \sigma = \begin{vmatrix} \sigma_e & 0 & 0 \\ 0 & \sigma_e & 0 \\ 0 & 0 & \sigma_d \end{vmatrix};$$

$k_1 = \omega/v = \omega\sqrt{\varepsilon\mu}$ — волновое число решетки.

Подставив выражение (9.31) в (9.30) и используя теорему Гаусса $\nabla\mathbf{D} = \rho$, придем к уравнению для плотности заряда

$$\dot{D}_e^2\nabla^2\dot{\rho}_1 - (u_1\nabla)\dot{\rho}_1 - (\omega_м + j\omega)\dot{\rho}_1 = (\sigma_d - \sigma_e)\frac{\partial\dot{E}_{1z}}{\partial z}, \tag{9.32}$$

где $\omega_м = \varepsilon/\sigma_e$ — «максвелловская» частота релаксации.

Исключив из уравнения (9.29) ток с учетом (9.30) и (9.31), получим уравнения для компонент электрического поля:

$$\begin{aligned}
\nabla^2\dot{E}_{1x} + k_1^2\left(1 - j\frac{\omega_м}{\omega}\right)\dot{E}_{1x} &= \frac{1}{\varepsilon}\left(1 - \frac{u_0^2}{v^2}\frac{\omega}{\omega_D}\right)\frac{\partial\dot{\rho}_1}{\partial x} \approx \frac{1}{\varepsilon}\frac{\partial\dot{\rho}_1}{\partial x}; \\
\nabla^2\dot{E}_{1y} + k_1^2\left(1 - j\frac{\omega_м}{\omega}\right)\dot{E}_{1y} &= \frac{1}{\varepsilon}\left(1 - \frac{u_0^2}{v^2}\frac{\omega}{\omega_D}\right)\frac{\partial\dot{\rho}_1}{\partial y} \approx \frac{1}{\varepsilon}\frac{\partial\dot{\rho}_1}{\partial y}; \\
\nabla^2\dot{E}_{1z} + k_1^2\left(1 - \frac{\mu_d}{\mu_a}\frac{\omega_м}{\omega}\right)\dot{E}_{1z} &= \frac{1}{\varepsilon}\left(1 - \frac{u_0^2}{v^2}\frac{\omega}{\omega_D}\right)\frac{\partial\dot{\rho}_1}{\partial z} + j\frac{u_0^2}{v^2}\frac{\omega}{\varepsilon u_0}\dot{\rho}_1 \approx \frac{1}{\varepsilon}\frac{\partial\dot{\rho}_1}{\partial z},
\end{aligned} \tag{9.33}$$

где $\omega_D = u_0^2/D_e$ — «диффузионная частота» (для реальных полупроводников в диапазоне СВЧ $\omega \leqslant \omega_D$ и $u_0 << v$).

Уравнения (9.32) и (9.33) применимы для описания волновых процессов как в безграничном полупроводнике, так и в тонкопленочных структурах. В случае однородной плоской волны они упрощаются:

$$(k^2-\eta_\perp^2)\varepsilon\dot{E}_{1x}=jk_x\dot{\rho}_1; \tag{9.34}$$

$$(k^2-\eta_z^2)\varepsilon\dot{E}_{1z}=jk_z\dot{\rho}_1; \tag{9.35}$$

$$(k^2-\eta_z^2)\dot{\rho}_1=-jk_z(1-k_0)\beta_м\beta_D\varepsilon\dot{E}_{1z}, \tag{9.36}$$

где для волнового числа полупроводника с плазмой $k^2=k_x^2+k_y^2+k_z^2=$ $=k_\perp^2+k_z^2$; $\eta_\perp^2=k_1^2(1-jk_0\,\omega_м/\omega)$; $\eta_z^2=k_1^2(1-k_0\,\omega_м/\omega)$; $\eta_D^2=\beta_D[jk_z-$ $-(\beta_м+j\beta_e)]$, $\beta_e=\omega/u_0$; $\beta_м=\omega_м/u_0$; $\beta_D=\omega_D/u_0$; $k_0=\mu_d/\mu_e$.

Приравнивая к нулю определитель системы (9.34)—(9.36), получим дисперсионное уравнение

$$(k^2-\eta_\perp^2)[(k^2-\eta_z^2)(k^2-\eta_D^2)-(1-k_0)\beta_м\beta_Dk_z^2]=0,$$

которое распадается на два независимых уравнения

$$k^2-\eta_\perp^2=0; \tag{9.37}$$

$$(k^2-\eta_z^2)(k^2-\eta_D^2)-(1-k_0)\beta_м\beta_Dk_z^2=0. \tag{9.38}$$

Уравнение (9.37) соответствует вихревым (соленоидальным) волнам в плазме, когда $\rho_1=0$ и $E_{1z}=0$. Перепишем это выражение в виде

$$k^2=k_1^2(1-j\,\omega_м/\omega).$$

Оно не содержит σ_e, u_0 и D_e, т. е. эти волны не «чувствуют» ни скорость дрейфа, ни диффузию, ни анизотропию среды (k_0). Свободные носители заряда приводят лишь к затуханию волны.

Дисперсионное уравнение (9.38) соответствует двум типам волн: быстрым вихревым и медленным квазистатическим.

Быстрые электромагнитные волны с фазовой скоростью $v_\phi\approx v_2$ не имеют продольной компоненты $H_{1z}=0$. В этом случае $k_z^2\ll\beta_e^2$, $k^2<<\eta_D^2\approx-\beta_D(\beta_м+j\beta_e)$ и

$$k^2=\eta_z\frac{\omega_м+j\omega}{(k_0\cos^2\theta+\sin^2\theta)\omega_м+j\omega}=\begin{cases}\eta_\perp^2 & \text{при } \theta=0;\\ \eta_z^2 & \text{при } \theta=90°,\end{cases}$$

где θ — угол между направлением волнового вектора **k** и направлением дрейфа носителей (осью z). Следовательно, быстрые волны не «чувствуют» дрейфа (u_0) и диффузии (D_e), но зависят от разогрева электронов (k_0) и степени анизотропии среды (θ). При увеличении отклонения волны (θ)

от продольного направления ($\theta = 0$) происходит переход от затухания к усилению волны.

Наибольший интерес представляет квазистатическое решение уравнения (9.38), соответствующее *медленным волнам* пространственного заряда (ВПЗ) с фазовыми скоростями $u_\phi \approx u_0$. В этом случае $k^2 \gg \eta_=^2 \sim \eta_\perp^2$. Тогда согласно (9.38),

$$k^2 = \eta_D^2 + (1 - k_0)\beta_м\beta_D \cos^2\theta;$$
$$\gamma^2 + \beta_D\gamma\cos\theta - (k_0\cos^2\theta + \sin^2\theta) - j\beta_e\beta_D = 0; \qquad (9.39)$$
$$\gamma^2 + \beta_D\gamma\cos\theta - \beta_D\gamma_0\cos\theta = 0,$$

где $\gamma = jk = \alpha + j\beta$ — коэффициент распространения волны пространственного заряда; $\gamma_0 = \alpha_0 + j\beta_0 = \beta_м(k_0\cos^2\theta + \sin^2\theta)/\cos\theta + j\beta_e/\cos\theta$ — коэффициент распространения этой волны при нулевой диффузии ($D_e \to 0$, $\beta_D \to \infty$).

Если диффузия достаточно слаба, т. е. $|\gamma_0| << \beta_D\cos\theta = u_0\cos\theta/D_e$, то корни уравнения (9.39) имеют вид

$$\gamma_1 = \gamma_0 - \frac{\gamma_0^2}{\beta_D\cos\theta} \approx \alpha_0 + \frac{\beta_e^2/\cos^2\theta}{\beta_D\cos\theta} + j\beta_0; \qquad (9.40)$$

$$\gamma_2 = -\beta_D\cos\theta - \gamma_1 \approx -\beta_D\cos\theta - \gamma_0 \approx -\beta_D\cos\theta - j\beta_0. \qquad (9.41)$$

Корень уравнения (9.40) соответствует прямой волне, которую называют *сносовой*. Она будет слабо затухающей или нарастающей, если выбором проводимости среды обеспечено условие $|\alpha_1| \ll |\beta_1|$. При наличии отрицательной дифференциальной проводимости возникает пространственное усиление волны с максимальным коэффициентом усиления

$$|\alpha_1|_{max} = |k_0|\beta_м = \frac{|\mu_d|}{\mu_e}\frac{\omega_м}{u_0}$$

вдоль направления дрейфа. По мере отклонения от этого направления усиление уменьшается, доходя до нуля при критическом угле $\theta_{кр}$, а затем сменяется затуханием, которое увеличивается вплоть до $\alpha \to \infty$ при $\theta = 90°$.

Корень уравнения (9.41) соответствует обратной волне, которая распространяется с той же фазовой скоростью, что и прямая волна. Обратная волна порождена диффузионным размытием зарядового возмущения в направлении, обратном дрейфу носителей, заряда, который переносит эти возмущения в виде волны пространственного заряда. При нулевой диффузии ($D_e = 0$) корень γ_2 отсутствует. Поэтому обратную волну называют

диффузионной. Эта волна всегда (в том числе и при $\mu_d < 0$) является сильно затухающей, поскольку, согласно (9.41), для нее $|\alpha_2| \gg |\beta_2|$, когда $\beta_D >> \beta_e$.

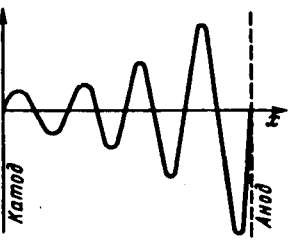


Рис. 9.21. Нарастающая волна объемного заряда

Следовательно, волнам пространственного заряда свойственна невзаимность: на большие расстояния от области возбуждения зарядовые возмущения переносятся лишь прямыми волнами, обеспечивая «развязку» в обратном направлении. Исключение составляют поперечные волны пространственного заряда ($\theta = 90°$), так как в этом направлении действует диффузионный механизм передачи возмущения, а механизм сноса не работает. При $D_e \to 0$, $\alpha \to \infty$, $v_\phi \to 0$.

Таким образом, волну объемного заряда с нарастающей амплитудой (рис. 9.21) создает только среда с отрицательной проводимостью; именно этот механизм нарастания вызывает усиление электромагнитных колебаний, без которого невозможен и процесс генерации.

Для нарастающих волн объемного заряда эквивалентное сопротивление полупроводникового образца можно найти из выражения для переменной составляющей плотности полного тока

$$j_{1\text{полн}} = j_1 + \varepsilon \frac{\partial e_1}{\partial t},$$

если при этом учесть, что она (вследствие непрерывности полного тока) не должна зависеть от координаты x и соответствовать переменной составляющей i_1 тока во внешней цепи:

$$j_1 + \varepsilon \frac{\partial e_1}{\partial t} = \frac{i_1}{S}, \qquad (9.42)$$

где S — площадь поперечного сечения образца.

Переменные составляющие, как и решение всякого неоднородного дифференциального уравнения (с производными по координате x), можно представить как сумму принужденной и свободной составляющих $\dot{J}_1 = \dot{J}_{\text{пр}} + \dot{J}_{\text{св}}$, $\dot{E}_1 = \dot{E}_{\text{пр}} + \dot{E}_{\text{св}}$, из которых свободные составляющие

$$\dot{J}_{\text{св}} = A e^{j(\omega t - k x)}, \qquad \dot{E}_{\text{св}} = \dot{B} e^{j\omega t}, \qquad (9.43)$$

зависят от координат, а принужденные составляющие

$$\dot{E}_{\text{пр}} = \dot{E}_m e^{j\omega t}, \qquad \dot{J}_{\text{пр}} = \dot{J}_m e^{j\omega t},$$

от них не зависят.

Не учитывая в первом приближении диффузионную составляющую и учитывая, что принужденная составляющая $J_{пр} = \sigma_d E_{пр}$, согласно (9.43) и (9.42), получим

$$\sigma_d \dot{E}_m + A e^{-jkx} + j\omega\varepsilon(\dot{E}_m + B e^{-jkx}) \approx \dot{I}_1 / S,$$

причем из граничного условия $E_m = 0$ при $x = 0$ $A + j\omega\varepsilon B = \dot{I}_1 / S$. Следовательно,

$$\dot{E}_m \approx \frac{1 - e^{-jkx}}{\sigma_d + j\omega\varepsilon} \frac{\dot{I}_1}{S}$$

и переменная составляющая напряжения на образце

$$\dot{U}_m = \int_0^l \dot{E}_m \, dx = \frac{jkl - 1 + e^{-jkl}}{jk(\sigma_d + j\omega\varepsilon)S} \dot{I}_1.$$

Отсюда малосигнальное комплексное сопротивление однородного образца

$$Z(\omega) \approx \frac{\dot{U}_m}{\dot{I}_1} = \frac{\alpha - 1 + e^{-\alpha}}{\alpha(\sigma_d + j\omega\varepsilon)S} l, \qquad (9.44)$$

где

$$\alpha = (1/\tau_d - j\omega) l / v = T_0 / \tau_d - j2\pi T_0 / T, \qquad (9.45)$$

T_0 — пролетное время; $T = f^{-1}$ — период колебаний тока.

Выражение (9.44) позволяет оценить границу между режимами усиления и генерации. Она будет определяться границей устойчивости образца при питании его от генератора напряжения. В этом случае для устойчивости образца необходимо, чтобы в верхней полуплоскости не было полюсов проводимости $Y(\omega) = 1/Z(\omega)$, т. е. нулей Z. Определенные численными методами корни α_m уравнения $\alpha - 1 + e^{-\alpha} = 0$ приведены в табл. 9.1 (m — номер корня).

Таблица 9.1

m	α_m
1	$-2{,}09 \pm j7{,}46$
2	$-2{,}69 \pm j13{,}88$
3	$-3{,}03 \pm j20{,}22$
4	$-3{,}22 \pm j26{,}54$
5	$-3{,}50 \pm j32{,}85$

Из сравнения (9.45) с данными, приведенными в табл. 9.1, легко видеть, что $(T_0/T)_m = \mathrm{Im}\,\alpha_m/(2\pi) \approx m$, т. е. нули $Z(\omega)$ и связанная с ними генерация возникают вблизи частоты пролета и ее гармоник. Кроме того, для возникновения генерации необходимо, чтобы выражение $(T_0/\tau_d)_m = \mathrm{Re}\,\alpha_m < 0$, т. е. $\tau_d = \varepsilon/\sigma_d$ было отрицательно и по порядку величин соизмеримо с временем пролета. Расчет $Z(\omega)$ с учетом диффузии показывает, что диффузия уменьшает усиление тем больше, чем выше частота; для первой гармоники ее влияние пренебрежимо мало.

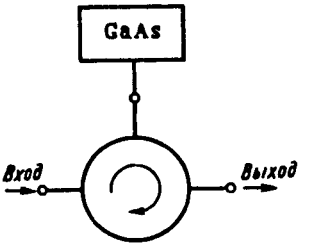


Рис. 9.22. Схема усилителя с циркулятором

Эффект усиления волн пространственного заряда лежит в основе работы полупроводниковых усилителей. Если в усилителе используется нарастающая волна объемного заряда, то такой режим иногда называют *режимом устойчивого отрицательного сопротивления*. В усилителях с отрицательным сопротивлением обычно используют циркуляторы (рис. 9.22) и двойные тройники.

Режим бегущей волны можно рассматривать как частный случай работы усилителя с устойчивым отрицательным сопротивлением, когда посредством устройства связи около катода образца со стационарным распределением электрического поля возбуждается волна объемного заряда, которая экспоненциально растет по мере движения и возбуждает в выходном устройстве у анода усиленные (относительно входа) колебания. Таким образом, прибор может обеспечить однонаправленное усиление без применения циркуляторов или двойных тройников. В усилителях с бегущей волной обычно используется полупроводниковый брусок длиной в несколько длин волн объемного заряда, что позволяет выполнить их в виде функциональной интегральной схемы (рис. 9.23).

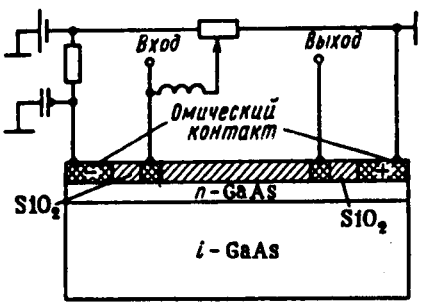


Рис. 9.23. Функциональная схема усилителя бегущей волны

Интересным свойством таких усилителей является то, что поскольку сигналы от входа к выходу передаются со скоростью дрейфа объемного заряда, то временна́я (фазовая) задержка сигналов на выходе обратно пропорциональна дрейфовой скорости носителей заряда. Таким образом, полупроводниковые усилители с бегущей волной можно использовать как линии задержки с электронной перестройкой или

фазовращатели с усилением. Регулировка усиления и фазы достигается изменением напряжения на приборе.

§ 9.10. Ограничение накопления объемного заряда

В определенной области полей диод Ганна с отрицательной дифференциальной проводимостью может служить генератором. Существенного увеличения КПД и частотного диапазона можно добиться при его работе в настраиваемой LCR-цепи. В диапазоне СВЧ такой цепью служит резонатор. При работе диода в резонаторе к нему помимо постоянного напряжения смещения U_0 (от внешнего источника) оказывается приложенным также напряжение СВЧ U_m, устанавливающееся в резонаторе за счет колебаний тока в диоде:

$$u = U_0 + U_m \sin \omega t = (E_0 + E_m \sin \omega t) l.$$

Если диод используется совместно с резонансным контуром или в полом резонаторе с высокой добротностью, колебания в объеме кристалла происходят в режиме больших сигналов. При этом один и тот же диод может генерировать на частотах в диапазоне от долей герца до сотен гигагерц. Этот диапазон перекрывается несколькими различными режимами работы.

Режимы работы генераторов, как и моды полупроводниковой плазмы, определяются концентрацией носителей заряда, их температурой, размерами образца, скоростью междолинного переноса и схемой, в которую включен кристалл. Диффузия также накладывает свои условия на тот предел n и l, ниже которого пространственный заряд рассасывается.

При использовании нарастающей волны объемного заряда генераторы, как и усилители, работают в режиме *устойчивого отрицательного сопротивления.*

При большой скорости нарастания бегущая волна пространственного заряда большой амплитуды образует стационарный домен в течение небольшой части периода колебаний. Если амплитуда переменного поля мала по сравнению с $E_0 - E_p$, то осуществляется *пролетный режим* (транзитная мода). Пролетный режим характеризуется наличием полностью сформировавшихся доменов, которые движутся через диод Ганна. В этом режиме эквивалентная схема активной части кристалла может быть представлена последовательным соединением двух элементов. Один из них, соответствующий внедоменной области, представляет собой сопротивление R_a в слабом электрическом поле и емкости C образца. Доменная область аппроксимируется параллельным соединением емкости домена

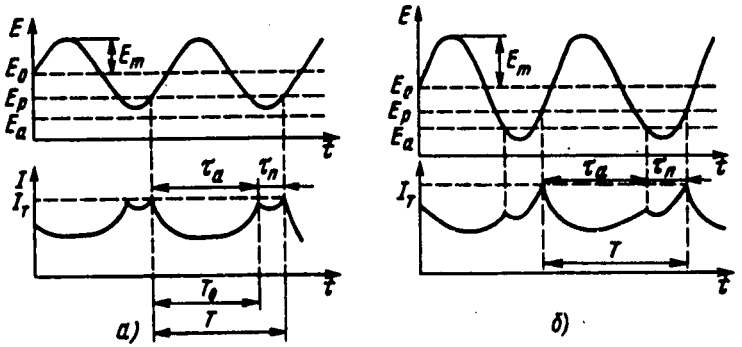


Рис. 9.24. Колебания напряжения и тока в запаздывающем режиме (*а*) и режиме гашения (*б*)

$C_д$ и его сопротивления $R_D < 0$. КПД генератора обычно низок (не превышает долей процента).

Частота генерации f определяется временем прохождения домена от места образования до соответствующего контакта. Поскольку домены, как правило, образуются у катода, это время равно времени прохождения домена через весь образец: $T = l/v_D = 1/f$. Скорость движения домена 10^7 см/с, поэтому при толщине образца 10 мкм время прохождения составит приблизительно 10^{-10} с и, следовательно, частота генератора — 10^{10} Гц.

Возможны и другие механизмы генерации, позволяющие получить колебания более высокой частоты, не связанные непосредственно с временем прохождения домена через весь образец. Если амплитуда переменной составляющей поля возрастает настолько, что выполняется условие $E_0 - E_p < E_m < E_0 - E_a$ (рис. 9.24, *а*), то домен, как и в пролетном режиме, исчезает, достигнув анода. Однако новый домен образуется после этого не сразу, поскольку напряженность поля сначала оказывается меньше порогового значения E_p, при котором возникает домен. Такой режим называют *запаздывающим* (режим с запаздыванием формирования домена). Частота колебаний при этом меньше, чем пролетная частота, она может быть перестроена путем изменения собственной частоты резонатора.

При дальнейшем повышении переменной составляющей поля до уровня $E_m > (E_0 - E_a)$ домен существует только часть периода T, после чего исчезает (при $E = E_a$), не дойдя до анода (рис. 9.24, *б*). Этот режим называют *режимом гашения* (режим с подавлением домена). Частоту колебаний в этом режиме также можно перестраивать с помощью резонатора, однако она может быть не только меньше, но и больше пролетной. Теоретический КПД генераторов, работающих в режиме гашения, может достигать 13%, а в режиме запаздывания — 27%.

Иногда все три рассмотренных режима (пролетный, запаздывающий и гашения) называют *ганновской* или *пролетной модой колебаний*. Для нее время формирования домена должно быть незначительным по сравнению с периодом колебаний. Если же период колебаний резонатора имеет порядок времени формирования домена и $E_m > (E_0 - E_a)$, то в диоде возникает *гибридный режим*. От режима гашения он отличается только тем, что домен начинает рассасываться, не успев сформироваться окончательно. В гибридном режиме на частоте 8,3 ГГц при импульсной мощности около 19 Вт генератор работает с КПД, равным 22%. Мощность 34 Вт была получена при КПД, равном 15%.

Если частота резонатора столь велика, что время формирования домена оказывается много больше периода электромагнитных колебаний с амплитудой $E_m > (E_0 - E_т)$, домены не успевают сформироваться и зависимость тока от напряженности поля повторяет статическую-характеристику для однородного образца (рис. 9.25). Этот режим получил название режима *ограничения накопления объемного заряда* (ОНОЗ). Он является одним из основных генераторных режимов работы диодов Ганна. Эффективность режима ОНОЗ на частотах порядка 10 ГГц и выше определяется отсутствием ограничений, связанных не только с временем пролета домена, но и с временем его формирования.

Диод отдает мощность во внешнюю цепь СВЧ, когда суммарное поле больше порогового ($E > E_т$), и является пассивным элементом, когда $E < E_т$. За отрезок времени τ_s (рис. 9.25) происходит рассасывание объемного заряда, накопленного за время активной части периода ($T - \tau_s$). Если объемный заряд не успевает рассасываться за время τ_s, то он может накопиться за несколько периодов, и генератор выйдет из режима ОНОЗ.

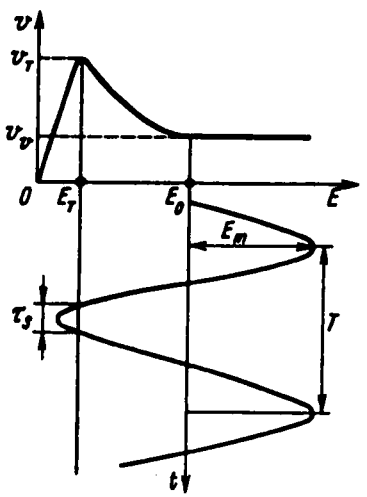


Рис. 9.25. Зависимость напряженности электрического поля от времени в режиме ОНОЗ

Предельные частоты генерации в режиме ОНОЗ ограничены уровнем, при котором разогрев носителей заряда и междолинный переход, определяющие установление статической характеристики, происходят за время, большее периода колебаний. Для GaAs зависимость параметров кривой $v(E)$ от частоты становится заметной при частотах свыше 20 ГГц. При дальнейшем повышении частоты КПД начинает

снижаться и падает до нуля при частотах порядка 200 ГГц, хотя использование напряжения сложной формы может несколько расширить частотный диапазон режима ОНОЗ и повысить КПД. Так, использование одновременно первой и второй гармоник переменного поля позволяет повысить КПД до 31%.

В настоящее время существует несколько типов полупроводниковых приборов СВЧ: туннельные диоды, СВЧ-транзисторы (полевые и биполярные), лавинно-пролетные диоды, диоды Ганна.

Если первые три типа приборов основаны на явлениях в *p-n*-переходе (туннельный эффект через слой объемного заряда, транзисторный эффект или лавинное умножение), то в диодах Ганна используются явления в объеме однородного полупроводника.

У большинства полупроводниковых приборов с *p-n*-переходом существенным недостатком является сильная зависимость их параметров от качества изготовления и защиты *p-n*-перехода. Это приводит к сравнительно низкой воспроизводимости и стабильности параметров. Активной частью подобных приборов является очень узкая область слоя объемного заряда *p-n*-перехода, поэтому создание таких приборов на большие мощности вызывает затруднения.

Использование в диодах Ганна для усиления и генерации всего объема полупроводника упрощает проблему рассеяния теплоты выделяющейся при работе, тем самым позволяя значительно увеличить выходную мощность.

Предполагается, что благодаря улучшению технологии и конструкции в будущем мощность лавинно-пролетных и диодов Ганна будет в непрерывном режиме примерно одинаковой и достигнет у серийных генераторов 10 Вт. Наиболее мощными генераторами на твердом теле являются в настоящее время генераторы, работающие в импульсном режиме ОНОЗ. У экспериментальных образцов их мощность доходит до 6 кВт.

§ 9.11. Резонансные свойства полупроводникового материала

По мере увеличения частоты электромагнитных колебаний и постепенного перехода от диапазона сверхвысоких частот к инфракрасному, а затем и к оптическому диапазону все больше начинает сказываться дисперсия параметров полупроводникового материала, т. е. зависимость его электромагнитных свойств от частоты. Зависимость диэлектрической проницаемости полупроводникового материала от частоты в классической физике получена при рассмотрении упрощенной модели электронов, квазиупруго связанных с положениями равновесия и реагирующих на изменения поля подобно линейным гармоническим осцилляторам.

Поляризация возникает из-за смещения электронов из положения равновесия ($x = 0$) под действием поля электромагнитной волны. Действующая на *связанные электроны* (внутренних оболочек) со стороны электрического поля сила уравновешивается тремя основными силами: инерции, трения и упругости. Первая пропорциональна ускорению, вторая — скорости, третья — смещению.

Уравнение движения связанного электрона приобретает вид

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + mg_0\frac{\mathrm{d}x}{\mathrm{d}t} + m\omega_0^2 x = eE, \qquad (9.46)$$

где g_0, ω_0 — коэффициенты, характеризующие трение (например, радиационное) и упругость.

Для гармонической волны $E = E_m \sin\omega t$ решение уравнения в установившемся режиме позволяет получить комплексное значение амплитуды колебаний

$$\dot{X}_m = \frac{e}{m}\frac{1}{\omega_0^2 - \omega^2 + j\omega g_0}\dot{E}_m.$$

Тогда индуцированный электрический момент осциллятора $\dot{P}_э = e\dot{X}$, поляризация $\dot{\mathbf{P}} = n_0\dot{\mathbf{P}}_э$ и диэлектрическая восприимчивость

$$\underline{\chi}_0 = \frac{\dot{P}}{\dot{E}} = \frac{n_0 e^2}{m}\frac{1}{\omega_0^2 - \omega^2 + j\omega g_0}, \qquad (9.47)$$

где n_0 — число связанных электронов. Для кремния с $n_0 = 5{,}2\cdot10^{22}$ см$^{-3}$, $\omega_0 = 6\cdot10^{14}$, $\lambda_0 = 0{,}5$ мкм.

При рассмотрении компоненты поляризации, обусловленной *свободными электронами* в предыдущих уравнениях следует положить $\omega_0 = 0$, так как на эти электроны не действует удерживающая сила. Для них, согласно (9.47),

$$\underline{\chi}_n = \frac{ne^2}{m_n}\frac{1}{\omega - \omega + jg_n}.$$

Если рассматривать движение свободных электронов как движение заряженных частиц в вязкой среде, то в случае постоянного поля $E(t) = \mathrm{const}$, электроны будут двигаться без ускорения $(\mathrm{d}^2x/\mathrm{d}t^2) = 0$ с постоянной скоростью $\overline{v}_n = \mathrm{d}x/\mathrm{d}t = \mathrm{const}$. При этом

$$m_n g_n \frac{\mathrm{d}x}{\mathrm{d}t} = eE,$$

откуда с учетом (9.2) $g_n = eE/(m_n v_n) = \nu_n$.

Таким образом, для свободных электронов коэффициент вязкости g_n есть не что иное, как газокинетическая частота столкновений ν_n. Поэтому

$$\underline{\chi}_n = \frac{ne^2}{m_n\omega - \omega + j\nu_n}.$$

Аналогично для дырок

$$\underline{\chi}_p = \frac{pe^2}{m_p\omega - \omega + j\nu_p}.$$

Для частот, при которых поляризации от связанных и свободных зарядов соизмеримы, комплексная диэлектрическая проницаемость

$$\underline{\varepsilon} = \varepsilon_0 + \underline{\chi}_0 + \underline{\chi}_n + \underline{\chi}_p,$$

или

$$\underline{\varepsilon} = \underline{\varepsilon}_1 - \frac{ne^2}{m_n}\frac{1 + j\nu_n/\omega}{\omega^2 + \nu_n^2} - \frac{pe^2}{m_p}\frac{1 + j\nu_p/\omega}{\omega^2 + \nu_p^2}, \qquad (9.48)$$

где $\underline{\varepsilon}_1 = \varepsilon_0 + \underline{\chi}_0$ — диэлектрическая проницаемость без учета свободных носителей заряда.

В классической физике для системы из N осцилляторов с разными концентрациями n_s, собственными частотами ω_s, ν_s и эффективными массами m_s в общем случае комплексная проницаемость

$$\underline{\varepsilon} = \varepsilon_0 + \sum_{s=1}^{N}\underline{\chi}_s = \varepsilon_0 + \sum_{s=1}^{N}\frac{n_s e^2/m_s}{\omega_s^2 - \omega^2 + j\omega\nu_s}. \qquad (9.49)$$

Формула дисперсии (9.49) позволяет хорошо предсказать форму спектральной линии (рис. 9.26).

Однако из классических представлений нельзя правильно определить ни интенсивность этой линии, ни резонансную частоту, поскольку явления в атоме и взаимодействие элементарных частиц с электромагнитным полем требуют квантовомеханического подхода.

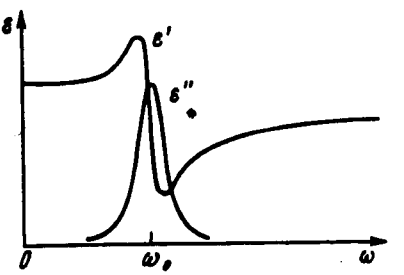


Рис. 9.26. Дисперсия диэлектрической проницаемости

Согласно квантовомеханической теории, взаимодействие между полем и поглощающим атомом можно описать с помощью набора линейных осцилляторов, каждый из которых имеет резонансную частоту,

соответствующую разрешенному переходу между двумя энергетическими состояниями $\omega_s = \omega_{ij} = (\mathscr{E}_i - \mathscr{E}_j)/\hbar$. При этом число классических электронов дисперсии n_s заменяется разностью концентраций осцилляторов в соответствующих состояниях: $n_s = N_i - N_j$. Влияние осциллятора на электродинамические параметры вещества для каждого перехода определяется *силой осциллятора*

$$f_s = f_{ij} = \frac{1}{m\hbar\omega_{ij}}\left|\mu_{ij}\right|^2 = \tau A_{ij},$$

где μ_{ij} — матричный элемент, характеризующий вероятность перехода из состояния i в состояние j; τ — среднее время жизни носителя заряда в состоянии j до возвращения в состояние i; A_{ij} — вероятность данного перехода.

Теперь суммирование в уравнении (9.49) должно производиться по всем N разрешенным переходам:

$$\underline{\varepsilon} = \varepsilon_0 + \sum_{s=1}^{N} \frac{f_s n_s e^2/m_s}{\omega_s^2 - \omega^2 + j\omega g_s}.$$

Благодаря большому сходству результатов квантовомеханического рассмотрения и простого классического решения последнее дает качественное согласие с результатами экспериментов.

§ 9.12. Преломление и поглощение электромагнитных волн в полупроводнике

Распространение плоской волны в веществе вдоль оси x сопровождается изменением напряженности поля $\dot{\mathbf{E}} = \dot{\mathbf{E}}_0 e^{-\gamma x} = \dot{\mathbf{E}}_0 e^{-\alpha x} e^{-j\beta x}$, где $\gamma = j\omega\sqrt{\underline{\varepsilon}\underline{\mu}} = \alpha + j\beta = j\underline{k}$ — коэффициент распространения; α — коэффициент ослабления.

Интенсивность электромагнитного излучения соответствует плотности потока мощности. Она пропорциональна квадрату напряженности:

$$I(x) = P \doteq E^2 = \dot{\mathbf{E}}\dot{\mathbf{E}}^* = E_0^2 e^{-2\alpha x} \doteq I(0) e^{-Kx},$$

где $K = 2\alpha$ — показатель поглощения. Приведенное выражение известно как закон *Бугера—Ламберта.*

Преломляющую способность плоской поверхности можно характеризовать комплексным показателем преломления $\underline{n} = \sqrt{\underline{\varepsilon}_r\underline{\mu}_r} = n' - jn''$. В

полупроводниках $\underline{\mu}_r = 1$. Поэтому в дальнейшем будем полагать, что $\underline{n}^2 = n'^2 - n''^2 - 2jn'n'' = \underline{\varepsilon}_r = \varepsilon'_r - j\varepsilon''_r$. Следовательно,

$$2n'n'' = \varepsilon''_r, \quad n'^2 - n''^2 = \varepsilon'_r. \tag{9.50}$$

В часто встречающемся случае, когда $n'' \ll n'$, можно считать, что $\underline{n}^2 \approx n'^2 = \varepsilon'_r$. Комплексное волновое число

$$\underline{k} = \omega\sqrt{\underline{\varepsilon}\underline{\mu}} = \frac{\omega}{c_0}\sqrt{\underline{\varepsilon}_r\underline{\mu}_r} = \frac{\varepsilon}{c_0}(n' - jn'') = \beta - j\alpha, \tag{9.51}$$

откуда коэффициент фазы

$$\beta = \omega n'/c_0, \tag{9.52}$$

показатель поглощения

$$K = 2\alpha = 2\omega n''/c_0 = \omega\varepsilon''_r/(c_0 n'). \tag{9.53}$$

В полупроводниках поглощение связано с взаимодействием поля волны с основными четырьмя типами зарядов: 1) электронами внутренних оболочек атома; 2) свободными носителями; 3) электронами на локальных примесных центрах или дефектах; 4) электронами валентной зоны.

Поглощение электронами внутренних оболочек может играть заметную роль при тех частотах, которые примыкают к низкочастотной границе внутреннего фотоэффекта (в дальней инфракрасной области спектра). Если спектр этого поглощения состоит только из одной линии на резонансной частоте ω_0, то для нее согласно (9.47),

$$\underline{\varepsilon}_1 = \varepsilon_0 + \underline{\chi}_0 = \varepsilon_0 + \frac{n_0 e^2/m}{\omega_0^2 - \omega^2 + j\omega g_0}.$$

Наибольшее поглощение происходит при резонансной частоте $\omega = \omega_0$, когда $\underline{\varepsilon}_1 = 1 - jn_0e^2/(m\omega g_0)$ и $K_{1\max} = e^2 n_0/(mg_0 c_0 n'\varepsilon_0)$.

Смещение связанных электронов и поляризация атомов под действием электромагнитного поля означает в конечном счете взаимодействие с диполями, составляющими кристаллическую решетку. Поглощаемая при этом энергия электромагнитного поля передается энергии колебаний кристаллической решетки. Поэтому рассматриваемый вид поглощения называют *решеточным*. Наиболее сильное поглощение электромагнитного излучения происходит при частотах, равных частотам собственных колебаний диполей и соответствующих резонансным частотам ω_s. Согласно закону сохранения импульса происходит поглощение только таких фотонов, импульс которых равен квазиимпульсу фононов. Малость квазиимпульса фонона по сравнению с импульсом фотона приводит к

тому, что при поглощении фотона испускается несколько фононов. Поэтому спектр поглощения в далекой инфракрасной области характеризуется возникновением рядом пиков сложной структуры.

Приведенное выражение для коэффициента поглощения не учитывает взаимодействия света с кристаллом, сопровождаемого переходами электронов внутренних оболочек в свободное состояние зоны проводимости. Эти процессы происходят в области далекого ультрафиолетового и рентгеновского диапазонов, когда энергия фотонов намного превышают ширину запретной зоны. Коэффициент поглощения, связанный с этими переходами, велик (порядка 10^5 см$^{-1}$), однако он соответствует небольшой области длин волн и эти переходы не вносят большого вклада в длинноволновый показатель преломления.

Вдали от края основной полосы поглощения K_1 незначителен ($\underline{\chi}_0 \approx \chi_0$) и определяющую роль играет поглощение свободными носителями заряда. В этом случае согласно (9.48)

$$\underline{\varepsilon}_2 = \varepsilon_0 + \chi_0 + \underline{\chi}_n + \underline{\chi}_p = \varepsilon_1 - \frac{ne^2}{m_n\omega}\frac{\omega + j\nu_n}{\omega^2 + \nu_n^2} - \frac{pe^2}{m_p\omega}\frac{\omega + j\nu_p}{\omega^2 + \nu_p^2}.$$

При достаточно больших частотах ($\omega \gg \nu$) поглощение, согласно (9.48) и (9.53), пропорционально квадрату длины волны:

$$K_2 = \frac{ne^2\nu_n}{\varepsilon_0 m_n c_0 n'\omega^2} + \frac{pe^2\nu_p}{\varepsilon_0 m_p c_0 n'\omega^2} \sim \frac{1}{\omega^2} \sim \lambda^2. \qquad (9.54а)$$

При $\omega \ll \nu$

$$K_2 = \frac{ne^2}{\varepsilon_0 m_n c_0 n'\nu_n} + \frac{pe^2}{\varepsilon_0 m_p c_0 n'\nu_p}, \qquad (9.54б)$$

т. е. поглощение перестает зависеть от длины волны и частоты.

Поглощение фотонов свободными носителями заряда сопровождается их переходом в состояние с большей энергией в пределах той же подзоны проводимости с изменением квазиимпульса. Это обеспечивается взаимодействием фотонов и зарядов с фононами (кристаллической решеткой) либо рассеянием на ионизированных примесях. Учет рассеяния приводит к более сложной частотной зависимости затухания. Так, при рассеянии акустическими фононами поглощение меняется по закону $\lambda^{1,5}$, при рассеянии оптическими фононами — по закону $\lambda^{2,5}$, а при рассеянии ионизированными примесями — $\lambda^{3,5}$. В общем случае $K_2 = A\lambda^{1,5} + B\lambda^{2,5} + C\lambda^{3,5} = D\lambda^p$. В зависимости от концентрации примесей соотношение между константами A, B, C и D меняется, а показатель p возрастает по мере увеличения легирования или степени компенсации.

Затухание электромагнитной волны в примесном полупроводнике сопровождается ионизацией или возбуждением атомов примеси при поглощении кванта соответствующей энергии. Поглощение на примесях и дефектах может быть вычислено, если известно сечение поглощения $\varkappa$: $K_3 = \varkappa N$, где N — плотность поглощающих центров. Для типичных полупроводников $\varkappa \approx 10^{16}$ см^2.

При изучении оптических свойств полупроводников наиболее важную роль играет поглощение электронами валентной зоны, сопровождаемое поглощением кванта с энергией, достаточной для перехода электрона через запрещенную зону в незаполненную зону проводимости. Такое поглощение называют *собственным* или *фундаментальным*. Спектр собственного поглощения при малых длинах волн изображают непрерывной кривой, которая более или менее круто спадает в области больших длин волн; при еще больших длинах волн материал относительно прозрачен.

Переходы из валентной зоны в зону проводимости делятся на прямые и непрямые в зависимости от того, меняется ли при этом волновой вектор $\mathbf{k}$ электрона. Межзонные переходы, при которых волновое число не меняется, называют прямыми. Переходы с изменением импульса называют непрямыми. Вид перехода зависит от формы границ энергетических зон $\mathscr{E}(\mathbf{k})$.

Пусть положения минимума зоны проводимости $\mathscr{E}_c$ и максимума валентной зоны $\mathscr{E}_v$ совпадают ($k_c = k_v$). Тогда край полосы поглощения определяется условием $\hbar\omega_0 = \mathscr{E}_c - \mathscr{E}_v = \mathscr{E}_g$, если при переходе выполняется условие четности начальной и конечной волновых функций. Согласно квантовомеханическому правилу отбора, волновой вектор меняется на «квант поглощения» $\mathbf{q} = \mathbf{k}_j - \mathbf{k}_i$, где $\mathbf{k}_i$, $\mathbf{k}_j$ — волновые векторы электрона в начальном и конечном состояниях. Волновое число фотона $q = 2\pi/\lambda$ с длиной волны порядка 1 мкм мало по сравнению с волновым числом электрона $k \approx 2\pi/a$ (постоянная решетки a составляет несколько ангстрем). В этом случае правило отбора можно упростить: $\mathbf{k}_j = \mathbf{k}_i$. Следовательно, при всех межзонных переходах, связанных с поглощением или излучением квантов рассматриваемого диапазона, волновое число должно сохраняться практически неизменным. Таким образом, в данном случае разрешаются только вертикальные переходы. Расчеты показывают, что если переход разрешен при $\mathbf{k} = 0$, коэффициент поглощения

$$K_4^{(d)} \approx \frac{e^2}{c_0 n' \hbar^2 m_n} \left(2 \frac{m_p m_n}{m_p + m_n} \right)^{3/2} (\hbar\omega - \mathscr{E}_g)^{1/2}.$$

Если переход при $k = 0$ запрещен, то его вероятность становится намного ниже. При этом

$$K_4^{(d)} = \frac{4e^2}{3c_0 n' m_n m_p h^3} \left(\frac{m_p m_n}{m_p + m_n} \right)^{5/2} \frac{(h\nu - \mathscr{E}_g)^{3/2}}{h\nu}.$$

Запрет непрямых переходов не означает, что такие переходы совсем не происходят. Практически всегда существуют причины, смягчающие правила отбора, в результате чего такие переходы оказываются допустимыми, хотя их вероятность меньше вероятности прямых или вертикальных переходов. Сохранение импульса электрона при непрямом (невертикальном) переходе может быть объяснено взаимодействием с фононом.

При поглощении света возможно такое возбуждение электрона валентной зоны, при котором он не переходит в зону проводимости, а образует с дыркой систему, связанную кулоновским притяжением. Такую систему называют экситоном.

Область поглощения, связанная с прямыми переходами электронов на экситонные уровни, находится вблизи длинноволновой границы спектра основного поглощения, где $(\hbar\omega - \mathscr{E}_g) < 0$ с резким пиком поглощения, который расширяется с повышением температуры.

В заключение необходимо упомянуть о явлении *двухфотонного поглощения*, возникающего при больших плотностях фотонных потоков. При этом две интенсивные электромагнитные волны с частотами ω_1 и ω_2 обеспечивают переход на частоте $\omega_0 = \omega_1 + \omega_2$, соответствующей в полупроводнике переходу между валентной зоной и зоной проводимости. Подобные переходы осуществляются благодаря тому, что согласно квантовой теории имеется конечная вероятность того, что часть энергии обеих волн будет поглощена вследствие переходов на суммарной частоте ω_0.

§ 9.13. Полупроводниковые оптические модуляторы

С появлением квантовых генераторов, вырабатывающих узкие когерентные пучки электромагнитной энергии в оптическом диапазоне, возникла реальная возможность создания систем передачи и обработки информации, работающих на длинах волн порядка нескольких микрометров, т. е. на частотах 10^{14} Гц и выше. Поскольку модуляция излучения внутри таких источников связана часто со значительными трудностями, то в системах связи возникла необходимость изыскания эффективных методов модуляции луча на выходе источника с частотой поднесущей порядка 10^{10} Гц. Именно эти причины привлекли внимание к возможности использования для этой цели полупроводниковых материалов. Не-

смотря на то, что полупроводниковые системы в видимом диапазоне пока менее эффективны, чем электрооптические и ультразвуковые, знакомство с принципами их работы представляет интерес в связи с возможностью их применения в смежных диапазонах, а также тенденцией использования оптического диапазона в интегральных схемах и все большем распространении интегральной оптики.

Модуляция излучения путем изменения концентрации свободных носителей заряда. Если световая волна с начальной интенсивностью I_0 распространяется в однородном полупроводнике, то после прохождения слоя толщиной d ее интенсивность уменьшится до $I = I_0 e^{-Kd}$. Следовательно, пропускание данного слоя $A = I/I_0 = e^{-Kd}$, а поглощение $B = I_0/I = e^{Kd}$. В том случае, когда поглощение обусловлено только свободными носителями заряда, показатель поглощения (9.54) оказывается прямо пропорциональным концентрациям свободных носителей n, p. Поэтому, изменяя их концентрацию, можно получить соответствующую модуляцию проходящего излучения. Десятикратное изменение концентрации приводит к достаточно глубокой модуляции, если при этом Kd будет изменяться от 0,2 (A = 0,82) до 2,0 (A = 0,14).

Модуляция концентрации свободных носителей заряда обычно осуществляется с помощью p-n-перехода, смещенного в прямом направлении (рис. 9.27). Если в отсутствие тока через переход пропускание $A_0 = e^{-K_0 d} = I_1/I_0$, а при прямом смещении $A = e^{-Kd} = e^{-(K_0+\Delta K)d} = A_0 e^{-\Delta K d} = I_2/I_0$, то глубина модуляции

$$m = \frac{I_1 - I_2}{I_1} = \frac{A_0 - A}{A_0} = 1 - e^{-\Delta K d}$$

и соответствующий показатель поглощения неосновных неравновесных дырок с учетом (9.54а)

$$\Delta K = -\frac{1}{d}\ln(1-m) = \frac{e^2 \nu_p \Delta p}{c_0 \varepsilon_0 n' m_p \omega^2} = \alpha_s \Delta p, \quad (9.55)$$

где $\alpha_s = \nu_p e^2/(c_0\varepsilon_0 n' m_p \omega^2)$ — сечение захвата введенного центра поглощения.

Поскольку концентрация неравновесных дырок $\Delta p(x) = \Delta p(0) e^{-x/L_p}$, а диффузионный ток

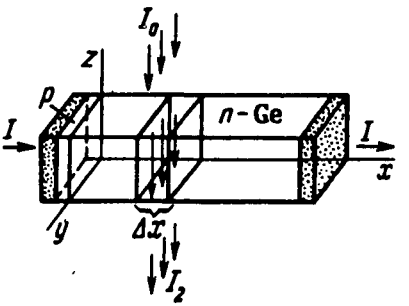


Рис. 9.27. Модулятор инфракрасного излучения с p-n-переходом

$$I_p = -eSD_p\left(\frac{\mathrm{d}\Delta p}{\mathrm{d}x}\right)_{x=0} = eSD_p\,\Delta p(0)/L_p = \gamma_p I,$$

то $\Delta p(x) = [\gamma_p/(ev_pS)]I\,\mathrm{e}^{-ax}$, где γ_p — коэффициент инжекции дырок; I — ток в p-n-переходе; a — постоянная затухания концентрации носителей заряда (равная в этом случае величине, обратной диффузионному пробегу); v_p — скорость носителей заряда (в случае диффузионного тока $v_p = D_p/L_p$); S — площадь p-n-перехода.

Таким образом, связь между глубиной модуляции электромагнитного излучения и током инжекции может быть представлена в виде

$$\ln(1-m) = -\frac{\gamma_p \alpha_s I d}{ev_p S}\mathrm{e}^{-ax}. \qquad (9.56)$$

Линейная зависимость модуляции (9.56) от плотности тока справедлива только при сравнительно небольших плотностях тока, пока током проводимости можно пренебречь. При больших плотностях тока общее напряжение на диоде $U \approx El$ оказывается приложенным в основном к области базы (длиной l) и глубина проникновения определяется путем, проходимым дыркой за время жизни τ_p, т. е. $a = 1/(\tau_p \mu_p E)$, а $v_p = \mu_p E$. Тогда глубина модуляции

$$\ln(1-m) \approx -\frac{\alpha_s \gamma_p l_d}{e\mu_p S} G\,\mathrm{e}^{-x/(\tau_p \mu_p E)},$$

т. е. пропорциональна проводимости базы G.

Экспериментальная зависимость глубины модуляции от плотности тока в германиевом диоде приведена на рис. 9.28, *а*. Модуляционная ха-

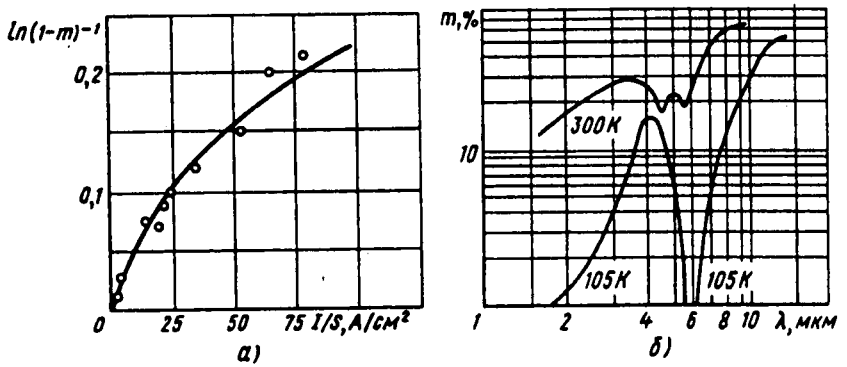


Рис. 9.28. Модуляционная и спектральные характеристики германиевого модулятора

рактеристика при больших уровнях инжекции стремится к насыщению вследствие уменьшения коэффициента инжекции, подвижности и времени жизни носителей заряда при больших плотностях тока. Максимальная амплитуда выходного сигнала ($I_1 - I_2$), согласно (9.55), получается при толщине модулятора $d_{опт} = \Delta K^{-1} \ln(1 + \Delta K / K_0)$.

В качестве модуляторов инфракрасного излучения с использованием *p-n*-переходов применяются германиевые переходы, поскольку в других полупроводниках диффузионные длины неосновных носителей заряда значительно меньше. Это не позволяет получить достаточно большую область Δx, в которой происходила бы эффективная модуляция. В германии эффективная модуляция реализуется для излучения с длинами волн выше 1,8 мкм. На рис. 9.28, *б* приведена спектральная характеристика модулятора при двух температурах, а на рис. 9.29 — одна из конструкций германиевого модулятора со сплавным *p-n*-переходом.

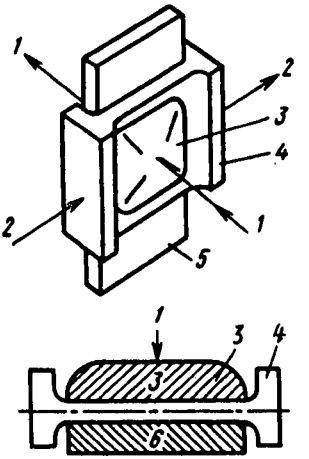


Рис. 9.29. Оптический модулятор:

1 — направление тока; *2* — направление излучения; *3* — индиевый электрод; *4* — германиевая пластинка; *5* — коваровая пластинка; *6* — оловянный электрод

Модуляция излучения путем изменения подвижности носителей заряда. Учитывая соотношение между частотой столкновений и подвижностью (9.2), уравнение (9.54а) можно переписать в виде

$$K_2 = \frac{ne^3}{c_0 n' \varepsilon_0 m_n^2 \mu_n \omega^2} + \frac{pe^3}{c_0 n' \varepsilon_0 m_p^2 \mu_p \omega^2}. \qquad (9.57а)$$

Аналогично перепишем уравнение (9.54б):

$$K_2 = \frac{ne\mu_n}{c_0 n' \varepsilon_0} + \frac{pe\mu_p}{c_0 n' \varepsilon_0}. \qquad (9.57б)$$

Таким образом, в области высоких частот показатель поглощения обратно пропорционален подвижности носителей заряда, а в области относительно низких частот прямо пропорционален ей. Поэтому, меняя подвижность, можно осуществлять модуляцию оптического пропускания полупроводника.

Управление подвижностью удобнее всего производить путем применения сильного электрического поля в той области, где происходит нарушение закона Ома ($\mu = \text{const}$).

Носителям зарядов в дополнение к их энергии теплового равновесия сообщается приложенным полем некоторая избыточная энергия. Время, затраченное на рассеяние этой избыточной энергии при соударениях носителей, определяет постоянную времени процесса. Например, для германия *n*-типа при комнатной температуре время релаксации составляет $3 \cdot 10^{-13}$ с, постоянная времени такого модулятора должна быть примерно 10^{-11} с.

Полупроводниковые модуляторы, в которых используется изменение подвижности носителей заряда, были применены для модуляции волн СВЧ.

Внутризонное поглощение. Оптическое поглощение может возрастать за счет перехода носителей заряда из одной подзоны в другую. Например, в германии *p*-типа полная валентная зона состоит из подзоны тяжелых дырок *1*, подзоны легких дырок *2* и подзоны *3*, отделенной от первых двух спин-орбитальным промежутком шириной 0,29 эВ.

Поглощение в интервале 3...4 мкм обусловлено переходами носителей заряда из подзоны *3* в подзоны *1* и *2*. Величина поглощения определяется функцией распределения свободных дырок, т. е. зависит от температуры. Функцию распределения можно изменять при воздействии на полупроводниковый материал электрического поля. Сильные поля увеличивают эффективную температуру полупроводника и вызывают появление в нем «горячих» носителей.

Постоянная времени данного метода модуляции (как и в случае модуляции посредством изменения подвижности) определяется временем спада энергии носителей заряда. Соответствующие измерения фотопроводимости германия *p*-типа при температуре жидкого воздуха показали, что эта величина колеблется от $5 \cdot 10^{-12}$ до $15 \cdot 10^{-12}$ с.

Сдвиг спектральной границы поглощения. В некоторых случаях поглощение на валентных электронах может определять не только ширина запретной зоны, но и характер заполнения энергетических состояний равновесными электронами и дырками. Например, в случае вырожденного полупроводника с электронной электропроводностью равновесные электроны в зоне проводимости занимают почти все состояния в некоторой области, примыкающей к дну этой зоны. Поэтому энергия кванта $\hbar\omega_0$, необходимая для перевода электрона из валентной зоны в зону проводимости, должна превышать ширину запретной зоны $\mathscr{E}_g$ на эффективную ширину $\Delta\mathscr{E}$ участка, занятого электронами вблизи дна зоны проводимости: $\hbar\omega_0 = \mathscr{E}_g + \Delta\mathscr{E}$. Ширину этой области легко определить для материалов со сферической изоэнергетической поверхностью при плотности квантовых состояний

$$N(\mathscr{E}) = \frac{4\pi}{h^3}(2m_n)^{3/2}(\mathscr{E} - \mathscr{E}_c)^{1/2}$$

и низких температурах ($T \approx 0$ K), когда функцию распределения

$$f_n(\mathscr{E}, T) = \left(\exp\frac{\mathscr{E} - \mathscr{E}_F}{k_B T} + 1\right)^{-1} \qquad (9.58)$$

можно принять равной единице. Тогда концентрация свободных электронов

$$n = \int_{\mathscr{E}_c}^{\mathscr{E}} f \, \mathrm{d}Z = \int_{\mathscr{E}_c}^{\mathscr{E}_F} \frac{4\pi}{h^3}(2m_n)^{3/2}(\mathscr{E} - \mathscr{E}_c)^{1/2} \, \mathrm{d}\mathscr{E} = \frac{8\pi}{3h^3}(2m_n)^{3/2}(\mathscr{E}_F - \mathscr{E}_c)^{3/2}. \quad (9.59)$$

В данном случае

$$\Delta\mathscr{E} = \mathscr{E}_F - \mathscr{E}_c = (3h^3/8\pi)^{2/3} n^{2/3}/(2m_n),$$

$$\hbar\omega_0 = \mathscr{E}_g + \Delta\mathscr{E} = \hbar\omega_{\min} + (3/\pi)^{2/3}\hbar^2 n^{2/3}/(8m_n)$$

и изменение пороговой частоты

$$\Delta\omega = \omega_0 - \omega_{\min} = \pi(3/\pi)^{2/3}\hbar^2 n^{2/3}/(4m_n)$$

непосредственно зависит от концентрации и эффективной массы электронов. Таким образом, длинноволновую границу поглощения на валентных электронах можно смещать, изменяя концентрацию равновесных электронов. Впервые это явление было обнаружено в антимониде индия (InSb), в котором эффективная масса электрона намного меньше, чем у дырок ($m_n \approx 0{,}03m_0$, $m_p \approx 0{,}20m_0$). Поэтому сравнительно небольшое изменение концентрации приводит к заметному изменению спектральной границы поглощения.

При комнатной температуре явление усложняется, так как происходит «размытие» верхнего уровня заполненной электронами части зоны проводимости. Вследствие этого становятся возможными переходы и при $\omega < \omega_0$, однако их вероятность быстро уменьшается по мере удаления частоты ω от ω_0. Внешнее электрическое поле, повышая эффективную температуру свободных электронов, увеличивает «размытие» функции распределения (9.58) и повышает вероятность поглощения квантов с энергией $\hbar\omega < \hbar\omega_0$. Таким образом, при воздействии на полупроводник электрического поля происходит смещение края полосы поглощения материала в сторону более длинных волн.

Зависимость эффективности данного метода модуляции от длины волны излучения характеризует кривые, приведенные на рис. 9.30. На длине волны, соответствующей наибольшему наклону кривой пропускания, наблюдается максимальная модуляция.

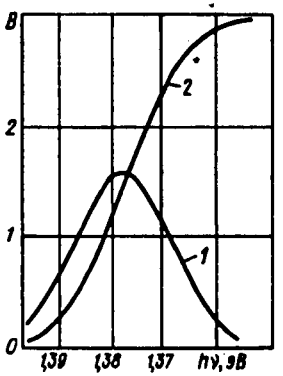


Рис. 9.30. Относительное поглощение в арсениде галлия:

1 — в отсутствии поля; *2* — при воздействии переменного поля напряженностью 4,5 кВ/см

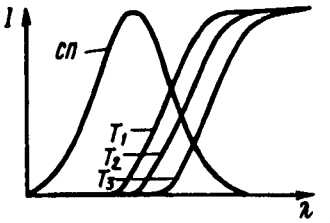


Рис. 9.31. Смещение края полосы поглощения при изменении температуры *T*:

СП — спектр источника

Смещение края полосы поглощения света в зависимости температуры может быть также использовано для измерения температуры в сенсорном световолоконном устройстве, в котором полупроводниковый преобразователь в виде кристалла арсенида галлия или теллурида кадмия с помощью световодов присоединяется к источнику и приемнику света. С повышением температуры край полосы поглощения смещается в сторону более длинных волн (рис. 9.31), вызывая уменьшение общей интенсивности света *I*, проходящего через полупроводник.

Эффект Келдыша—Франца. Этот эффект заключается в сдвиге края полосы поглощения в результате наклона границ энергетических зон в сильном электрическом поле *E*. Наклон границ создает треугольный барьер толщиной $d' = \mathscr{E}_g/(eE) = (\mathscr{E}_c - \mathscr{E}_v)/(eE)$ для туннельного перехода электронов из валентной зоны в зону проводимости (при $\mathscr{E} = \text{const}$). Поглощение электроном валентной зоны фотона с энергией $h\nu < \mathscr{E}_g$ приводит для этого электрона к уменьшению толщины барьера до $d = (\mathscr{E}_g - h\nu)/(eE)$, и туннельный переход становится более вероятным. Благодаря этому происходит сдвиг границы полосы поглощения в сторону меньших энергий.

В модуляторе, использованном в оптроне при напряжении 100 В коэффициент модуляции составлял 26%. При переходе к тонкопленочным конструкциям достаточные для модуляции поля могут быть получены при напряжениях 1...10 В. Характерная особенность эффекта Келдыша—Франца заключается в его практической безынерционности (время релаксации $10^{-13}...10^{-14}$ с).

Пространственно-временно́й модулятор света на основе эффекта Келдыша—Франца реализован в линейке из 16 элементов структуры приборов с зарядовой связью (рис. 9.32). На поверхности структуры создается периодическая последовательность электродов, образующих с полупроводником *4* барьер Шотки. Каждый период последовательности включает два непрозрачных электрода *2* и один полупрозрачный *3*. В исходном состоянии под действием приложенного напряжения область под полупрозрачным электродом обеднена электродами и электрическое поле спадает в глубину полупроводника. В этой области сильное поглощение и интенсивность светового пучка *6* на выходе из подложки *5* (из арсенида галлия) минимальна.

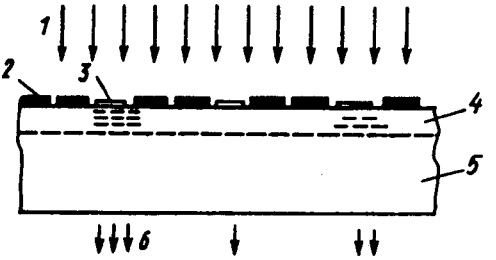


Рис. 9.32. Модулятор на основе эффекта Келдыша—Франца

При распространении по полупроводниковому слою последовательности зарядовых пакетов поле под полупрозрачными электродами может быть компенсировано, что приведет к уменьшению поглощения светового потока *1*. Таким образом, оптический импульс оказывается способным зафиксировать мгновенную картину пропускания света структурой. При изменении потенциала в области прозрачного электрода на 30 В пропускание увеличивалось в 2,7 раза. Тактовая частота посылки электрических сигналов составляла 1,6 МГц.

Набор подобных линеек позволил создать матричный модулятор, в котором адресация осуществляется последовательностями электрических сигналов. Расчет показывает возможность создания матрицы размером 1000×1000 элементов.

Рассмотренные методы модуляции электромагнитного излучения основаны на управлении поглощением полупроводника. Они не исчерпывают всех способов модуляции. В частности, имеется ряд материалов, оптические свойства которых могут изменяться (модулироваться) непосредственно под воздействием светового облучения. К их числу относят пленки халькогенидных стеклообразных полупроводников. Под воздействием света в них происходит изменение как показателя поглощения, так и показателя преломления, что дает возможность использовать их в качестве пространственного модулятора света и, в частности, записывать амплитудную и фазовую составляющие интерференционной двухмерной картины, т. е. амплитудно-фазовые голограммы. При этом запись и считывание можно производить световыми потоками разных спектральных диапазонов. Например, в пленках As_2S_3 и $As_{10}S_{90}$ запись голограммы про-

изводилась синим светом, а считывание — красным. Однако наибольшие дифракционные эффективности получались при времени записи порядка минут. Для оптической модуляции в полупроводниках могут быть использованы и магнитооптические явления.

§ 9.14. Магнитооптические явления

Прохождение плоскополяризованного излучения через диэлектрик, находящийся в продольном магнитном поле, сопровождается поворотом плоскости поляризации. Как и в случае ферритов, гиротропные свойства диэлектриков объясняются различием скоростей распространения волн, поляризованных по кругу в противоположных направлениях.

Эффект Фарадея в полупроводниках в зависимости от частоты электромагнитных колебаний может быть обусловлен поляризацией как связанных, так и свободных электронов. Дальнейшее изложение ограничено последним случаем, специфическим для полупроводников и привлекшим внимание в связи с обсуждением возможностей использования данного явления для модуляции электромагнитного излучения.

Эффект Фарадея на свободных носителях заряда. При взаимодействии электрического поля со свободными носителями, находящимися в постоянном магнитном поле, движение электрона будет определяться силой Лоренца и в классическом приближении

$$m_n \frac{\partial^2 \mathbf{R}}{\partial t^2} + m_n \nu_n \frac{\partial \mathbf{R}}{\partial t} = e\mathbf{E} + e\frac{\partial \mathbf{R}}{\partial t} \times \mathbf{B}.$$

Если совместить вектор индукции с осью z и считать, что вектор напряженности электрического поля плоской волны параллелен плоскости xy, то

$$\frac{\partial^2 x}{\partial t^2} + \nu_n \frac{\partial x}{\partial t} = \frac{e}{m_n} E_x + \frac{e}{m_n} B \frac{\partial y}{\partial t},$$

$$\frac{\partial^2 y}{\partial t^2} + \nu_n \frac{\partial y}{\partial t} = \frac{e}{m_n} E_y - \frac{e}{m_n} B \frac{\partial x}{\partial t}.$$

После умножения второго уравнения на j и суммирования с первым

$$\frac{\partial}{\partial t^2}(x + jy) + \nu_n \frac{\partial}{\partial t}(x + jy) + j\frac{e}{m_n} B(x + jy) = \frac{e}{m_n}(E_x + jE_y).$$

Если обозначить комплексное смещение $\dot{s} = x + jy$ и учесть, что для поляризованных по кругу в разных направлениях волн

$$\dot{E}(t) = E_x(t) + jE_y(t) = \dot{E}\,e^{\pm j(\omega t - k_\pm x)}, \quad \dot{s}(t) = \dot{s}\,e^{\pm j(\omega t - k_\pm z)},$$

то из последнего уравнения

$$-\omega^2 \dot{s} \pm j\omega\nu_n \dot{s} \mp \frac{e}{m_n}\omega B\dot{s} = \frac{e}{m_n}\dot{E}$$

или

$$\dot{s} = \frac{e\dot{E}}{m_n\omega}\frac{1}{\omega \mp eB/m_n \mp j\nu_n}.$$

Диэлектрическая восприимчивость $\underline{\chi}_n = ne\dot{s}/\dot{E}$, откуда диэлектрическая проницаемость

$$\underline{\varepsilon} = \varepsilon_0 + \chi_0 + \underline{\chi}_n = \varepsilon_1 - \frac{ne^2/(m_n\omega)}{\omega \mp eB/m_n \mp j\nu_n} = \varepsilon' - j\varepsilon''.$$

Для немагнитных диэлектриков ($\underline{\mu}_r = 1$) показатель поглощения (9.53)

$$K = \frac{ne^2}{c_0 n' \varepsilon_0 m_n}\frac{\nu_n}{(\omega \mp eB/m_n)^2 + \nu_n^2}.$$

Согласно формуле (9.50),

$$n'^2 - n''^2 = \varepsilon_1 - \frac{e^2 n}{\varepsilon_0 m_n \omega}\frac{\omega \mp eB/m_n}{(\omega \mp eB/m_n)^2 + \nu_n^2}.$$

В случае малого поглощения ($n''^2 \ll n'^2$) и больших частот, когда $\omega^2 \gg \nu_n^2$,

$$n_\pm'^2 = \varepsilon_1 - \frac{e^2 n}{\varepsilon_0 m_n \omega}\frac{1}{\omega \mp eB/m_n}.$$

Разные знаки соответствуют разным направлениям круговой поляризации и дают два различных значения коэффициента преломления, причем в слабом магнитном поле $n' \approx \dfrac{n'_- + n'_+}{2}$ и

$$n'_- - n'_+ = \frac{n'^2_- - n'^2_+}{2n'} = \frac{ne^2}{\varepsilon_0 m_n n' \omega^3}\frac{e}{m}B.$$

Угол поворота плоскости поляризации, выраженный в радианах на единицу длины пути луча в данной среде, согласно уравнениям (6.32) и (9.52)

$$F = \frac{\beta_+ - \beta_-}{2} = \frac{\omega}{c_0}\frac{n'_- - n'_+}{2} = \frac{ne^3}{2\varepsilon_0 m_n^2 c_0 n_0 \omega^2}B.$$

Зависимость угла поворота плоскости поляризации от напряженности магнитного поля позволяет использовать эффект Фарадея в полупроводниковом оптическом модуляторе. Если перед полупроводниковым

кристаллом и после него установить под углом 45° скрещенные поляризатор и анализатор, то пропускание будет возрастать или уменьшаться в зависимости от величины и знака приложенного магнитного поля.

Поглощение на свободных электронах характеризуется, согласно (9.57), показателем

$$K = \frac{ne^3}{c_0 n' \varepsilon_0 m_n^2 \mu_n \omega^2},$$

поэтому

$$\frac{Fd}{Kd} = \frac{1}{2}\mu_n B.$$

Произведение Kd определяет пропускание $A = e^{-Kd}$. При пропускании в 37% угол поворота плоскости поляризации составит $0{,}5\mu_n B$ рад. Следовательно, для получения достаточно глубокой модуляции требуется выполнение условия $\mu_n B \geqslant 1$. Иными словами, если полупроводник обладает подвижностью носителей заряда порядка 1 м^2/(В·с), необходимо обеспечить уровень магнитной индукции B = 1 Тл. Весьма высокая подвижность носителей у антимонида индия [при комнатной температуре μ_n = 7 м^2/(В·с)] позволяет получить сравнительно большие углы вращения плоскости поляризации.

Ферромагнитные полупроводники обладают несравненно бóльшим спонтанным фарадеевским вращением. Например, в EuO оно достигает $8{,}5 \cdot 10^5$ град/см (при длине волны λ = 610 нм), т. е. превосходит максимальное фарадеевское вращение в металлах. Значительной величины оно достигает и у халькогениевых шпинелей, например, $CdCr_2Se_4$ (свыше $2 \cdot 10^5$ град/см при T = 80 К). Для этих материалов характерна исключительно высокая магнитооптическая добротность (отношение угла поворота к затуханию света). Наиболее высокая добротность достигается в окнах прозрачности , т. е. для длин волн, которые не могут вызвать межзонный переход при ширине запрещенной зоны, существенно большей энергии фотонов. Например, в EuO при длине волны 2,5 мкм получена добротность $1{,}4 \cdot 10^4$ град/дБ, тогда как в металлах она меньше 0,1 град/дБ.

Циклотронный резонанс. Если вдоль направления постоянного магнитного поля распространяется электромагнитная волна, то происходит поглощение ее энергии, причем величина поглощения достигает максимума, когда частота волны равна частоте (9.6). Явление поглощения электромагнитной волны на циклотронной частоте (9.6) называют циклотронным резонансом и частоту (9.6) называют иногда частотой циклотронного резонанса.

Согласно уравнению (9.6), частота циклотронного резонанса зависит только от магнитной индукции и эффективной массы. В связи с этим

наиболее непосредственным методом определения эффективной массы в настоящее время является измерение частоты циклотронного резонанса. Для свободного электрона при $B = 0{,}5$ Тл вычисленная величина $\omega_c \approx$ $\approx 9 \cdot 10^{10}$ рад/с соответствует частоте около 15 ГГц.

Методом циклотронного резонанса можно оценить и время свободного пробега путем измерения ширины резонансной линии, поскольку уменьшение показателя поглощения по сравнению с его максимальным значением (при $\omega = \omega_c$) в два раза происходит при отклонении частоты от резонансной на величину $\Delta\omega = \nu_n = 1/\tau_n$.

Ранее предполагалось, что направление распространения электромагнитной волны совпадает с направлением внешнего постоянного магнитного поля. Аналогично можно рассмотреть явления в полупроводнике при распространении электромагнитного излучения перпендикулярно внешнему полю. Если волна поляризована параллельно магнитному полю, то на параметры волны магнитное поле влияния не оказывает; в материале с электронной электропроводностью при $\omega^2 \gg \nu_n^2$

$$\varepsilon_{\parallel}'^2 = \varepsilon_1 - \frac{ne^2}{m_n}\frac{1}{\omega^2 + \nu_n^2} \approx \varepsilon_1\left(1 - \frac{\omega_p^2}{\omega^2}\right),$$

где $\omega_p^2 = ne^2/(\varepsilon_1 m_n)$ — плазменная частота.

Если волна поляризована перпендикулярно направлению магнитного поля, то

$$\varepsilon_{\perp}'^2 \approx \varepsilon_1\left[1 - \frac{\omega^2(\omega^2 - \omega_p^2 - \omega_c^2)}{\omega_p^2(\omega^2 - \omega_c^2)}\right].$$

Различие электрических проницаемостей $\varepsilon_{\perp}'$ и $\varepsilon_{\parallel}'$ означает, что в полупроводнике, изотропном при $B = 0$, в магнитном поле возникают явления двойного лучепреломления, а, следовательно, и эффект Коттона—Мутона (или эффект Фохта), и магнитооптические эффекты Керра. Их особенностью в полупроводниках является возможность появления наряду с циклотронным также и плазменного резонанса.

Дальнейшее изложение продолжает знакомить с явлениями, которые возникают под воздействием электромагнитного излучения. Их можно объединить под общим названием *фотоэлектрические*. Они проявляются в изменении электрических свойств полупроводника под воздействием электромагнитного излучения. К одному из таких явлений относится *эффект фотонного давления*.

§ 9.15. Эффект фотонного давления

Эффект фотонного давления возникает под воздействием светового давления, которое перемещает свободные равновесные носители заряда от освещенной поверхности к задней стенке приемного элемента, в результате чего возникает ЭДС в направлении излучения. Это явление обычно наблюдается при облучении полупроводника длинноволновым мощным (лазерным) излучением, когда поглощение светового потока происходит на свободных носителях заряда.

Количество световой энергии, поглощаемой в единицу времени в слое с единичной поверхностью и толщиной dx ($\mathbf{e}_x$ — направление распространения света), пропорционально интенсивности света I и толщине слоя dx: $dI = -KI\,dx$, где K — показатель поглощения. Световая энергия, поглощаемая в единицу времени в единице объема,

$$-\frac{dI}{dx} = KI = Ki_0 e^{-Kx}, \qquad (9.60)$$

где I_0 — интенсивность падающего излучения.

В условиях равновесия, когда цепь разомкнута, импульс от падающего излучения уравновешивается силой, действующей на носитель заряда со стороны электрического поля напряженностью E_x:

$$KI_0 e^{-Kx} = c_0 enE_x.$$

После интегрирования последнего уравнения по всей толщине полностью поглощающего образца получим напряжение $U_x = -I_0/(c_0 en)$, которое возрастает с уменьшением концентрации электронов. В полупроводнике с дырочной электропроводностью напряжение меняет знак: $U_x = I_0/(c_0 ep)$. В полупроводнике, близком к собственному, необходимо учитывать влияние носителей обоих знаков:

$$U_x = I_0 \frac{p - \gamma bn}{c_0 e(p - bn)(p + \gamma n)},$$

где $b = \mu_n/\mu_p$; γ — отношение показателей поглощения электронов и дырок.

Эффект может быть использован для обнаружения и измерения излучения мощных лазеров, когда не требуется высокой чувствительности. Так как при длинноволновом излучении не возникают новые избыточные неравновесные свободные носители заряда, то скорость отклика не ограничивается временем жизни носителей заряда.

§ 9.16. Внутренний фотоэффект

Фотоэлектрические явления, при которых происходит перераспределение электронов по энергетическим состояниям в результате поглощения фотонов, называют *внутренним фотоэффектом*.

В отличие от внешнего фотоэффекта, заключающегося в полном отрыве и эмиссии фотоэлектронов, при внутреннем фотоэффекте происходит изменение только энергетического состояния электронов, приводящее к изменению концентрации свободных носителей заряда или их подвижности, а также к перераспределению их внутри кристалла. Если внешний фотоэффект наблюдается при освещении любых веществ, то внутренний фотоэффект характерен только для полупроводников и диэлектриков.

В случае внутреннего фотоэффекта первичным процессом является световая инжекция избыточных носителей заряда. Этот процесс зависит как от внешних (интенсивности и длины волны излучения), так и от внутренних (геометрии образца, показателя поглощения, диффузии и рекомбинации) факторов.

Образование избыточных носителей происходит при поглощении энергии световых квантов. Эта энергия запасается в основном избыточными электронами, в то время как средняя тепловая энергия кристаллической решетки остается практически неизменной. Поскольку в этих условиях нарушается тепловое равновесие между решеткой и электронами, носители заряда, возникшие в результате внутреннего фотоэффекта, являются неравновеенными. Число неравновесных носителей заряда обычно невелико, и запасенная ими избыточная энергия мала по сравнению с тепловой энергией кристаллической решетки. В связи с этим после прекращения облучения процесс установления равновесия между решеткой и электронами сводится к тому, что неравновесные электроны и дырки исчезают (рекомбинируя друг с другом), а температура решетки и всего кристалла (а следовательно, и концентрация равновесных носителей заряда) остается практически неизменной. Учитывая это, можно считать, что освещение и затемнение полупроводника, изменяя концентрацию неравновесных носителей заряда, не влияет на концентрацию равновесных и полная концентрация носителей заряда (n или p) равна сумме концентраций равновесных (n_0 и p_0) и неравновесных (Δn, Δp) носителей заряда: $n = n_0 + \Delta n$, $p = p_0 + \Delta p$.

Естественно предположить, что скорость генерации носителей заряда, т. е. число образуемых неравновесных электронов в единицу времени в единице объема ($\Delta n' = \Delta p' = G$), должно быть пропорционально поглощаемой за то же время в том же объеме световой энергии (9.60):

$$G = \Delta n' = \Delta p' = \beta K I = \beta K I_0 \, e^{-Kx}, \qquad (9.61)$$

где β — коэффициент пропорциональности.

Если интенсивность излучения I выражать числом квантов в секунду, то значение β будет определять число пар, образуемых одним квантом, и называться его *квантовым выходом*. Квантовый выход больше единицы, когда энергия кванта превышает двойную ширину запретной зоны. При этом возникающие носители заряда обладают достаточной кинетической энергией, чтобы в результате ударной ионизации образовать еще некоторое число неравновесных пар. Далее будет рассмотрен собственный внутренний эффект, в котором при поглощении каждого кванта, соответствующего собственной полосе, возникает пара свободных носителей заряда. При этом $\beta_n = \beta_p = \beta = 1$.

К основным уравнениям, описывающим внутренний фотоэффект, относятся уравнения электрического поля и движения зарядов под действием поля и диффузии, а также уравнение непрерывности:

$$\begin{aligned}
\frac{\partial p}{\partial t} &= G - \frac{p - p_0}{\tau} - \frac{1}{e}\operatorname{div}\mathbf{J}^+; \\
\frac{\partial n}{\partial t} &= G - \frac{n - n_0}{\tau} + \frac{1}{e}\operatorname{div}\mathbf{J}^-; \\
\mathbf{J}^+ &= e\mu_p p\mathbf{E} - eD_p \operatorname{grad} p; \\
\mathbf{J}^- &= e\mu_n n\mathbf{E} + eD_n \operatorname{grad} n; \\
\mathbf{J} &= \mathbf{J}^+ + \mathbf{J}^-; \\
\operatorname{div}\mathbf{E} &= e(\Delta p - \Delta n)/\varepsilon,
\end{aligned} \qquad (9.62)$$

где τ — время жизни носителей заряда; $\mathbf{J}^+, \mathbf{J}^-$ — плотности дырочного и электронного тока; $\mu_n = b\mu_p$; $D_p, D_n = bD_p$ — подвижности и коэффициенты диффузии дырок и электронов; b — отношение подвижностей.

В большинстве случаев (кроме низкоомных образцов) хорошей аппроксимацией является предположение, что во всем образце соблюдается условие электрической нейтральности, т. е. $\Delta n = \Delta p$.

Система основных уравнений, описывающих внутренний фотоэффект, слишком сложна для того, чтобы решать ее в общем виде, поэтому в каждой конкретной задаче приходится делать дополнительные упрощения.

В качестве примера рассмотрим стационарный ($\partial/\partial t = 0$) внутренний фотоэффект в собственном полупроводнике ($n_0 = p_0 = n_i$) в виде прямоугольной плоскопараллельной пластинки, длина и ширина которой намного превышают ее толщину d. Образец равномерно освещается по всей поверхности yz в направлении оси x, начало которой совпадает с

освещенной поверхностью, а теневая поверхность соответствует $x = d$. В этих условиях можно пренебречь краевыми эффектами и считать задачу одномерной. Тогда систему основных уравнений (9.62) с учетом (9.61) можно переписать в виде

$$\begin{gathered} KI_0 e^{-Kx} - \frac{\Delta p}{\tau} - \frac{1}{e}\frac{dJ_x^+}{dx} = 0; \\ J_x^+ = en_i\mu_p E_x - eD_p\frac{d\Delta p}{dx}; \\ J_x^- = en_i b\mu_p E_x + ebD_p\frac{d\Delta n}{dx}; \\ J_x^+ + J_x^- = 0; \\ \Delta p - \Delta n = \frac{\varepsilon}{e}\frac{dE_x}{dx}. \end{gathered} \tag{9.63}$$

При соблюдении нейтральности ($\Delta n = \Delta p$) и отсутствии внешнего электрического поля $E_x = 0$ совместное решение полученной системы уравнений приведет к дифференциальному уравнению

$$L^2\frac{d^2\Delta p}{dx^2} - \Delta p = -KI_0 e^{-Kx},$$

где $L = \sqrt{D\tau}$ — диффузионная длина.

Решение уравнения имеет вид

$$\Delta p = C_1 \operatorname{ch}\frac{x}{L} + C_2 \operatorname{sh}\frac{x}{L} + \frac{KI_0\tau}{1 - K^2L^2}e^{-Kx}. \tag{9.64}$$

Постоянные интегрирования C_1 и C_2 найдем из граничных условий $J_x^+ = -es\Delta p$ при $x = 0$ и $J_x^+ = es\Delta p$ при $x = d$, где s — скорость поверхностной рекомбинации, имеющая одинаковое значение на передней (освещаемой) и задней поверхностях.

При достаточной толщине пластины, когда $\operatorname{sh}(d/L) \approx \operatorname{ch}(d/L) \gg 1$,

$$\Delta p = \frac{KI_0\tau}{K^2L^2 - 1}\left[\frac{\alpha + KL}{\alpha + 1}e^{-x/L} - e^{-Kx}\right]. \tag{9.65}$$

В частности, на освещенной поверхности ($x = 0$)

$$\Delta p(0) = \frac{KI_0\tau}{(\alpha + 1)(KL + 1)}. \tag{9.66}$$

В зависимости от способа наблюдения внутренний фотоэффект может вызывать вторичные явления самого различного характера. Так, в

отсутствие внешних полей появление неравновесных носителей заряда приводит к *кристалл-фотоэффекту* (возникновение электрического поля в однородном неравномерно освещенном полупроводнике).

Заметное влияние внутреннего фотоэффекта на магнитные свойства магнитных полупроводников было обнаружено сначала в ферримагнитном ЖИГ легированным кремнием, а затем в $CdCr_2Se_4$, легированным галлием. Было обнаружено, что динамическая петля гистерезиса этого кристалла при частоте изменения магнитного поля 50 Гц и температуре 77 К существенно расширялась при освещении белым светом. При плотности мощности света 5 мВт/см^2 динамическая проницаемость уменьшается на порядок. В этом соединении свет влияет также на коэрцитивную силу, ферромагнитный резонанс и намагниченность. Подобные явления получили название *фотоферромагнитного эффекта*. Эффект представляет интерес с точки зрения создания принципиально новых приборов. Их перспективность связывают с тем, что магнитные характеристики соединений типа $CdCr_2Se_4$ претерпевают существенные изменения при малых интенсивностях света (10^{-3} мВт/см^2) и имеют времена релаксации на 3—5 порядков более короткие, чем для аналогичных явлений в ЖИГ и других ферритах. Затрудняет их использование необходимость применения криогенной техники (температура Кюри 130 К).

§ 9.17. Фотопроводимость

При наблюдении внутреннего фотоэффекта во внешнем электрическом поле можно обнаружить изменение электропроводности полупроводника под действием освещения. Это происходит в результате появления неравновесных свободных носителей заряда (Δn, Δp). При этом полная концентрация носителей заряда $n = n_0 + \Delta n$ и $p = p_0 + \Delta p$, а полная проводимость $\sigma = e(\mu_n n + \mu_p p) = \sigma_0 + \Delta\sigma$. Следовательно, избыточная проводимость или *фотопроводимость*, обусловленная фотоэффектом, $\Delta\sigma = e\mu_n \Delta n + e\mu_p \Delta p$.

Например, для достаточно толстого образца, согласно (9.65), фотопроводимость

$$\Delta\sigma \approx \frac{e(\mu_n + \mu_p) K I_0 \tau}{K^2 L^2 - 1}\left(\frac{\alpha + KL}{\alpha + 1} e^{-x/L} - e^{-Kx}\right)$$

зависит от расстояния x до освещенной поверхности.

Увеличение проводимости под действием освещения приводит к изменению тока, когда к образцу приложено внешнее напряжение. Если внешнее поперечное поле E_y столь мало, что не нарушает рассмотренного

ранее стационарного распределения неравновесных носителей заряда, то *фототок* через поперечное сечение ($\delta \times d$) всего образца

$$\Delta I = E_y \delta \int_o^d \Delta\sigma \, \mathrm{d}x = \delta \, \mathrm{d} \, \overline{\Delta\sigma} E_y,$$

где средняя фотопроводимость

$$\overline{\Delta\sigma} = \frac{1}{d} \int_o^d \Delta\sigma \, \mathrm{d}x = \frac{e\mu}{d} \Delta P.$$

В пределах применимости уравнения (9.63)

$$\Delta P = \int_0^d \Delta p \, \mathrm{d}x = \frac{\tau I_0}{K^2 L^2 - 1} \left[KL \frac{\alpha - KL}{\alpha + 1} \left(1 - \mathrm{e}^{-d/L}\right) - 1 + \mathrm{e}^{-Kd} \right]. \qquad (9.67)$$

В частном случае, когда в толстом образце ($Kd \gg 1$) благодаря хорошей прозрачности длина диффузии много меньше глубины проникновения ($KL \ll 1$) и толщины образца ($L \ll d$), $\Delta P \approx \tau I_0$. В общем случае $\Delta P = \tau_{эф} I_0$, где $\tau_{эф}$ — эффективное время жизни носителей заряда, которое определяется не только рекомбинацией в объеме, но и всеми процессами исчезновения носителей заряда, в частности рекомбинацией на поверхности толстого образца с учетом его толщины и биполярной диффузии. Например, согласно (9.67),

$$\tau_{эф} = \frac{\tau}{K^2 L^2 - 1} \left[KL \frac{\alpha + KL}{\alpha + 1} \left(1 - \mathrm{e}^{-d/L}\right) - 1 + \mathrm{e}^{-Kd} \right].$$

Тогда средняя фотопроводимость образца

$$\overline{\Delta\sigma} = e\mu\tau_{эф} I_0 / d$$

и фототок

$$\Delta I = \delta e \mu \tau_{эф} I_0 E_y. \qquad (9.68)$$

Полупроводниковые приборы, действие которых основано на явлении фотопроводимости, называют *фоторезисторами*. Они образуются пленкой или пластинкой фоточувствительного материала, снабженной двумя невыпрямляющими контактами для включения в электрическую цепь.

Удельное сопротивление полупроводников, применяемых для изготовления фоторезисторов, в неосвещенном состоянии обычно велико (до 10^9 Ом·см). Это вынуждает располагать электроды так, чтобы расстояние между ними было мало в направлении тока и велико в перпендикулярном направлении.

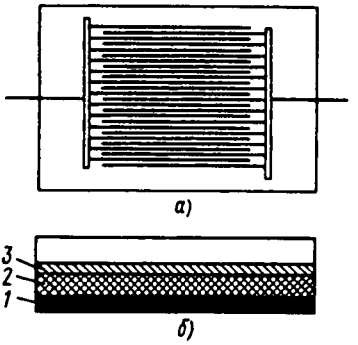


Рис. 9.33. Фоторезисторы:

а — с поперечным освещением (гребешковый); *б* — с продольным освещением

Различают фоторезисторы с поперечным и продольным освещением. В первом случае свет падает перпендикулярно направлению тока. Наилучшим использованием светочувствительной поверхности обладают фоторезисторы в форме гребешков (рис. 9.33, *а*). Фоторезисторы с продольным освещением имеют по меньшей мере один прозрачный электрод (рис. 9.33, *б*). На непрозрачный электрод *1*, образующий подложку, наносят тонкий слой светочувствительного полупроводника *2*, покрытый, в свою очередь, тонким слоем прозрачного металла *3*; последний представляет собой второй электрод.

Фоторезисторы характеризуются эффективностью, коэффициентом усиления и добротностью.

Эффективностью или *фоточувствительностью* называют отношение относительного изменения фотопроводимости резистора к световому потоку, падающему на всю поверхность. Если освещаемый нормально фотопроводящий слой имеет форму прямоугольника шириной δ, толщиной d и длиной l, то световой поток $\Phi = I_0 l\delta$, проводимость в неосвещенном состоянии $G_0 = e\mu n_0 \delta d/l$ и фотопроводимость $\Delta G = \overline{\Delta\sigma} S = e\mu\tau_{эф} I_0 \delta/l$. Тогда эффективность

$$\gamma = \frac{1}{\Phi}\frac{\Delta G}{G_0} = \frac{\tau_{эф}}{n_0 V}$$

оказывается обратно пропорциональной объему фотопроводника $V = l\delta d$.

Можно показать, что для тонких или прозрачных полупроводников ($Kd \ll 1$) эффективность обратно пропорциональна освещаемой поверхности $l\delta$. Все это заставляет при конструировании стремиться к уменьшению либо всех линейных размеров фоточувствительного элемента, либо освещаемой поверхности.

Коэффициентом усиления или *квантовым выходом* фоторезистора называют отношение числа неравновесных носителей заряда N_e, проходящих через сопротивление, к числу поглощенных за это же время фотонов N_ν. Поскольку, согласно (9.68), $N_e = \Delta I/e = \mu\tau_{эф} I_0 E_y \delta$ и $N_\nu = I_0 l\delta$, то коэффициент усиления

$$K = \frac{N_e}{N_\nu} = \frac{\mu\tau_{эф}}{l} E_y = \frac{\tau_{эф}}{T}, \qquad (9.69)$$

где $T = l/(\mu E_y)$ — время, затрачиваемое носителем заряда на прохождение всего пути между электродами.

Полосу пропускания фоторезистора можно приближенно определить как величину, обратную времени жизни носителя заряда ($\Delta B \approx \tau_{эф}^{-1}$). Тогда, согласно (9.69), *добротность*

$$K\Delta B \approx T^{-1} = \mu U/l^2 .$$

Из полученных формул, казалось бы, следует, что и коэффициент усиления и его произведение на полосу пропускания могут неограниченно возрастать при увеличении напряжения или при уменьшении расстояния между электродами. В действительности же пролетное время T может уменьшаться только до тех пор, пока оно не сравняется с временем релаксации неравновесных носителей заряда $\tau_{рел}$. Поскольку

$$\tau_{рел} = \frac{\varepsilon}{\Delta\sigma} = \frac{\varepsilon d}{e\mu\Delta P} = \frac{C}{e\Delta P_\Sigma}\frac{l^2}{\mu}$$

и $T = l^2/(\mu U)$, то при $T = \tau_{рел}$ должно выполняться равенство $UC = e\Delta P_\Sigma$, где $C = \varepsilon\delta d/l$ — межэлектродная емкость фоторезистора, $\Delta P_\Sigma = l\delta\Delta P$ — полное число пар возбужденных светом свободных носителей заряда.

Таким образом, после того как пролетное время сравняется с временем релаксации, внешнее напряжение начнет вводить в фотопроводник объемный заряд, равный заряду, генерированному светом. Следовательно, в области достаточно высоких напряжений $K\Delta B \approx 1/\tau_{рел}$ и время фотоответа равно времени релаксации. Поэтому чем меньшей инерционностью обладает фоторезистор, тем меньше его квантовый выход.

§ 9.18. Фотогальванический эффект

Возникновение ЭДС в *p-n*-переходе в результате разделения фотоносителей электрическим полем, обусловленным неоднородностью полупроводника, называют *фотогальваническим эффектом*.

Поглощение фотонов с энергией, большей ширины запрещенной зоны, сопровождается образованием электронно-дырочных пар с обеих сторон *p-n*-перехода. При этом существенно поведение неосновных носителей заряда, поскольку именно их плотность может меняться в широких пределах, в то время как плотность основных носителей остается практически неизменной. Под действием поля перехода неосновные носители заряда быстро переносятся полем *p-n*-перехода и становятся уже основными носителями заряда на другой его стороне. Поскольку *p-n*-переход

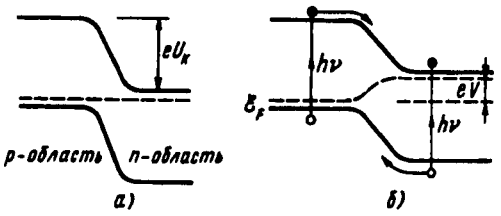


Рис. 9.34. Схемы энергетических уровней *p-n*-перехода:

а — без освещения; *б* — с фотовозбуждением

действует как сток для генерированных светом неосновных носителей заряда, то возникает дополнительный градиент концентрации, обусловливающий диффузию неосновных носителей заряда, находящихся на расстоянии диффузионной длины от перехода. В результате нейтрализации некоторой части объемного заряда *p-n*-перехода происходит уменьшение потенциального барьера V и смещение уровней Ферми относительно друг друга по обе стороны от перехода (рис. 9.34). Разность потенциалов V между двумя областями достигает наибольшего значения U_x при разомкнутой цепи, при закороченной цепи ток достигает наибольшего значения $I_к$ (рис. 9.35).

Фото-ЭДС является источником напряжения в полупроводниковых преобразователях световой энергии в электрическую (солнечные батареи), а также в измерительных и индикаторных приборах (фотодиоды). В этих приборах обычно используются низкоомные полупроводники, в которых внутреннее электрическое поле E_x очень мало и поэтому им можно пренебречь, считая, что для неосновных носителей заряда

$$J^+ = e\mu_p p E_x - eD_p \frac{\partial \Delta p}{\partial x} \approx -eD_p \frac{\partial \Delta p}{\partial x};$$
$$J^- = e\mu_n n E_x + eD_n \frac{\partial \Delta n}{\partial x} \approx eD_n \frac{\partial \Delta n}{\partial x}, \qquad (9.70)$$

причем в стационарных условиях

$$KI_0 e^{-Kx} = \frac{\Delta p}{\tau_p} + \frac{1}{e}\frac{\partial J^+}{\partial x};$$
$$KI_0 e^{-Kx} = \frac{\Delta n}{\tau_n} - \frac{1}{e}\frac{\partial J^-}{\partial x}, \qquad (9.71)$$

где τ_n, τ_p — время жизни электронов и дырок.

Общее решение уравнений (9.70) и (9.71) для *n*-области имеет вид (9.64), аналогичный вид имеет выражение Δn для *p*-области. Постоянные интегрирования могут быть найдены из гранич-

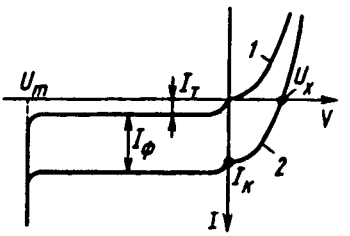


Рис. 9.35. Вольт-амперная характеристика фотодиода до (кривая *1*) и после (кривая *2*) освещения

ных условий в предположении, что приповерхностная область толщиной d_n обладает электропроводностью n-типа (как это реализуется в кремниевых солнечных элементах) и в p-области толщиной d_p происходит полное поглощение светового потока. Тогда из граничных условий на освещенной поверхности ($x = 0$) $J^+ = -es\Delta p$, на p-n-переходе ($x = d_n$) $\Delta p = p_0 e^{eV/(k_Б T)}$, $\Delta n = n_0(e^{V/(k_Б T)} - 1)$* и на теневой поверхности ($x = d_n + d_p$) $\Delta n = 0$ плотности токов через p-n-переход

$$\frac{J^+}{e} = \frac{KL_p I_0}{K^2 L_p^2 - 1}\left\{\frac{KL_p + \alpha[\alpha\,\mathrm{ch}(d_n/L_p) + \mathrm{sh}(d_n/L_p)]e^{-Kd_n}}{\alpha\,\mathrm{sh}(d_n/L_p) + \mathrm{ch}(d_n/L_p)} - KL_p\, e^{-Kd_n}\right\} -$$

$$- \frac{p_0 D_p}{L_p}\frac{\alpha\,\mathrm{ch}(d_n/L_p) + \mathrm{sh}(d_n/L_p)}{\alpha\,\mathrm{sh}(d_n/L_p) + \mathrm{ch}(d_n/L_p)}\left(e^{eV/(k_Б T)} - 1\right), \qquad (9.72)$$

$$\frac{J^-}{e} = \frac{KL_n I_0}{K^2 L_n^2 - 1}\left(KL_n - \mathrm{cth}\frac{d_p}{L_n} - e^{-Kd_p}\,\mathrm{cosech}\frac{d_p}{L_n}\right)e^{-Kd_n} -$$

$$- \frac{n_0 D_n}{L_n}\left(e^{eV/(k_Б T)} - 1\right)\mathrm{cth}\frac{d_p}{L_n}, \qquad (9.73)$$

где $L_p = \sqrt{D_p \tau_p}$ и $L_n = \sqrt{D_n \tau_n}$.

Сумма выражений (9.72) и (9.73) представляет собой плотность полного тока через переход

$$J = J_ф - J_0(e^{eV/(k_Б T)} - 1), \qquad (9.74)$$

где плотность тока короткого замыкания ($V = 0$) равна фототоку

$$\frac{J_ф}{eI_0} = \frac{KL_p + \alpha - e^{-Kd_n}[\alpha\,\mathrm{ch}(d_n/L_p) + \mathrm{sh}(d_n/L_p)]}{(KL_p - 1/KL_p)[\alpha\,\mathrm{sh}(d_n/L_p) + \mathrm{ch}(d_n/L_p)]} + \frac{1}{KL_n - 1/KL_n} \times$$

$$\times [KL_n - \mathrm{cth}(d_p/L_n) - e^{-Kd_p}\,\mathrm{cosech}(d_n/L_p)]e^{-Kd_n} - \frac{KL_p}{KL_p - 1/KL_p}e^{-Kd_p} = \nu,$$

а напряжение холостого хода ($J = 0$) равно фото-ЭДС

$$V_ф = \frac{k_Б T}{e}\ln\left(1 + \frac{J_ф}{J_0}\right).$$

При большом уровне освещения $eV_ф = k_Б T \ln(e\nu I_0/J_0)$. При малом уровне возбуждения ($J_ф \ll J_0$) фото-ЭДС $V_ф = k_Б T \nu I_0/J_0$ пропорциональна интенсивности света I_0, где ν — коэффициент собирания, равный

---

* p_0 и n_0 — темновые концентрации дырок и электронов в n- и p-областях.

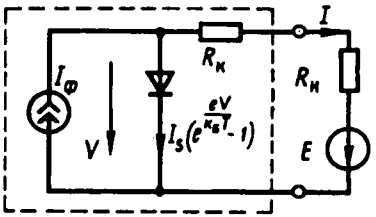


Рис. 9.36. Эквивалентная схема фотодиода

относительной доле неравновесных носителей, доходящих до *p-n*-перехода без рекомбинации.

Эквивалентная схема фотодиода, соответствующая уравнению (9.74), изображена на рис. 9.36, вместе с внешней цепью, включающей резистор нагрузки $R_н$ и внешний источник E. В *вентильном режиме* этот источник отсутствует ($E = 0$). Поэтому ток нагрузки

$$I = I_ф - I_s(e^{eV/(k_Б T)} - 1) = V/R,$$

где $R = R_н + R_к$. Ток достигает максимального значения $I_ф$ при коротком замыкании ($V = 0$). Если $I_ф \ll I_s$, то $I_ф \approx V_ф/(k_Б T/eI_0) = V_ф/R_к$, где $R_к = k_Б T/(eI_0)$ — внутреннее сопротивление фотодиода. В *диодном режиме*, когда $E > 0$ и $V < 0$, при $I_ф \ll I_0$ ток нагрузки $I \approx I_ф + I_s e|V|/k_Б T = I_ф + |V|/R_к$ оказывается больше, чем максимальный ток в вентильном режиме. К тому же в диодном режиме он может быть получен на большом резисторе нагрузки. Поэтому снимаемое с него напряжение сигнала намного больше, чем в вентильном режиме. Чувствительность при обратном смещении увеличивается также вследствие расширения области барьерного слоя и генерации большего числа фотоносителей в области сильного поля. Уменьшение емкости *p-n*-перехода при обратном смещении позволяет намного повысить быстродействие диода по сравнению с вентильным режимом. В этом случае большее быстродействие также связано с тем, что время отклика ограничивается временем пролета носителей в узкой *n*-области, которое намного меньше времени рекомбинации дырок, ограничивающее быстродействие в вентильном режиме. Для повышения чувствительности и быстродействия необходимо использование максимального обратного напряжения, которое однако не должно превышать пробойного значения U_m. Быстродействие современных фотодиодов доходит до 10^{-9} с.

Диодные структуры составляют основу большинства разновидностей фотоприемников, применяемых в оптоэлектронике. Одной из них являются *p-i-n-фотодиоды*. Для этой структуры характерно наличие высокоомной *i*-области (с собственной электропроводностью), разделяющей *p*- и *n*-области с толщиной, значительно превышающей толщину переходов на границах сильно легированных областей. Уже при небольших обратных смещениях в высокоомной области (ее сопротивление в 10^6—10^7 раз превышает сопротивление легированных областей) возникает сильное электрическое поле, обеспечивающее быстрое рассасывание (собира-

ние) генерируемых светом электронов и дырок, малые рекомбинационные потери и высокое быстродействие. Этому способствуют также малые значения *RC*, которые для кремниевых диодов площадью 0,02 мм^2 оказываются порядка $10^{-11}...10^{-12}$ с. У типичных быстродействующих *p-i-n*-диодов полоса пропускания составляет 10^9 Гц.

В полупроводниках, в которых не удается создать *p-n*-переходы, фотодиод может быть изготовлен на основе контакта полупроводника с металлом (фотодиод с барьером Шотки или *фотодиод Шотки*). Его быстродействие доходит до 10^{-10} с.

Основной недостаток рассмотренных диодов — сравнительно малая фоточувствительность, обусловленная отсутствием усиления. Одним из путей повышения коэффициента усиления является использование лавинного усиления фотоносителей в *p-n*-переходе, который нашел свою реализацию в *лавинных фотодиодах*. В нем лавинное усиление начинается с инжекции неосновных носителей заряда, возникающих в результате внутреннего фотоэффекта. Время переключения лавинных фотодиодов приблизительно составляет $10^{-8}...10^{-9}$ с.

Кроме лавинного механизма, для усиления тока фотодиода могут быть использованы транзисторные структуры. Основной тип фотоприемника с внутренним усилением (*фототранзистор*) можно рассматривать как комбинацию фотодиода с транзистором. Поэтому характеристики $i_к = f(u_{кэ})$ фототранзистора *n-p-n-* или *p-n-p*-типа аналогичны характеристикам соответствующих фотодиодов, масштаб по оси токов которых увеличен в $G_ф = (1-\alpha)^{-1}$ раз ($G_ф$ — коэффициент усиления фототока; α — коэффициент передачи по току). Недостатком фототранзисторов является сравнительно высокая инерционность (10^{-4}—10^{-6} с).

Увеличить выходную мощность сигнала в 10^4 раз по сравнению с входной мощностью можно за счет большего времени накопления заряда в *МДПДМ-фотоприемниках*[*].

В этой структуре между электродами расположен изолированный диэлектриком полупроводниковый кристалл с высоким удельным сопротивлением. После включения напряжения формируется обедненный слой, занимающий всю толщу полупроводника. Генерируемые излучением в объеме полупроводника фотоносители движутся к границе с диэлектриком, где они накапливаются, уменьшая толщину обедненного слоя. Связанное с этим перераспределение напряжения в многослойной структуре регистрируется на нагрузочном резисторе в виде электрического импульса, амплитуда которого пропорциональна суммарному числу падающих фотонов. Время запоминания (накопления) оптического сигнала 5...25 мкс.

---

[*] Металл—диэлектрик—полупроводник—диэлектрик—металл.

Подобные структуры являются перспективными для создания малоинерционных высокочувствительных фотоприемников (инерционность 10^{-10} с, пороговая чувствительность при температуре жидкого азота 10^{-13} Вт/см^2). Время хранения информации при температуре жидкого азота превышает 10^3 с, а при комнатной температуре — десятки миллисекунд.

§ 9.19. Элементы фотопамяти на МНОП-структурах

В настоящее время ни одна фоточувствительная среда не удовлетворяет всей совокупности требований к запоминающим устройствам. Дальнейшее рассмотрение ограничено лишь одним из наиболее перспективных структур: металл—нитрид кремния—диоксид кремния—кремний (МНОП).

Способность этой структуры к запоминанию информации основана на наличии в запрещенной зоне Si_3N_4 специфических уровней ловушек $\mathcal{E}_D$ и $\mathcal{E}_A$. Нитрид кремния играет роль накопительной среды для носителей заряда, источником которых служит полупроводник (Si). Прослойка SiO_2 между ними необходима для создания дополнительного потенциального барьера. Она позволила повысить время хранения заряда в нитриде кремния до десятков лет. Оптимальная толщина прослойки около 2 мм. Зонная диаграмма структуры в отсутствие напряжения изображена на рис. 9.37.

В аналогичном элементе памяти на основе МНОП-транзистора с электрическими записью и считыванием запись «1» осуществляется подачей положительного импульса, переключающим прибор в состояние с высоким пороговым напряжением. При этом электроны переходят из кремниевой подложки в подзатворную область диэлектрика. Для переключения структуры в состояние с низким пороговым напряжением («0») на затвор подается отрицательный импульс напряжения, возвращающий электроны из нитрида в кремниевую подложку. Считывание производится по уровню тока при импульсах напряжения с амплитудами, меньшими напряжений перезарядки центров памяти в нитриде.

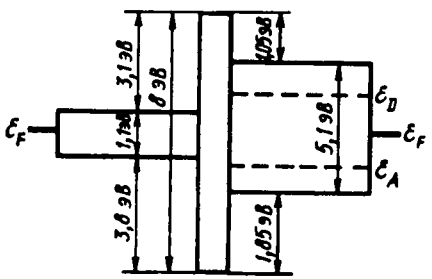


Рис. 9.37. Зонная диаграмма МНОП-структуры

После приложения к МНОП-структуре импульса инвертирующей полярности в полупроводнике образуется толстый обедненный слой, на который падает большая часть приложенного напряжения. При этом электрическое поле в диэлектрике невелико

и перезарядки центров памяти в нитриде не происходит. По мере формирования инверсионного слоя напряжение между областью приповерхностного заряда (ОПЗ) полупроводника и диэлектриком начинает перераспределяться, поле в диэлектрике увеличивается. Когда оно превысит пороговое значение, возникает инжекция носителей в Si_3N_4 и изменение зарядов центров памяти в нитриде. Освещение структуры через полупрозрачный электрод существенно ускоряет формирование инверсного слоя, стимулируя переключение элемента памяти из одного логического состояния в другое. Плотности светового потока, необходимые для переключения, были получены в пределах $(1,5...5,0)\cdot10^{-8}$ Дж/мм^2 на длине волны 630 мм. Время переключения обычно составляет 0,2...5,0 мкс.

Использование МНОП-структуры в качестве элемента фотопамяти основано на двух ее характерных свойствах. Во-первых, накопление заряда в слое диэлектрика имеет пороговый характер: происходит только при напряжении, превосходящем пороговое. Во-вторых, значительная плотность туннельного тока (несколько ампер на квадратный сантиметр) в тонком слое SiO_2 при напряженностях поля около 10^7 В/см позволяет записывать и считывать информацию достаточно быстро. Стимулированное светом переключение структуры происходит благодаря увеличению напряжения на диэлектрике до порогового значения за счет созданной светом поверхностной плотности заряда σ на границе полупроводника и диэлектрика. При поглощении света полупроводником вблизи его границы с диэлектриком возникают электронно-дырочные пары. Они разделяются электрическим полем ОПЗ, в результате чего у поверхности диэлектрика происходит накопление неосновных носителей, формирующих заряд плотностью σ. Скорость его накопления зависит не только от интенсивности потока фотонов I, но и от диффузионной длины $\mathcal{L}$ носителей заряда в кремнии:

$$\frac{d\sigma}{dt} = \beta_1 eI\left(1 - \frac{e^{-KW}}{1 + K\mathcal{L}}\right),$$

где β_1 — квантовый выход; W — толщина обедненного слоя. При этом предполагается, что время разделения зарядов в ОПЗ пренебрежимо мало (10^{-10} с).

Зависимость напряжения на диэлектрике от плотности поверхностного заряда можно найти приближенно, решив уравнение Пуассона $\frac{\partial^2\varphi}{\partial x^2} = -\rho/\varepsilon$ при следующих допущениях: заряд неравновесных носителей σ рассматривается как плоский заряженный слой; изменением заряда на поверхностных состояниях можно пренебречь, заряд в диэлектрике отсутствует (исходное состояние), объемная плотность $\rho = 0$ при

$0 < x < d$ и $\rho = eN$ при $x > d$, где d — толщина диэлектрика; N — концентрация основных носителей заряда в кремнии. Интегрирование уравнения Пуассона при граничных условиях

$$\varphi(0) = U, \quad \varphi_g(d) = \varphi_n(d), \qquad \varepsilon_d \left(\frac{\partial \varphi}{\partial x}\right)_{x=d} - \varepsilon_n \left(\frac{\partial \varphi}{\partial x}\right)_{x=a} = \sigma$$

приводит к напряжению на диэлектрике

$$U = \varphi(0) - \varphi(d) = \frac{eNd}{\varepsilon_d}\left[l - \sqrt{l^2 - \frac{2l}{eN}\left(\frac{\varepsilon_g}{d}U + \sigma\right)}\right] - \frac{d}{\varepsilon_d}\sigma, \qquad (9.75)$$

где U — напряжение на структуре; $l = \varepsilon_n d / \varepsilon_g$ — «приведенная толщина» диэлектрика; ε_g, ε_n — диэлектрические проницаемости диэлектрика и полупроводника. Из полученного выражения следует, что при $\sigma \approx -\varepsilon_d U / d$ заряд инверсного слоя равен заряду металлического электрода и все напряжение (с точностью до напряжения на инверсном слое) оказывается приложенным к диэлектрическому слою.

В зависимости от типа электропроводности полупроводниковой подложки можно управлять с помощью света инжекцией в диэлектрике электронов либо дырок, так как обедненный слой возникает только при одной полярности управляющего напряжения. Когда необходимо иметь управление светом при обеих полярностях переключающего напряжения, используют МДП-структуры, содержащие p-n-переход в полупроводниковой подложке.

Считывание информации удобнее всего производить путем регистрации фототока J, протекающего во внешней цепи структуры при освещении (рис. 9.38). При считывании на структуру $C_{стр}$ сначала подается импульс напряжения $E_г$, приводящий к образованию в полупроводнике обедненного слоя. После завершения связанного с этим переходного процесса в структуру подается импульс света, генерирующий носители,

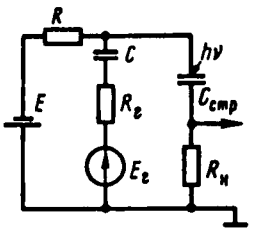


Рис. 9.38. Схема считывания информации

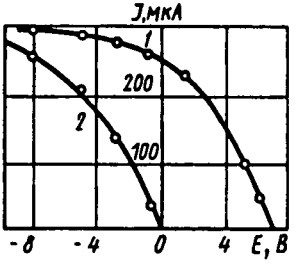


Рис. 9.39. Зависимость сигнала считывания от напряжения смещения для двух зарядовых состояний

которые разделяются в поле пространственного заряда и изменяют плотность поверхностного заряда неосновных носителей σ на границе Si—SiO_2. Это приводит к подзарядке структуры током от внешнего источника E.

Амплитуда регистрируемого фототока при считывании определяется только зарядовым состоянием ловушек на освещенном участке структуры. Из представленных на рис. 9.39 зависимостей фототока для участков с записанной информацией (кривая *1*) и без нее (кривая *2*) видно, что можно подобрать напряжение постоянного смещения E так, чтобы фототок возникал только при наличии записанной ранее информации. Частота считывания может быть доведена до 10 МГц.

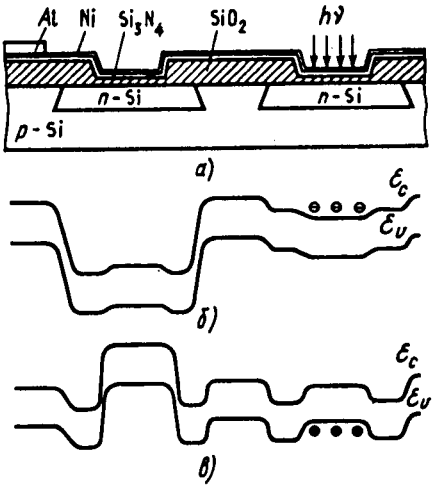


Рис. 9.40. МДП-структура с утолщениями в зазорах между ячейками памяти (*а*) и изменение потенциала границы полупроводник-диэлектрик при $U > 0$ (*б*) и $U < 0$ (*в*)

Плотность записи информации определяется дрейфом носителей заряда вдоль границы Si—SiO_2 и энергетической чувствительности к свету. Как следует из (9.75), поверхностный потенциал полупроводника $\varphi(d)$ может быть изменен выбором соответствующей толщины диэлектрика d. Изменяя $\varphi(d)$, можно создать потенциальные барьеры, препятствующие растеканию неосновных носителей зарядов, путем профилирования толщины диэлектрика (рис. 9.40). Подобными конструктивными и технологическими приемами плотность записи информации может быть доведена до $5 \cdot 10^4$ бит/мм^2.

Диэлектрическая изоляция ячеек допускает возможность применения третьего измерения, что позволяет увеличить плотность записи информации еще на порядок. При этом предполагается использование света, проникающего достаточно глубоко в подложку, например, излучения инжекционного лазера. В освещенном участке структуры адресация осуществляется одновременно по всем слоям (рис. 9.41), а кодирование информации — напряжением, подаваемым на электроды каждого слоя.

Для записи информации в ячейку памяти с площадью поверхности 10 мкм^2 достаточно светового импульса с энергией около 10^{-12} Дж. В МНОП-структурах с p-n-переходом в полупроводниковой подложке чувствительность к свету может быть увеличена еще на один-два порядка. В режиме считывания энергия светового импульса в 3—5 раз меньше, чем в

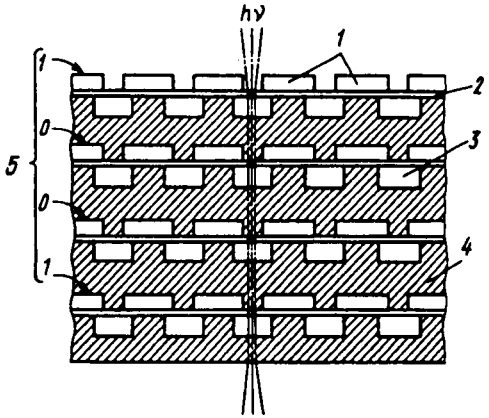


Рис. 9.41. Структура с трехмерной интеграцией МДПДП-ячеек памяти:

1 — управляющие электроды; *2* — поляризующийся диэлектрик; *3* — фоточувствительный полупроводник; *4* — изолирующий диэлектрик; *5* — числовой код

режиме записи и стирания. Существенно, что структура позволяет сохранять информацию в течение нескольких лет, не потребляя при этом энергию. Высокая чувствительность среды к свету позволяет записывать и считывать информацию большими массивами, например, байтами, освещая одновременно девять дорожек. Прогнозируемая емкость подобных оптоэлектронных ЗУ составляет 10^{10} бит.

Реверсивные МДП-структуры в сочетании с лазерной электронно-лучевой трубкой позволяют проводить режим обработки информации, характерный для дисковых накопителей. Возможности подобного оптоэлектронного реверсивного запоминающего устройства превосходят на три порядка магнитные накопители с механическим перемещением запоминающей среды.

Время накопления заряда на ловушках в результате туннелирования из полупроводника в МНОП-структуре составляет десятки микросекунд. Однако эффективная скорость обмена данных оказывается достаточно большой при вводе информации блоками. Например, для блока емкостью 10^3 бит эффективная скорость обмена данных составляет 10^8 бит/с. Если же использовать передачу информации по параллельным оптическим каналам, то она может достигнуть 10^{12} бит/с. Время доступа к блоку информации по произвольному адресу составляет $10^{-6}...10^{-5}$ с.

В обычных электронных запоминающихся устройствах увеличение информационной емкости ограничено ростом числа электрических соединений. Электрические соединения снижают тактовую частоту устройств и увеличивают уровень перекрестных помех. В современных сверхбольших ИС площадь, занимаемая проводниками, почти равна площади кристалла, а общая длина проводников почти в 10^3 раз превышает линейные размеры кристалла. На перезарядку проводников затрачивается 70...75% от всех энергетических затрат.

В рассмотренном ранее оптоэлектронном запоминающем устройстве благодаря адресации и передачи команд управления по оптическим каналам число электрических соединений внутри структуры резко сокращается. Таким образом, создание светочувствительных МНОП-структур

открывает качественно новый уровень развития реверсивных запоминающих сред на основе МДП-интегральной технологии.

§ 9.20. Оптически управляемые пространственно-временны́е модуляторы света на МДП-структурах

В МДП-структуре отсутствует *p-n*-переход и разделение носителей заряда на границе полупроводника первоначально возникает под действием внешнего поля. Однако другие проявления внутреннего фотоэффекта аналогичны фотогальваническому эффекту: разделение фотоносителей происходит в электрическом поле неоднородного полупроводника и сопровождается возникновением фототока, изменением электрического поля объемного заряда на границе с диэлектриком и появлением соответствующей ему ЭДС. Изменение концентрации свободных носителей заряда в приповерхностном слое полупроводника изменяет его электропроводность, т. е. возникает также явление, аналогичное эффекту поля в полупроводниках (изменение электропроводности приповерхностного слоя полупроводника под воздействием электрического поля).

Высокая чувствительность МДП-структур к световым сигналам находит всевозрастающее применение в оптически управляемых модуляторах света с однородной модулируемой средой. Одним из примеров может служить используемый для оптической обработки информации модулятор на основе двухслойной структуры из фотопроводника и электрооптического кристалла (рис. 9.42).

Модуляция света при считывании методом отражения основана на эффекте Поккельса в кристалле DKDP (KD_2PO_4) при температуре, несколько превышающей точку Кюри. При рабочих температурах (от –65 до –50 °С) время хранения информации на кристалле до считывания превышает 1 ч, а полуволновое напряжение составляет только 200 В.

Запись, считывание и адресация информации осуществляются оптическим путем. Для линейного отклика в отраженном свете к фотопроводнику прикладывается напряжение 80 В. При этом на фотопроводник проектируется или сканируется входное изображение. При достаточной экспозиции большая часть общего напряжения падает на DKDP, что вызывает пространственную модуляцию показателя преломления, а, следовательно, и считы-

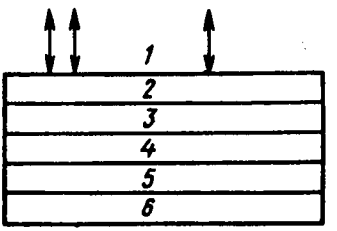


Рис. 9.42. Структура пространственного модулятора «Фототитус»:

1 — считывающий свет; *2* — прозрачный электрод; *3* — DKDP; *4* — диэлектрическое зеркало; *5* — фотопроводник; *6* — прозрачный электрод

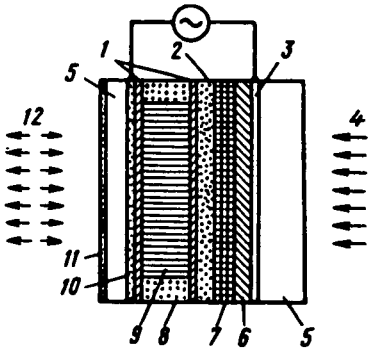


Рис. 9.43. Пространственный модулятор с жидкокристаллическим отражателем:

1 — нейтральный изолирующий слой; *2* — диэлектрическое зеркало; *3* — прозрачный электрод; *4* — записывающий свет; *5* — стекло; *6* — фотопроводник (CdS); *7* — светопоглощающий слой; *8* — прокладка; *9* — нематический жидкий кристалл; *10* — прозрачный электрод; *11* — просветляющее покрытие; *12* — считывающий свет

вающего света, отраженного от диэлектрического зеркала. Время записи составляет 0,3 мкс, время считывания — 30...300 мкс, контраст — 100:1, чувствительность — 10 мкДж/см^2. Разрешение (150 лин/мм) сравнимо с разрешением системы PROM, но требуемое напряжение (80 В) значительно меньше. Другим преимуществом системы «Фототитус» является возможность использования одной и той же длины волны при записи и считывании.

Изменяя напряжение, поданное на кристалл, можно вычесть из записанного изображения постоянную амплитуду, т. е. осуществить операцию вычитания. Система «Фототитус» применяется в когерентных оптических корреляционных системах как на входе, так и в плоскости пространственного фильтра.

Уровень рабочих напряжений в оптоэлектронном пространственном модуляторе с МДП-структурой можно снизить на порядок, используя в отражающем слое гибридный полевой эффект в *жидком кристалле*. Закрытое состояние определяется эффектом скручивания, а открытое — деформацией ориентированных фаз. В таком приборе (рис. 9.43) записывающий луч генерирует неравновесные носители заряда в фотопроводнике, осуществляя пространственную модуляцию его сопротивления. Поэтому возникающее под действием приложенного напряжения поле в жидком кристалле изменяется в пространстве пропорционально интенсивности падающего света. Пока не приложено напряжение, молекулы ориентированы параллельно электродам (гомогенная структура) и изогнуты под углом 45°. Линейно поляризованный считывающий свет, распространяясь по ячейке слева направо, испытывает поворот плоскости поляризации на 45°; после отражения от диэлектрического зеркала плоскость поляризации поворачивается обратно на тот же угол. Поэтому при скрещенных поляризаторе и анализаторе свет не проходит (закрытое состояние), а вращение плоскости поляризации определяется только эффектом скручивания.

Если к жидкому кристаллу с положительной диэлектрической анизотропией приложено значительное напряжение, то оси молекул повернутся перпендикулярно электродам (возникнет гомеотропная структура); для

распространяющегося вдоль них света двупреломление станет равным нулю и при скрещенных поляризаторе и анализаторе свет вновь не пройдет. Однако при промежуточных значениях напряжения имеются молекулы, наклоненные под острым углом к электродам (промежуточное состояние между гомогенным и гомеотропным), причем примерно половина молекул ориентируется в направлении одного электрода, а другая — в направлении другого (под углом 45° относительно молекул первой половины). Поэтому двупреломление станет максимальным, когда направление поляризации падающего луча будет параллельно направлению, определяемому входным электродом, и молекулы первой половины ячейки не будут влиять на поляризацию падающего света.

Преобразователь с жидким кристаллом толщиной 2 мкм при напряжении 6 В с частотой 10 кГц и чувствительностью 10 мкДж/см^2 имел разрешение 114 лин/мм, контраст 100:1 и быстродействие 10 мс. Для прозрачных электродов использована пленка из оксида олова и индия, для изолирующих слоев — оксид кремния, для фотопроводника — сульфид кадмия, для светоблокирующего слоя — теллурид кадмия. Записанное изображение может быть превращено из негативного в позитивное изменением напряжения (от 5,5 до 3,6 В) и поворотом анализатора (на 22,5°). Промежуточные значения уровня прохождения отраженного света и контраста могут быть получены при промежуточных значениях напряжения или уменьшения частоты (с 10 до 2 кГц).

Широкому использованию рассмотренных оптически управляемых модуляторов препятствуют их недостатки. Структуру PROM отличают невысокие разрешающая способность и чувствительность, необходимость освещения при стирании информации, сравнительно высокое питающее напряжение. Для структуры «Фототитус» требуется меньшее рабочее напряжение. Однако остаются необходимым дополнительное освещение для стирания информации, наличие фазовых искажений в преобразованном изображении и особенно охлаждение и вакуумирование, приводящие к увеличению габаритов и энергопотребления. Фоторезистивный слой в предыдущем транспаранте обычно изготовляют из поликристаллических и аморфных полупроводников, которые недостаточно технологичны, обладают малой квантовой эффективностью и ограниченным спектральным диапазоном фоточувствительности (сине-зеленая и ультрафиолетовая области спектра). Указанные полупроводниковые материалы в этих структурах также обладают малой эффективностью управления (активным током чисто емкостной нагрузки жидкокристаллического слоя) и малой чувствительностью.

Отмеченные недостатки удалось преодолеть в результате применения в качестве фоточувствительного слоя симметричной структуры МДП (МДПДМ).

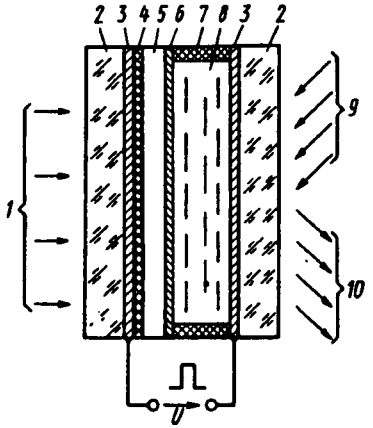


Рис. 9.44. Схематический разрез оптически управляемого модулятора на основе слоев жидкого кристалла и МДПДМ-структуры:

1 — входное (управляющее) изображение; *2* — подложка; *3* — прозрачный электрод; *4* — диэлектрический слой; *5* — полупроводник; *6* — диэлектрическое зеркало; *7* — прокладка; *8* — жидкий кристалл; *9* — считывающее (управляемое) излучение; *10* — преобразованное изображение

Подобная структура отличается от предыдущей наличием дополнительного тонкого (0,5—0,6 мкм) слоя диэлектрика между фоторезистивным слоем и электродом. От другого электрода полупроводник по-прежнему отделен диэлектрическим зеркалом и жидким кристаллом (рис. 9.44). В качестве полупроводника используются соединения $A^{II}B^{VI}$, а также силикат и германат висмута ($Bi_{12}SiO_{20}$ и $Bi_{12}GeO_{20}$). Применение узко- и широкозонных полупроводников позволяет изготовлять оптически управляемые модуляторы, чувствительные к излучению в инфракрасной, видимой и ультрафиолетовой областях спектра.

После приложения к структуре импульсного напряжения оно распределяется в основном между слоями полупроводника и жидкого кристалла (вклад в перепад напряжения второго достаточно тонкого диэлектрического слоя сравнительно большой емкости незначителен). При этом в полупроводнике возникает неравновесное состояние, при котором происходит обеднение объема свободными носителями заряда. За время, называемое *временем экранирования*, генерация свободных носителей заряда постепенно приводит к тому, что поверхностный заряд заэкранирует внешнее электрическое поле в полупроводнике и практически все приложенное к модулятору напряжение окажется подведенным к жидкому кристаллу.

Время действия импульса напряжения в модуляторе устанавливают меньше времени экранирования, причем соответствующим выбором амплитуды и длительности импульса напряжения добиваются, чтобы к концу импульса жидкий кристалл находился в предпороговом состоянии. Тогда при включении управляющего освещения в результате генерации дополнительных свободных неравновесных носителей заряда и их накопления может быть обеспечен дополнительный прирост напряжения на жидком кристалле, при котором напряжение на нем станет выше порогового и будет осуществлена модуляция считывающего луча.

При этом генерация неравновесных носителей заряда не успевает восстановить равновесную концентрацию носителей обедненного поверхностного слоя. Поэтому его активное сопротивление остается весьма

большим (10^9...10^{11} Ом·см) и необходимый перепад напряжения на диэлектрическом слое жидкого кристалла обеспечивается в результате управляемого светом изменения емкости полупроводникового слоя.

Чувствительность модулятора при использовании слоя высокоомного кремния толщиной 200 мкм достигала 10^8 Дж$^{-1}$·см^2, а разрешение — 6 лин/мм. При применении арсенида галлия эти параметры были равны 10^9 Дж$^{-1}$·см^2 и 11 лин/мм при быстродействии до 10 нс. Поэтому время смены изображения ограничивалось только вязкостью жидкого кристалла и в обычном режиме (при питании синусоидальным напряжением) не могло быть сделано меньше 10 мс. Использование широкозонных полупроводников, в частности силиката и германата висмута, позволило достигнуть разрешения до 300 лин/мм. Двухчастотный режим управления жидким кристаллом уменьшил время включения и выключения до 1 мс. Разработчики предполагают, что совершенствование этого режима даст возможность в дальнейшем улучшить быстродействие еще на 1—2 порядка. Напряжение питания в первых макетах колебалось от 10 до 100 В.

К одному из возможных применений модуляторов на основе жидких кристаллов и МДП-структур относится преобразование инфракрасного изображения в видимое и когерентное. Это позволит отказаться от использования для этой цели электровакуумных приборов (электрических преобразователей и видиконов), требующих высоких напряжений.

Новые возможности оптоэлектронных методов обработки информации открывают использование нелинейных свойств оптически управляемых модуляторов с внутренней обратной связью, благодаря которой в модуляторе появляются оптические гистерезис и многостабильность.

Одна из возможных схем подобных бистабильных оптически управляемых модуляторов, разработанных в ФИАНе изображена на рис. 9.45. Слоистая структура объединяет модулирующий слой из жидкого кристалла, реализующего эффект скручивания (твист-ячейка), и однородный фотоэлектрический слой в виде МДПДМ-структуры. После их последовательного расположения на пути светового луча и подключения элементов структуры к источнику электрического напряжения они образуют замкнутую цепь оптоэлектрической обратной связи.

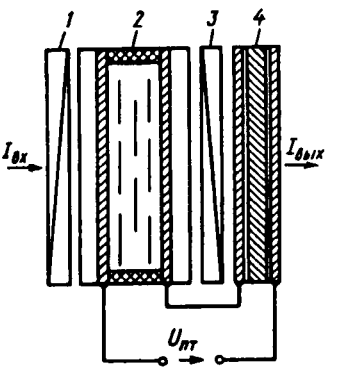


Рис. 9.45. Схема бистабильного пространственно-временно́го модулятора:

1 — поляризатор; *2* — твист-ячейка; *3* — анализатор; *4* — МДПДМ-структура

Пропускание $T = f(U)$ твист-ячейки (рис. 9.46), расположенной между скрещенными поляризатором и анализатором, увеличивается с ростом приложенного к

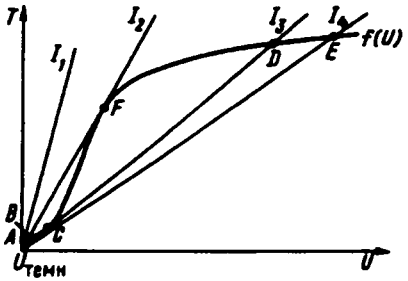


Рис. 9.46. К графическому расчету рабочих режимов бистабильного оптического модулятора

ней напряжения $U = U_{пт} - U_{МДПДМ}$. Это напряжение меньше напряжения питания $U_{пт}$ на напряжение, создаваемое на МДПДМ-структуре $U_{МДПДМ} = U_0 - \alpha T I_{вх}$, которое тем меньше, чем больше интенсивность $I_{вх}$ света, поступающего в твист-ячейку (α — чувствительность МДПДМ-структуры; U_0 — напряжение на ней в затемненном состоянии). Следовательно, благодаря освещению прирост напряжения на твист-ячейке $\Delta U = U - (U_{пт} - U_0) = \alpha T I_{вх}$, откуда $T = (U - U_{темн})/(\alpha I_{вх})$, где $U_{темн} = U_{пт} - U_0$ — напряжение на твист-ячейке в затемненном состоянии. Линейная зависимость пропускания от напряжения приведена для четырех значений $I_{вх}$: $I_1 < I_2 < I_3 < I_4$. Пересечения этих прямых с кривой $f(U)$ определяют возможные рабочие напряжения на жидком кристалле при интенсивностях I_1, I_2, I_3 и I_4. При $I_2 < I_{вх} < I_3$ существует три рабочих точки, из которых устойчивым состояниям соответствуют две, определяющие бистабильный режим.

Если в исходном состоянии сквозь анализатор проходит малая часть светового потока и напряжение на жидком кристалле недостаточно для просветления твист-ячейки, то для того, чтобы напряжение на жидкокристаллическом слое превысило пороговое значение, необходимо воздействие светового импульса либо импульса напряжения. В результате твист-ячейка просветляется, вызывая под воздействием обратной связи дальнейшее возрастание на ней напряжения. Этот процесс развивается лавинообразно. Бистабильные свойства находят отчетливое проявление в гистерезисном характере зависимости оптического пропускания от интенсивности падающего света (рис. 9.47).

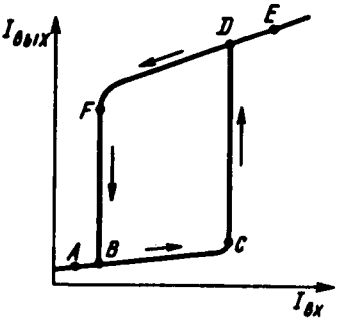


Рис. 9.47. Гистерезисная характеристика бистабильного оптического модулятора

Отличительными особенностями нелинейного оптически управляемого модулятора света с обратной оптоэлектронной связью являются отсутствие электронного усилителя и многоканальность. Благодаря высокой чувствительности фотослоев и использованию

жидких кристаллов удалось довести энергию переключения до 10^{-13} Дж на элемент (бит) при плотности элементов около 10^5 см$^{-2}$.

Рассмотрением фотогальванического эффекта и его применений заканчивается знакомство с наиболее существенными фотоэлектрическими явлениями, связанными с поглощением фотонов и перемещением свободных носителей заряда. В следующем параграфе начинается рассмотрение той группы явлений в полупроводниках, которая связана с генерацией фотонов, т. е. с излучательными переходами.

§ 9.21. Электролюминесценция

Поглощение веществом энергии с последующим электромагнитным излучением в видимой области спектра или близкой к ней называют *люминесценцией*. Точнее люминесценцию определяют как избыточное по сравнению с тепловым излучение тела, если это избыточное излучение обладает длительностью, значительно превышающей период световых колебаний (10^{-15} с). Люминесцентное излучение сопровождается переходом микросистемы с верхнего энергетического уровня на более низкий уровень, т. е. является процессом, обратным поглощению. При рассмотрении люминесценции предполагают, что она определяется спонтанными переходами.

Различные виды люминесценции (фотолюминесценция, катодолюминесценция, рентгенолюминесценция, радиолюминесценция, триболюминесценция и т. д.) отличаются не процессом излучения, а только способами возбуждения. *Электролюминесценция* связана с преобразованием электрической энергии в световое излучение. По методам возбуждения электролюминесценцию можно разделить на два вида: предпробойную, возникающую в сильных полях, близких к тем, при которых образуется пробой и инжекционную (на *p-n*-переходе при прямом смещении).

Предпробойная электролюминесценция. Она наблюдается, в частности, в порошкообразных цинкосульфидных электролюминофорах, погруженных в диэлектрик.

Типичная электролюминесцентная ячейка (рис. 9.48) состоит из подложки *1* с проводящим покрытием *2*, электролюминофора, взвешенного в диэлектрике *3*, и второго электрода *4*.

Обычно в качестве подложки используют стекло, а в качестве первого электрода — прозрачный слой оксида олова с удельным поверхностным сопротивлением 50...2000 Ом·см^2. Люминофором обычно служит активированный сульфид цинка с пластиковой или стеклянной связкой тол-

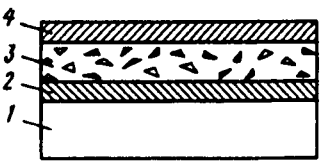


Рис. 9.48. Схема электролюминесцентной ячейки

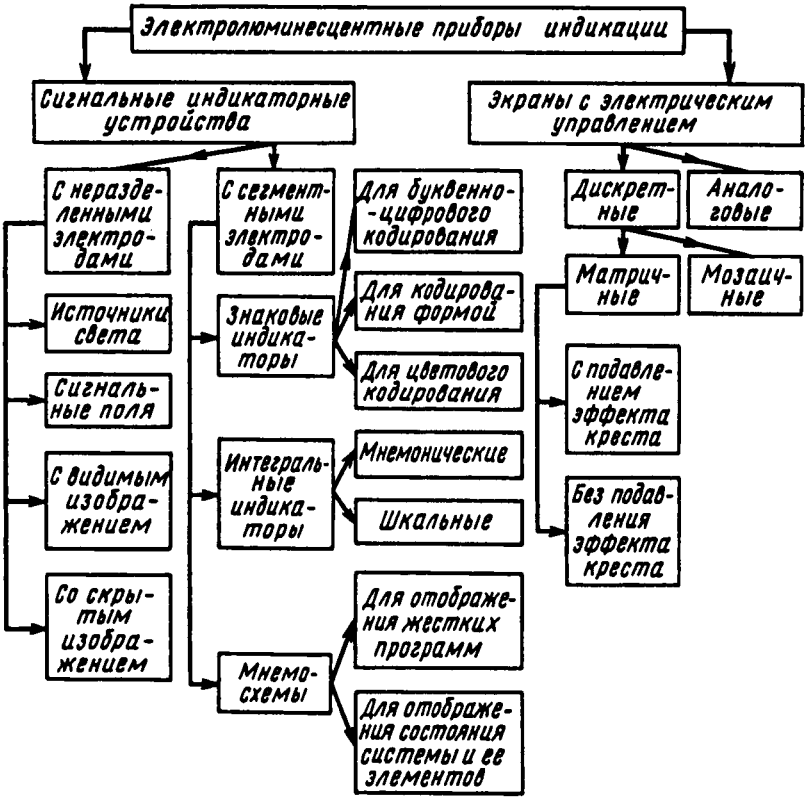


Рис. 9.49. Классификация электролюминесцентных приборов индикации

щиной 50...100 мкм. Изолятор, в котором диспергирован электролюминофор, выполняет две функции: увеличивает локальное поле на зернах люминофора и препятствует прохождению постоянного тока. Использование несущей среды с большой диэлектрической проникаемостью приводит к эквивалентной схеме электролюминесцентного элемента в виде емкости с коэффициентом потерь порядка 0,1.

Реактивный ток электролюминесцентной ячейки много больше тока, совпадающего по фазе с напряжением. Однако компенсирующие индуктивные элементы обычно не используют, поскольку при этом сводится на нет конструктивная простота оптоэлектронных матриц.

Светотехнические свойства ячеек характеризуются начальной яркостью, сроком службы по половинной яркости и полным сроком службы. Полный срок службы электролюминесцентной ячейки нельзя определить числом рабочих часов, поскольку в каждом конкретном случае он зави-

сит от уровня окружающего освещения. Ячейка, работающая при низком уровне освещения, более долговечна по сравнению с ячейкой, работающей при высоком уровне. Увеличивая постепенно приложенное напряжение и его частоту, можно увеличивать яркость и эффективную долговечность.

Электролюминесцентные источники излучения на основе цинкосульфидных и им подобных порошков получили наиболее широкое применение в качестве световых индикаторов различных типов (рис. 9.49). Ломинесцентные приборы индикации отличают высокая надежность (до 10^{-6} ед/ч), большой срок службы (до 5000 ч), малые габариты, малое потребление мощности по сравнению с индикаторами на лампах накаливания при относительно высокой яркости (20...30 кд/м^2), большой угол обзора, возможность многоцветного свечения и изменения цвета свечения, обеспечение практически всех категорий и числа уровней (внутри категории) кодирования, рекомендованных инженерной психологией; относительно низкая стоимость.

Сигнальные и индикаторные устройства предназначаются для отображения кодированной информации в виде алфавита знаков, определяемого конфигурацией электродов и типом люминофоров. Они выполняются из светящихся электролюминесцентных ячеек, которым можно придавать любую форму. Из них можно составлять практически любые двумерные формы: фигуры, буквы, символы и знаки. В опознавательном индикаторном устройстве электроды имеют форму, соответствующую определенному рисунку.

Инжекционная электролюминесценция. При этом виде люминесценции для возбуждения используется инжекция неравновесных носителей заряда *p-n*-переходом, включенным в прямом направлении, а излучение имеет рекомбинационную природу. Межзонная излучательная рекомбинация происходит с большой вероятностью в областях с высокими концентрациями электронов и дырок. Поэтому излучательная рекомбинация сосредоточена в непосредственной близости ($x \approx \pm L$) от *p-n*-перехода или даже в самом обедненном слое, где электроны, диффундирующие в одну сторону, рекомбинируют с дырками, диффундирующими в другую сторону. Баланс между рекомбинацией и диффузионным потоком двух типов носителей определяет стационарное значение тока диода.

Вероятность тех или иных рекомбинационных процессов зависит от структуры энергетических зон кристалла. Наряду с межронными переходами излучение может сопровождать рекомбинацию через промежуточные энергетические уровни (центры рекомбинации), расположенные внутри запрещенной зоны.

Если центром рекомбинации является центр акцепторного типа (рис. 9.50), то сначала электрон переходит с этого центра на свободное место в валентной зоне *1*, что можно рассматривать также как захват дырки цен-

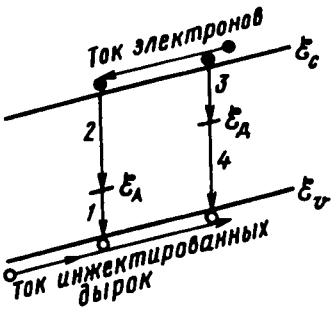


Рис. 9.50. Схема рекомбинации через донорные и акцепторные уровни

тром рекомбинации, а затем положительно ионизированный акцепторный центр захватывает электрон из зоны проводимости *2*, завершая процесс рекомбинации в целом. В случае донорного центра рекомбинации сначала имеет место захват электрона из зоны проводимости *3*, а затем переход его с донорного уровня в валентную зону *4*. Каждый из четырех переходов может быть излучательным или безызлучательным в зависимости от природы центра.

Инжекционная электролюминесценция имеет целый ряд преимуществ перед предпробойной электролюминесценцией:

1) для возбуждения люминесцентного диода достаточны напряжения порядка одного или нескольких вольт (по сравнению с десятками вольт в цинкосульфидных ячейках);

2) более высокая энергетическая эффективность, излучаемая энергия при некоторых условиях может быть даже выше энергии, отбираемой у источника тока (за счет охлаждения кристалла, когда энергия для инжекции одного носителя заряда меньше ширины запрещенной зоны), тогда как электролюминесценция в ZnS является многостадийной, причем каждая стадия связана с тепловыми потерями;

3) меньшая инерционность (при собственном излучении кристалла она определяется временем жизни неосновных носителей заряда и может достигать 10^{-10} с).

Инжекционный источник света, заключенный в корпус или герметизированный, называют *светодиодом*.

Светодиоды обычно изготовляют из полупроводников группы $A^{III}B^{IV}$ (GaAs, InAs, InSb и т. д.), у которых вероятность прямых переходов много больше, чем у германия и кремния. В связи с этим материалы группы $A^{III}B^{V}$ обладают высоким коэффициентом поглощения, малым временем жизни свободных носителей заряда и большой интенсивностью рекомбинационного излучения. Наибольший практический интерес представляют полупроводники, способные излучать видимый свет, т. е. полупроводники с $\mathcal{E}_g > 1{,}6$ эВ. Однако получили распространение материалы, излучающие и в ближней инфракрасной области, поскольку имеются эффективные приемники такого излучения. При комнатной температуре ширина запрещенной зоны арсенида галлия 1,43 эВ, фосфида галлия 2,26 эВ. Ширина запрещенной зоны у карбида кремния может изменяться от 2,3 до 4,1 эВ в зависимости от структуры кристаллической решетки. Максимум в спектре излучения можно смещать, изменяя соотношение между

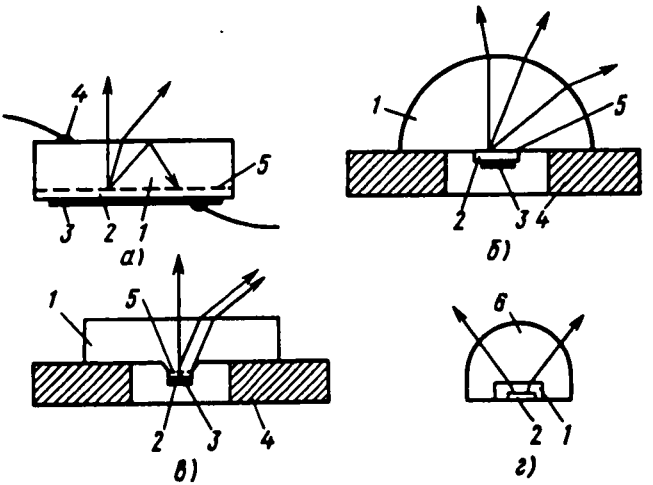


Рис. 9.51. Структуры светодиодов:

а — прямоугольная; *б* — полусферическая; *в* — мезоструктура; *г* — со стеклянным покрытием; *1* — полупроводник *n*-типа; *2* — слой *p*-типа; *3* — контакт к *p*-слою; *4* — контакт к базе; *5* — *p-n*-переход; *6* — стеклянное покрытие

компонентами в твердых растворах. Например, в $GaAs_xP_{1-x}$ свечение перекрывает диапазон от 1,4 до 2,2 эВ. Это дает возможность согласовывать спектры излучения со спектрами поглощения фотоприемников.

Хотя внутренняя квантовая эффективность инжекционной электролюминесценции может быть близка к единице, внешняя эффективность светодиода обычно значительно меньше, поскольку при обычной его конструкции большая часть светового потока испытывает на поверхности полное внутреннее отражение и затем поглощается (рис. 9.51, *а*). Через поверхность выходит та небольшая доля лучей, которая падает под углом меньше критического (около 16°). В простой плоской структуре проходящая доля генерируемого света

$$\eta \approx \frac{1}{2}\left(1-\sqrt{1-\frac{1}{n'^2}}\right)\frac{4n'}{(1+n')2} \approx \frac{1}{n'(1+n')^2},$$

где множитель 1/2 предполагает, что может выйти лишь излучение, эмитируемое в сторону верхней поверхности. В формуле не учитывается поглощение выходящего света в верхней области, хотя предполагается, что весь отраженный свет в конечном счете теряется в результате самопоглощения. При $n' = 3{,}5$ из плоской структуры можно вывести только 1,4% светового потока. Использование полусферы (рис. 9.51, *б*) позволяет улучшить этот результат на порядок. Если излучательный переход

достаточно мал, то свет, эмитируемый в полусферу, падает на поверхность под углом, меньшим критического, так что эффективность вывода света равна $2n'/(1+n')^2$, т. е. $\eta \approx 0{,}35$ при $n' = 3{,}5$.

Внешний квантовый выход можно несколько увеличить, если излучательный переход расположить на столообразном выступе, так называемой *мезе* (рис. 9.51, *в*). В этом случае часть светового потока отражается от боковых стенок мезы и поступает на плоскую поверхность базы по направлению, близкому к нормали. Это увеличивает выход света в два–три раза. Другой метод улучшения условий вывода света, свободный от самопоглощения, заключается в использовании покрытия из стекла с высоким показателем преломления (рис. 9.51, *г*). Применение стекла с $n' = 2{,}2...2{,}5$ позволяет повысить эффективность метода примерно в два раза. Создавая на поверхности просветляющее покрытие, можно увеличить пропускание полупроводника примерно на 30%.

Оптическая эффективность светодиодов может быть повышена с помощью многослойных структур, создающих направленные световые потоки и суженную диаграмму направленности излучения, например, двойных гетероструктур, в которых в области вблизи *p-n*-перехода показатель преломления на 0,15...0,20 больше, чем в прилегающих слоях. Возникающий при этом волноводный эффект определяет оптическое ограничение излучения, уменьшает дифракционные потери, а также сужает диаграмму направленности в направлении, перпендикулярном плоскости *p-n*-перехода.

В эмиттере, в котором показатель преломления плавно уменьшается по мере удаления от активного слоя, фотоны распространяются по криволинейным траекториям и выходят из диода под малыми углами относительно нормали к торцу. При квадратичном профиле показателя преломления выигрыш в эффективности может быть даже пятикратным.

Сужение диаграммы направленности особенно существенно для повышения эффективности ввода излучения в волокно систем оптической связи.

Светодиоды видимого диапазона удобны для вывода информации из вычислительных устройств, особенно бортовых, так как они могут работать неограниченно долго, миниатюрны, практически не выделяют теплоты, потребляют небольшие мощности и работают при тех же напряжениях и токах, что и обычные элементы транзисторных схем. Проблема поиска неисправностей в сложных схемах легко решается при введении светодиодов в эмиттерные цепи транзисторов. Светодиоды являются самыми экономичными индикаторами включения транзисторных радиоприемников и могут служить индикаторами их настройки на радиостанцию. Ведутся работы по созданию цветных плоских телевизионных экранов на базе светодиодов. Одним из их преимуществ будет огромный срок

службы. Это в сочетании с возможностью создать излучение в любой области спектра и исключительно высокой стабильностью позволит применять светодиоды в качестве удобных вторичных эталонов света в оптике и спектроскопии и, в частности, для калибровки абсолютной и спектральной чувствительностей различных фотоприемников и фотоэкспонометров. Малый разброс задержки световых импульсов ($3 \cdot 10^{-11}$ с) дает возможность использования светодиодов с малой длительностью послесвечения (доли наносекунды) для определения временных характеристик фотоприемников и для начала отсчета в схемах совпадений, а малый разброс порогового напряжения свечения (0,1%) позволяет применять их для систем стабилизации чувствительности импульсных фотоприемников.

Быстродействие, высокая яркость и световая мощность светодиодов позволяют успешно применять их и в системах связи, причем пропускная способность их такова, что с помощью одного диода можно одновременно передавать на значительные расстояния несколько десятков тысяч телефонных разговоров или несколько телевизионных программ. Такие системы связи полезны для внестудийных телевизионных передач или для телевизионной связи на предприятиях, где нежелательна или невозможна кабельная связь. Преимуществом светодиодов перед лазерами, обеспечивающими еще большее быстродействие, является большая простота изготовления и применения, долговечность и меньшая стоимость. Широкая полоса пропускаемых частот позволяет применять светодиоды в дальномерах, а также для связи при наличии высокочастотных помех. Кроме того, светодиоды используют для записи звука на пленку в любительском кино и титров в профессиональном кино. Преимуществами записи с помощью светодиодов являются отсутствие линз и зеркал, малые габариты, надежность и экономичность.

Вне конкуренции с другими источниками света находятся светодиоды при оптическом вводе информации в сверхбыстродействующие вычислительные машины. При считывании информации с перфокарт они обеспечивают разрешение в десятки раз выше, срок службы в сотни раз больше, чем при считывании с применением ламп накаливания. Такой же выигрыш дает использование светодиодов в преобразователях непрерывной информации в дискретную, преобразователях угла поворота антенн в электрический сигнал и во всех других устройствах, где механическое перемещение необходимо преобразовать в электрический сигнал.

Одной из основных проблем микроэлектроники в настоящее время является развязка электрических цепей для устранения высокочастотных наводок, которые не позволяют создавать интегральные схемы объемом более нескольких кубических сантиметров. Полностью решает эту проблему применение *оптических связей,* в основе которых лежит преобразование электрического тока в электролюминесцентное излучение и обратное преобразование светового потока в электрический ток за счет

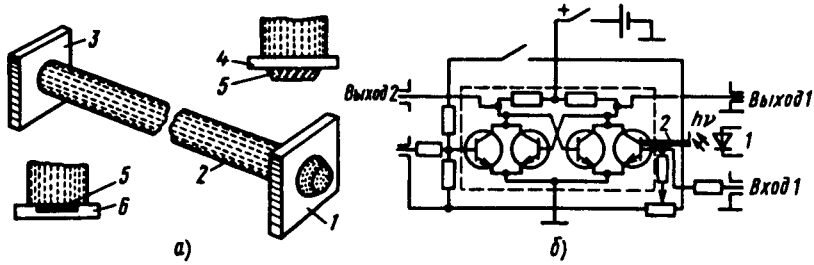


Рис. 9.52. Оптоэлектронное соединение:

а — конструкция; *б* — пример использования в микросхеме триггера; *1* — светодиод; *2* — световод; *3* — фотодиод; *4* — GaAs *n*-типа; *5* — диффузионная p^+-область; *6* — Si *n*-типа

внутреннего фотоэффекта. Использование таких связей между интегральными схемами (вместо миниатюрных шин или соединительных проводников) рассматривается в настоящее время как один из способов повышения надежности, упрощения конструкции и уменьшения размеров устройства. В простейшем виде система соединений (рис. 9.52, *а*) методом оптической связи включает диод-приемник *3*, принимающий через световод *2* излучение и преобразующий его в электрический сигнал. Диод-передатчик *1* может быть выполнен на *p-n*-переходе из арсенида галлия; диод-приемник — из кремния.

Во избежание потерь у поверхностей раздела и для обеспечения согласования между световодом и источником излучения, с одной стороны, и приемником излучения — с другой, необходимо выбрать показатель преломления материала световода, близкий к показателю преломления полупроводникового материала. При использовании кремния и арсенида галлия коэффициент преломления световода должен быть примерно равен 3,6. Наиболее пригодными материалами для практического применения являются неорганические стекла, особенно германатовое стекло и трисульфид мышьяка. Для света с длиной волны 1 мкм германатовое стекло имеет $n' = 1{,}74$. Лучшими данными ($n' = 2{,}47$) обладает трисульфид мышьяка, позволяющий получить световод хорошего качества.

Для создания оптических связей между собранными в общий пакет интегральными схемами требуется большое число световодов. Применение для каждой оптической линии связи отдельных световодов нецелесообразно. По-видимому, достаточно приемлемым решением является применение пластин волоконной оптики, в которых в одной матрице, выполненной из материала с малым показателем преломления, содержится большое число тонких стекловолокон. При использовании для каждой оптической связи отдельного световода требуется высокая точность его совмещения с диодом-излучателем и диодом-приемником. Применение

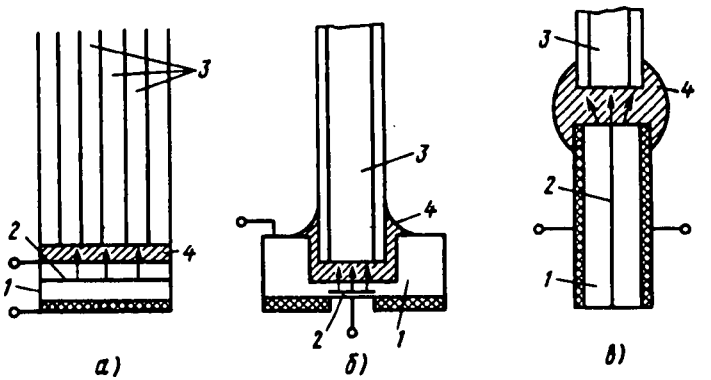


Рис. 9.53. Контактные способы соединения светодиодов со световодами:
1 — светодиодный кристалл; *2* — излучающая область; *3* — световод; *4* — оптический клей

пластин волоконной оптики не требует точного их размещения относительно полупроводниковых интегральных схем. Однако эффективность светопередачи при этом оказывается меньше, чем в случае отдельных световодов.

Эффективность оптической связи во многом зависит от способа соединения излучателя со световодами. Лишь при узкоколлимированном луче и равенстве его поперечного сечения поперечному сечению световода достигается почти 100%-ный ввод света. У светодиодов, имеющих широкую диаграмму направленности, дело обстоит наиболее сложно. Простейшее соединение получают, приклеивая плоский светодиод к торцу многоволоконного жгута (рис. 9.53, *а*). Эффективность использования площади торца, при этом колеблется от 0,4 до 0,7 (1,5...4 дБ), а внеапертурные потери составляют 5,4...17 дБ. Лучшее оптическое согласование достигается при малой площади излучателя и улучшенной диаграмме направленности светодиода (рис. 9.53, *б*), а также при торцевом излучении (рис. 9.53, *в*). При числовой апертуре волокна, меньшей 0,3, торцевые диоды оказываются эффективнее поверхностных. Различные фокусирующие системы (рис. 9.54) позволяют снизить внеапертурные потери светодиода до 5 дБ.

Примером интегральной схемы с оптическими связями является схема триггера на рис. 9.52, *б*. Схема работает на четырех транзисторах с малыми паразитными емкостями и малым напряжением эмиттер — коллектор. Световод *2* из трисульфида мышьяка диаметром 0,3 мм и длиной 12 мм приклеен к одной из сторон микросхемы. На другом конце световода с помощью эпоксидного клея прикреплен светодиод из арсенида галлия.

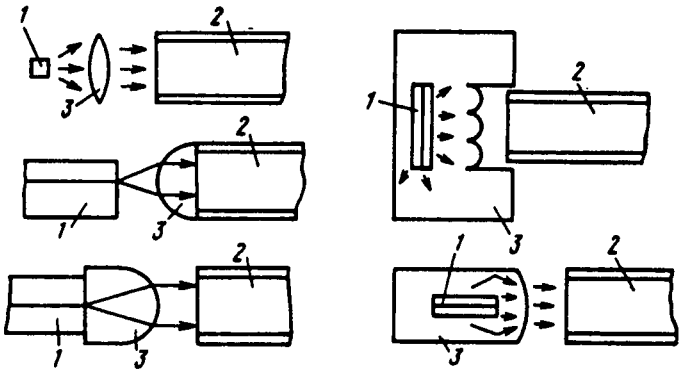


Рис. 9.54. Линзовые способы ввода излучения в световод:
1 — излучатель; *2* — световод; *3* — фокусирующий элемент

Применение оптических связей для создания междусхемных соединений имеет ряд существенных преимуществ. Размеры полупроводниковой пластины, на которой выполняют в настоящее время интегральную схему, во много раз меньше размеров корпуса схемы и ее внешних выводов. Использование оптических связей позволяет собирать узлы, состоящие из набора тонких полупроводниковых пластин, оформленных в одном общем корпусе, и за счет этого получать существенный выигрыш в размерах.

Однако основным преимуществом этого метода связи является полная развязка между входными и выходными цепями за счет одностороннего потока сигналов. Это может существенно упростить взаимодействие между элементами электронных устройств, что особенно важно в больших интегральных схемах ЭВМ.

Основу оптической связи в электронных устройствах, рассмотренной применительно к светодиодам, составляют процессы преобразования электрических сигналов в оптические и оптических сигналов в электрические, процессы распространения излучения в различных средах, а также эффекты взаимодействия электромагнитного излучения оптического диапазона с веществом. Эти процессы и связанные с ними приборы и устройства относятся к новой области электроники, получившей название *оптоэлектроники*. Наряду с индикаторами к числу базовых оптоэлектронных приборов относятся оптроны.

§ 9.22. Оптроны

Оптоэлектронный прибор, в котором имеются объединенные источники и приемник излучения, помещенные в один корпус, называют *оптроном* или *оптопарой*.

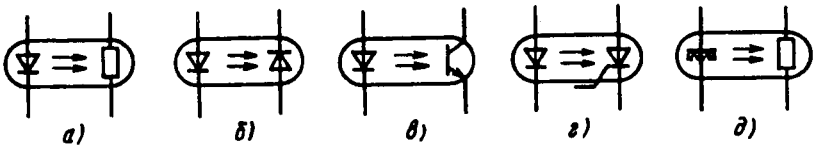


Рис. 9.55. Элементарные оптроны:

а — резистивный; *б* — диодный; *в* — транзисторный; *г* — тиристорный; *д* — резистивный с электролюминесцентным конденсатором

Элементарный оптрон. Такой оптрон состоит из источника света и фотоприемника (оптронной пары) и представляет собой электрооптический четырехполюсник. Он является структурным элементом цепей оптоэлектроники, подобно тому как транзистор или диод является структурным элементом цепей транзисторной электроники. Элементарные оптроны по типу фотоприемника подразделяют на фоторезисторные, диодные, транзисторные и тиристорные; в фоторезисторных оптронах используются наряду со светодиодами и электролюминесцентные конденсаторы (рис. 9.55). Быстродействие фоторезистивных фотоприемников лежит в пределах $10^{-2}...10^{-3}$ с, фототранзисторных — $10^{-4}...10^{-7}$ с, лавинных и *p-i-n*-фотодиодных — 1 нс, а лучших образцов — 0,2 нс. Высокое быстродействие фотодиода позволяет считать, что несмотря на его малый коэффициент усиления, для оптоэлектронных схем вычислительной техники наиболее перспективен оптрон типа светодиод-фотодиод.

Различие конструкций оптронов связано с желанием обеспечить высокие значения одного из параметров или дополнительные функциональные возможности. В частности, использование стеклянной прокладки (рис. 9.56, *а*) позволяет уменьшить проходную емкость до 0,01 пФ и повысить допустимое напряжение гальванической развязки. Введение длинного стеклянного световода (рис. 9.56, *б*) доводит напряжение гальванической развязки до десятков киловольт. Максимальное значение коэффициента передачи по току достигается, когда все световые лучи перпендикулярны границе раздела благодаря усложнению формы излучателя и фотоприемника (рис. 9.56, *в*) или с помощью дополнительного отражателя (рис. 9.56, *г*). Дальнейшее увеличение светопередачи, степени интеграции и микроминиатюризации связано с переходом на монолитные конструкции оптронов (рис. 9.57).

Схема элементарного оптрона (рис. 9.58), как функционального элемента, включает в себя управляемый источник света *ИС*, преобразующий входной электрический сигнал $X(u_1, i_1)$ в световой сигнал $B(u_1, i_1)$; канал оптической связи *ОС*, функция передачи которого $B(\gamma)$, и фотоприемник *ФП*, преобразующий оптический сигнал на его входе $B(u_1, i_1, \gamma)$ в электрический сигнал на его выходе $Y(u, i)$. В общем случае комплексное

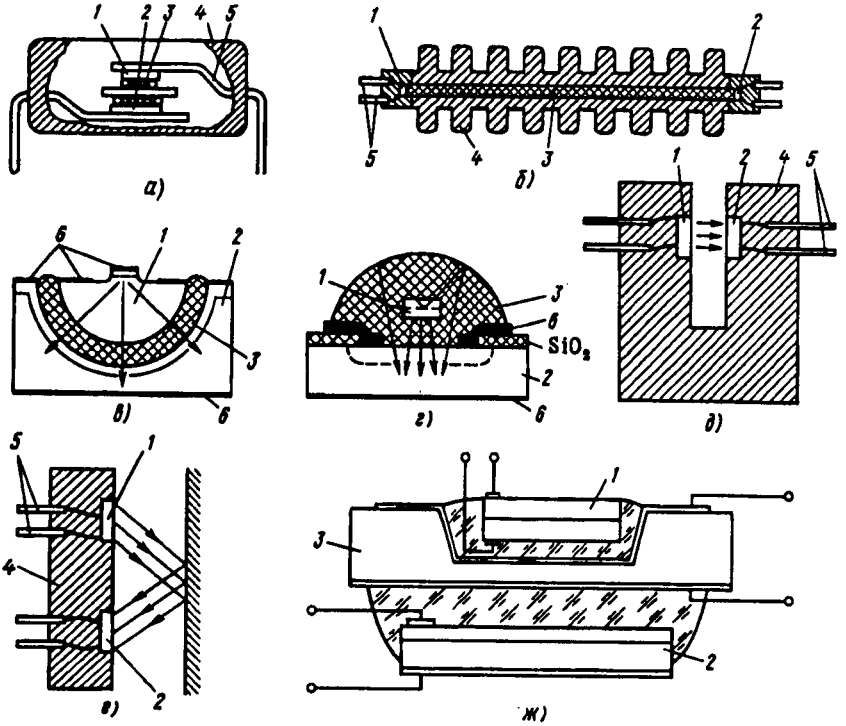


Рис. 9.56. Составные оптроны:

а — в DIP-корпусе; *б* — высоковольтный оптоизолятор; *в* — полусферический; *г* — с пластмассовой полусферой; *д* — оптопрерыватель; *е* — отражательный; *ж* — с управляемой оптической средой; *1* — излучатель; *2* — фотоприемник; *3* — оптическая среда; *4* — корпус; *5* — внешние выводы; *6* — металлические электроды

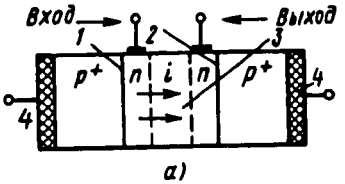
сопротивление фотоприемника $Z = f(B_1, u_2, i_2)$, где u_2, i_2 — смещающее напряжение и ток на втором независимом входе оптрона.

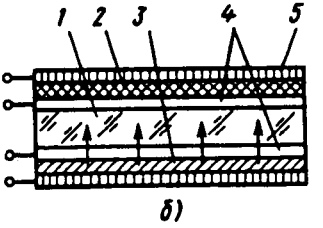
Коэффициент передачи элементарного оптрона

$$K_U = u/u_1 = \xi(K_{\text{ИС}} K_{\text{ФП}} K_{\text{ОС}}),$$

где $K_{\text{ИС}}$, $K_{\text{ФП}}$, $K_{\text{ОС}}$ — коэффициенты передачи *ИС*, *ФП* и *ОС*.

В частности, при $K_{\text{ИС}} K_{\text{ФП}} K_{\text{ОС}} = \text{const}$ для линейных излучателей и фотоприемников коэффициент передачи оптрона соответствует операциям умножения $K_U = C_u u_1 u_2$ и $K_i = C_i i_1 i_2$, где C_u, C_i — постоянные коэффициенты. В этом случае оптрон позволяет получить квадратичное преобразование $K_U = C_u u_0^2$ (при $u_1 = u_2 = u_0$), моделировать деление





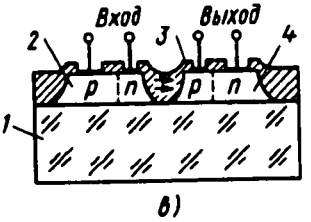


Рис. 9.57. Монолитные оптроны:

а — монокристаллический; *1*, *2* — излучающая и фотопреобразующая области; *3* — оптическая среда (световод) ; *4* — внешние выводы; *б* — тонкопленочный; *1* — стеклянное основание; *2* — фоторезистивная пленка; *3* — электролюминофор; *4* — прозрачные электроды; *5* — непрозрачные электроды; *в* — планарно-эпитаксиальный; *1* — сапфировая подложка; *2* — излучатель; *3* — оптическая среда; *4* — приемник

$K_U = C_u u_2/u_3$ (при $u_1 = 1/u_3$) и алгебраическое суммирование $K_U = C_0 \sum u_i$ (при $u_2 = \text{const}$ и $u_1 = \sum_i u_i$).

Функциональные характеристики оптрона определяются также видом функции $K_{OC}(\gamma)$, где γ — фактор управления по оптической среде, характеризующий функциональную нагрузку световых лучей и соответствующий третьему независимому входу оптрона. Преобразование информации под воздействием сигнала $\gamma(u_3, i_3)$ на этом входе может осуществляться различными процессами, в частности электрооптическими, магнитооптическими и фотоэлектрическими эффектами, которые воздействуют на интенсивность, направление, поляризацию, фазу и длину волны светового пучка. Таким образом могут быть осуществлены амплитудная и фазовая модуляции светового потока, его коммутация, переключение по каналам оптической связи, размножение каналов и пр.

В оптроне, изображенном на рис. 9.56, *ж*, оптической управляемой средой служит арсенид галлия, осуществляющий оптическую модуляцию с помощью эффекта Келдыша—Франца. К преимуществам такого модулятора относятся спектральная согласованность со светодиодами из арсенида галлия, простота изготовления, возможность микроминиатюризации, высокая эффективность и практическая безынерционность (10^{-14} с), ограничиваемая

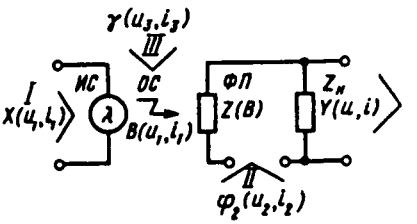


Рис. 9.58. Функциональная схема оптрона

лишь постоянной времени электрической цепи. Модуляция осуществляется напряжением от 0 до 100 В.

Принципиальная особенность оптронов как элементов связи заключается в использовании для переноса информации электрически нейтральных фотонов, что обусловливает ряд его преимуществ:

а) отсутствие гальванической связи между входом и выходом (достижимо сопротивление изоляции более 10^{14} Ом, а емкости связи — менее 10^{-14} Ф) и связанная с этим помехозащищенность;

б) высокая электрическая прочность («развязываемые» напряжения могут превышать десятки киловольт);

в) однонаправленность потока информации и, следовательно, отсутствие обратной реакции приемника на источник;

г) широкая полоса пропускания (от постоянной составляющей до гигагерцевого диапазона);

д) широта функциональных связей, включая бесконтактные (механические и электрические), например через воздух, вакуум (космос).

Для рассмотренных элементарных оптронов характерно наличие только *прямой оптической связи* между излучателем и фотоприемником. По назначению оптроны с прямой оптической связью можно разделить на четыре функциональные группы:

1) информационные, обеспечивающие высокоскоростную передачу информации по гальванически развязанной цепи (оптоэлектронное реле);

2) управляющие, предназначенные для бесконтактного управления электрическими цепями;

2) энергетические, реализующие изолированные вторичные источники питания;

4) линейные, предназначенные для аналогового преобразования сигналов.

Наиболее широко в качестве фотоприемников аналоговых оптронов применяются фоторезисторы благодаря их большому динамическому диапазону, разнообразию вольт-амперных и световых характеристик, а также простой схеме включения. Такой оптрон может быть использован как подстроечный элемент в мостовых схемах, схемах АРУ и как элемент составного потенциометра. Если выходными параметрами потенциометра являются напряжение и ток, то в оптроне осуществляются масштабное преобразование и, в частности, усиление. При наличии двух функциональных входов могут быть реализованы операции умножения, деления и алгебраического сложения. При моделировании операции дифференцирования *RC*-контур формируется обычно в цепи светодиода. По сравнению с обычными пассивными *RC*-дифференциаторами с глубокой положительной связью оптоэлектронный дифференциатор обладает значительно более простой схемой, большим запасом устойчивости и большим

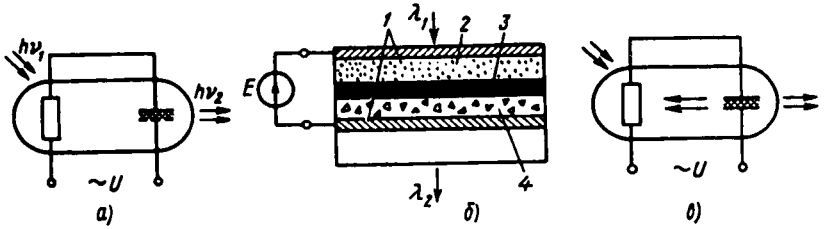


Рис. 9.59. Оптоэлектронный усилитель света:

а — схема; *б* — конструкция; *в* — регенеративный оптрон; *1* — прозрачный проводник; *2* — фотопроводник; *3* — непрозрачный проводник; *4* — электролюминофор

динамическим диапазоном при той же точности математического действия (2...3%).

Функциональные возможности оптрона могут быть существенно расширены при использовании *электрооптической связи*, т. е. когда между источником и приемником излучения имеется электрическая связь. К подобным оптронам относится и активный оптрон с обратной связью.

Активный оптрон. Простейшим представителем подобных оптронов является оптрон, работающий в режиме усилителя света. В нем имеется только электрическая связь между излучателем и приемником света за счет последовательного включения в общую электрическую цепь (рис. 9.59, *а*). При этом согласование характеристик оптоэлектронной пары по комплексному сопротивлению, току и напряжению наиболее просто можно осуществить только для электролюминесцентного конденсатора и резистора (рис. 9.59, *б*). В отсутствие освещения напряжение, приложенное к электролюминесцентному слою, ограничено высоким темновым сопротивлением фотопроводника. Полное напряжение на электролюминофоре появляется после освещения фотопроводящего слоя, когда его сопротивление уменьшается. При этом уровень выходного сигнала зависит от уровня входного сигнала. Усилитель света был использован для преобразования рентгеновского изображения в яркое видимое изображение (вместо малоэффективных флуоресцентных экранов). Различие спектральных характеристик приемника и излучателя позволяет осуществить аналог электронно-оптического преобразователя инфракрасного излучения в видимое, избавляя от присущих электронно-лучевым трубкам больших габаритов и высоких напряжений.

Схема регенеративного оптрона отличается от усилителя света наличием обратной положительной оптической связи между источником света и фотоприемником (рис. 9.59, *в*). Такую обратную связь в конструкции рис. 9.59, *б* можно осуществить, если электрод *3* сделать прозрачным. Тогда достаточно подать на фотопроводник импульсное возбуждение,

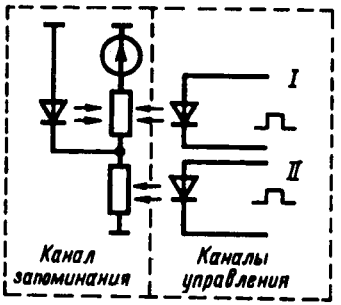


Рис. 9.60. Схема бистабильного оптрона с цепями записи *I* и стирания *II* информации

чтобы выходной световой сигнал со слоя электролюминофора поддерживал сопротивление фотопроводникового слоя на низком уровне. Подобный *усилитель яркости с запоминанием* может быть использован в системах цифровой индикации, где скорость поступления информации низка, а ее отображение осуществляется в течение длительного периода времени.

Регенеративный режим лежит в основе работы *бистабильных оптронов* (рис. 9.60), являющихся базовыми элементами электролюминесцентных экранов с памятью. Регенеративные оптроны позволяют обеспечить усиление света, спектральное преобразование, повышение контрастности и преобразование некогерентного излучения в когерентное (при использовании оптических квантовых генераторов).

В отличие от усилителя света, у которого входные и выходные сигналы световые, у *усилителя напряжения* (рис. 9.61, *а*) оба сигнала электрические. Регенеративный оптрон с усилителем осуществляет замкнутый цикл преобразований сигнала (рис. 9.61, *б*). Коэффициент регенерации этого функционального кольца $K = \eta BSJK_y$ представляет собой произведение передаточных функций каждого из звеньев: дифференциальные коэффициентов светоотдачи излучателя ($\eta = \partial P_{изл}/\partial i_{изл}$), светопропускания оптического звена ($B = \partial P_{ф}/\partial P_{изл}$), фоточувствительности приемника ($S = \partial i_{ф}/\partial P_{ф}$), внутреннего усиления фотоприемника ($G = \partial i_{фп}/\partial i$), усиления внешнего усилителя преобразования ($K_y = \partial i_{изл}/\partial i_{фп}$). Варьируя характеристики и способы включения излучателя, приемника и преобразователя *У*, можно получать устройства, управляемые током (с отрицательным сопротивлением), напряжением (с отрицательной

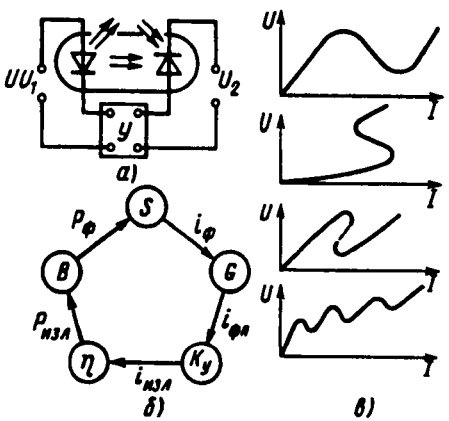


Рис. 9.61. Регенеративный оптрон с усилителем:

а — электрическая схема; *б* — структура передаточной функции; *в* — вольт-амперные характеристики

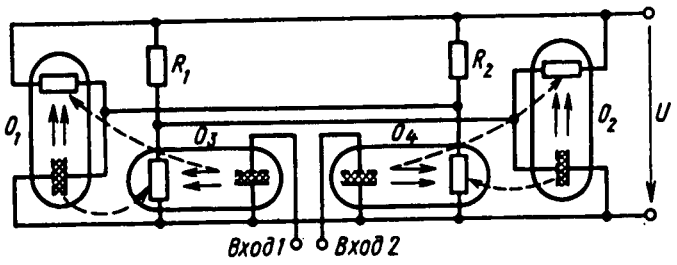


Рис. 9.62. Триггер на регенеративных оптронах

электропроводностью) или с полной регенерацией в нескольких точках вольт-амперной характеристики (рис. 9.61, *в*), т. е. создавать многоустойчивый оптрон.

Функциональные возможности регенеративного оптрона расширяет использование управляемого оптического канала с помощью, например, электрооптического эффекта. Применение для этой цели жидких кристаллов позволяет достигнуть усиления по мощности, равного 10^4, а с прозрачной сегнетоэлектрической ЦТСЛ-керамикой частотный диапазон оптрона доходит до 100 МГц.

Регенеративные оптроны удобны в оптоэлектронных логических схемах, элементах памяти, сдвигающих регистрах, полусумматорах и т. п., прежде всего как бистабильные элементы. Примером может служить *триггер* (рис. 9.62), где сопротивления резисторов R_1 и R_2 выбирают значительно меньшими, чем сопротивления утечки электролюминесцентных конденсаторов. Поэтому после включения напряжения питания излучатели оптронов O_1 и O_2 начинают возбуждаться. Однако вследствие флуктуации один из излучателей возбудится сильнее. Если первым возбудится оптрон O_2, то это приведет к гашению оптрона O_1, и триггер окажется в одном из устойчивых состояний. Подача запускающего импульса на вход *1* приведет к возбуждению оптрона O_1 и гашению излучателя оптрона O_2 — триггер переходит в другое устойчивое состояние. Подача запускающего импульса на вход *2* возвращает триггер в первоначальное состояние.

Функциональные оптоэлектронные схемы. Такие схемы могут решать большинство задач транзисторной и электровакуумной электроники, поскольку оптроны являются схемотехническими аналогами многих традиционных электрорадиокомпонентов, деталей, устройств. К простейшим функциональным схемам относятся логические схемы И, ИЛИ, НЕ. Три логические схемы рис. 9.63 составляют так называемый *полный набор логических элементов*, с помощью которого может быть реализована любая логическая функция. Если к названным трем схемам добавить оптоэлектронный триггер, то полученная система элементов будет достаточна для синтеза *произвольного конечного автомата*. На основе пленок

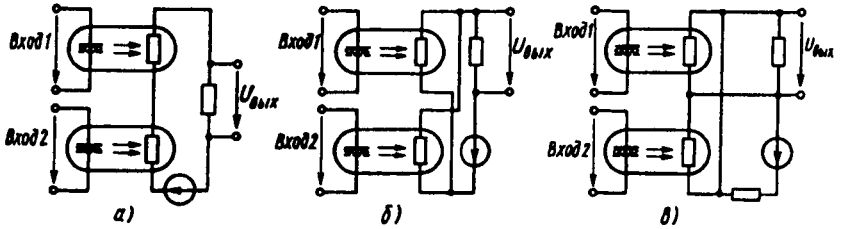


Рис. 9.63. Оптоэлектронные логические схемы:
а — схема И; *б* — схема ИЛИ; *в* — схема НЕ—ИЛИ

ZnSMn—CdS число входов в схемах может быть доведено до 10^3. Многоканальность входных цепей является одной из особенностей пленочной оптоэлектроники с использованием материалов $A^{II}B^{VI}$.

Совершенствование оптронов направлено на уменьшение рабочих и темновых токов, расширение рабочих температур, повышение быстродействия, долговечности, температурной стабильности и радиационной стойкости. Валовой объем производства оптронов сопоставим с объемом выпуска СВЧ-диодов и германиевых транзисторов.

Возможность на одних и тех же оптронах совмещать ввод и вывод информации с цифро-аналоговой и аналого-цифровой обработкой, хранением и индикацией (визуализацией) открывает новое перспективное направление функциональной оптоэлектроники, связанное с организацией однородных вычислительных сред в виде *операционных экранов* — некогерентных оптоэлектронных процессоров для обработки изображений.

Поиски нового принципа организации вычислительной системы вызваны тем, что при взаимодействии с ЭВМ человек почти 80% информации воспринимает по зрительному каналу. Визуализация информации — традиционная область оптоэлектроники. Однако общение человека с машиной затрудняет то, что образное мышление, присущее ему, имеет мало общего с законами булевой алгебры, мышлению человека присуще использование принципов картинной логики. Вместе с тем, современная элементная база СБИС не может решить проблем оперативного ввода—вывода и параллельной обработки больших массивов информации при работе с изображениями в быстро меняющихся ситуациях. Например, однокристаллический процессор (МРД721Д) обработки изображений 640×400 точек с быстродействием $5 \cdot 10^6$ оп/с и конвейерной организацией вычислений может провести обработку одного кадра только за 1,4 с.

В специализированных цифровых вычислительных системах обработка изображений основана на клеточной логике и естественном параллелизме вычислений. Быстродействие доходит до $10^{10}...10^{11}$ оп/с. Однако операции клеточной логики применяются в основном только для задач

предварительной обработки изображений и не могут решить интеллектуальные задачи. Одним из путей создания новых принципов передачи, обработки и хранения информации в реальном времени на иных принципах микроминиатюризации является формирование *оптоэлектронных логико-временных сред* [66]. Оптоэлектронные плоские и объемные экраны для обработки информации картинного типа предлагается создавать на основе индикаторов, объединяющих функции элемента оптической связи и визуализации состояния электрической цепи с его запоминанием.

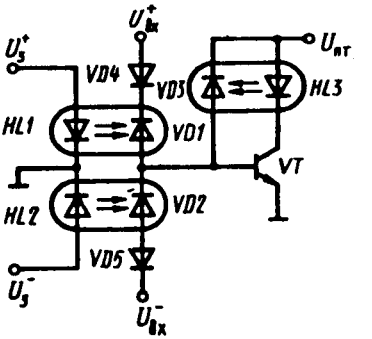


Рис. 9.64. Оптоэлектронная схема, совмещающая функции памяти и индикации

Структурный элемент однородной оптоэлектронной вычислительной среды, в которой аналоговая и цифровая обработки информации сопровождаются запоминанием и индикацией результатов без дополнительных аппаратурных затрат, изображен на рис. 9.64. Выключенное (нулевое) состояние схемы соответствует отсутствию входного сигнала, т. е. состоянию, когда ток светодиода меньше его порогового значения. Свечение выходного светодиода *HL*3 и появление светового сигнала на оптическом выходе соответствуют состоянию «включено». Светодиоды *HL*1 и *HL*2 играют роль входных; фотодиоды *VD*1 и *VD*2 предназначены для приема оптической информации, а *VD*3 для запоминания информации. Схема также содержит усилитель на транзисторе *VT* и развязывающие диоды *VD*4 и *VD*5. Оптический запуск триггера осуществляется электрическим сигналом $U_з^+$, который переводит коллекторный ток транзистора в область насыщения, и яркость свечения достигает максимального значения. Для выключения (обнуления) схемы необходимо подать электрический сигнал $U_з^-$. Тогда срабатывает оптрон *HL*2—*VD*2, малое сопротивление *VD*2 шунтирует переход эмиттер-база транзистора, выводя его из состояния насыщения в активный режим.

Структура оптоэлектронного микропроцессора предполагает соединение множества решеток однородной вычислительной среды с блоком многопрограммного управления посредством волоконно-оптических шин. Оптоэлектронный процессор включает *волограф*, представляющий собой волоконно-оптический терминал, на котором в реальном масштабе времени отображается состояние всех решеток.

§ 9.23. Интегральная оптоэлектроника

Для систем комплексной автоматизации производственных процессов и умственного труда в настоящее время требуются элементы, пропускная способность которых в тысячи раз превышает предельные возможности СБИС на транзисторах. Их пропускная способность ограничивается как технологическими пределами микроминиатюризации, так и взаимными помехами. Физическим ограничением СБИС является одномерный характер электрических сигналов, не позволяющий реализовать всех преимуществ трехмерной кристаллической среды. Использование *фотонной связи* существенно снижает помехи, дает возможность модулировать оптические сигналы не только во времени, но и в пространстве, вести параллельную обработку больших массивов информации по многим каналам. Поскольку оптоэлектроника в какой-то мере является альтернативой транзисторной электроники и в какой-то мере дополнением к ней, постольку развитие оптоэлектроники идет таким образом, чтобы цепи оптронов согласовывались с входными и выходными цепями интегральных схем, обладали технологической совместимостью с интегральными схемами, допускали микроминиатюризацию и интеграцию, а характеристики функциональных оптронных цепей не уступали достигнутым в интегральной электронике. Вследствие этого переход к интегральным оптоэлектронным парам и схемам стал основной тенденцией развития оптоэлектроники.

Оптоэлектронной интегральной схемой (ОИС) называют интегральную схему, в которой наряду с электрическими используют оптические сигналы. Принципиальная возможность построения ОИС возникла после того качественного скачка в оптоэлектронике, который был связан с разработкой в середине 60-х годов миниатюрных быстродействующих излучателей и фотоприемников на *p-n*-переходах. *Интегральная оптоэлектроника* изучает взаимные преобразования оптических и электрических сигналов в монолитных интегральных схемах.

Плотность элементов в ОИС ограничивается выделением и рассеянием теплоты. Реальные излучающие элементы имеют внутренний квантовый выход примерно 70%. Однако и при внутреннем квантовом выходе 50% только из-за отвода мощности световым потоком из зоны электрон-фотонного преобразования в ОИС достигается примерно вдвое большая плотность упаковки элементов, чем в таких же структурах с электронными связями. Потери излучения на границах в монолитных структурах значительно меньше, чем в оптоэлектронных устройствах на дискретных элементах (из-за отсутствия промежуточной среды с низким показателем преломления, вроде воздуха и стекла). Кроме того, структуры на основе арсенида галлия имеют повышенные рабочие температуры по сравнению с кремнием (473 К) и, следовательно, допускают более высокие плотно-

сти упаковки. В монолитных оптоэлектронных структурах с квантовым выходом, близким к 100%, т. е. с очень малым выделением тепловой энергии, и при размерах элементов порядка длины световой волны плотность упаковки определяется дифракционными явлениями.

Потенциальные возможности интегрально-оптических устройств позволяют их применять в волоконно-оптических линиях связи и системах оптической обработки информации. По сравнению со своими электронными аналогами и объемными оптическими устройствами их преимуществами являются большая полоса рабочих частот и возможность их разделения (мультиплексирование), малые размеры, масса, низкие управляющие напряжения, повышенная надежность и невосприимчивость к вибрациям.

Интегрально-оптические схемы могут включать в себя несколько функциональных элементов: источник излучения, фотоприемник, пассивный волновод, модулятор, устройство согласования. Интеграция элементов осуществляется либо в монолитной схеме, в которой все оптические элементы создаются на единой подложке, либо гибридным методом, при котором функциональные элементы получены на различных материалах и подложках. Преимуществом последнего метода является использование материалов, оптимальных для выполнения отдельных функций, недостатком — трудности обеспечения эффективной оптической связи различных элементов, необходимость юстировки волноводов микрометровых размеров с большой степенью точности. В процессе эксплуатации различные элементы такой системы могут рассогласоваться друг относительно друга, что приведет к отказу в работе устройства в целом.

На основе полупроводников возможно создание всех интегрально-оптических элементов. Поэтому именно они используются для построения монолитных интегральных схем, в которых оптические элементы располагаются на единой подложке. Наибольшее распространение получили волноводные устройства на основе эпитаксиальных слоев GaAlAs, GaAsP, InGaAsP. К перспективным материалам интегральной оптики относят сульфид кадмия и соединения на его основе.

В полупроводниковых соединениях оптические волноводы хорошего качества создаются путем изменения состава материала либо его электрических свойств, например, формированием высокоомной области на низкоомной подложке. Осуществляется также волноводное распространение света в обедненной области диффузионных *p-n*-переходов. Универсальной гибкостью в изготовлении структур с заданными оптическими параметрами обладают различные методы эпитаксиального выращивания тонкопленочных волноводных структур с изменением состава пленки по сравнению с составом подложки. Например, в тройных соединениях $Ga_{1-x}Al_xAs$, меняя концентрацию алюминия, значительно изменя-

ют показатель преломления $\Delta n' = -0{,}4x$ без существенного изменения постоянных решетки. Формирование оптических волноводов из высокоомных эпитаксиальных материалов позволяет полностью исключить оптические потери, обусловленные свободными носителями заряда. Полосковые оптические волноводы могут быть сформированы в процессе диффузии через окна в соответствующих масках.

Принципы электрон-фотонного преобразования. Проектирование оптимальных элементов ОИС основано на трех основных принципах: 1) сохранения объема сигнала; 2) оптимальности структуры интегрального оптоэлектронного элемента; 3) информативности оптоэлектронного преобразования.

Принцип сохранения объема сигнала — оптоэлектронное преобразование происходит без потерь информации, если в результате всех физических процессов испускания фотонов и генерации электронно-дырочных пар не изменяется или возрастает объем сигнала $V = TFL$, где T — длительность сигнала; F — ширина спектра частот; $L = \ln(P_c/P_ш)$ — динамический диапазон (P_c — мощность сигнала, $P_ш$ — мощность шума).

Этот принцип определяет возможность изменения параметров сигнала в процессе оптоэлектронного преобразования, позволяя, в частности, снижать мощность сигнала, необходимую для обработки информации, за счет накопления дискретных сигналов в течение определенного промежутка времени (например, в матричных устройства регистрации образов).

Принцип оптимальности структуры интегрального оптоэлектронного элемента основан на том, что в интегральном элементе с несколькими оптоэлектронными преобразованиями потери информации минимальны, если фотон-электронное преобразование происходит в атомной системе с меньшей энергией квантовых переходов, чем последующее электрон-фотонное преобразование, а передача оптического сигнала — в микросистеме с энергией квантовых переходов меньшей, чем его генерация.

Это означает, что фотоприемные звенья следует формировать в полупроводнике с меньшей шириной запрещенной зоны, чем излучающие. Этот принцип также выражает необратимость фотон-электронного и электронно-фотонного преобразования сигнала в одной системе атомов.

Если входной сигнал x преобразуется в некотором элементе в выходной сигнал y, то по Шеннону, пропускная информационная способность такого элемента

$$Q(y, x) = \lim_{T \to \infty} \frac{H(y) - H(y/x)}{T},$$

где $H(y)$ — энтропия входного сообщения при отсутствии связи входа и выхода; $H(y/x)$ — уменьшение энтропии выходного сообщения в резуль-

тате связи между x и y, а также изменения числа возможных состояний выходного сигнала y, зависящего от уровня шумов и их статистических свойств.

Электроны и фотоны, как носители информационных сигналов, имеют различные характеристики. Электрон обладает зарядом и может находиться в двух состояниях (связанном и свободном). Вектор скорости электрона определяется векторами электрического и магнитного полей. Фотон же не имеет заряда и движется в однородной среде с постоянной скоростью прямолинейно независимо от электромагнитных полей, которые имеются на его пути. После столкновения (взаимодействия) с микрочастицей фотон полностью передает ей свою энергию и импульс и перестает существовать. Это позволяет в каналах передачи оптических сигналов рассматривать фотоны как невзаимодействующие частицы, чего нельзя сказать про каналы электрической связи с носителями заряда. В результате электрон-фотонного преобразования сигнал представляет собой набор невзаимодействующих символов [$H(y/x) \approx 0$], тогда как наличие связей между символами при фотон-электронном преобразовании снижает энтропию [$H(y/x) > 0$].

Принцип информативности оптоэлектронного преобразования основан на том, что при равных объемах сигналов информативность полученного в результате оптоэлектронного идеального преобразования ансамбля фотонов больше информативности ансамбля электронов ($Q_{эф} > Q_{фэ}$). Потери информации в реальных преобразователях вносят дополнительные изменения в энтропию выходных сообщений.

Первый принцип основывается на законе сохранения количества информации; второй — на законе сохранения энергии импульса.

Монолитные оптроны с прямой оптической связью. Они имеют три звена: излучатель, изолирующую прозрачную среду и фотоприемник (рис. 9.65, *а*). Все звенья интегрального оптрона могут быть сформированы в одном кристалле соединения $A^{III}B^{V}$ (рис. 9.65, *б*), причем полуизолирующие слои получаются легированием (хромом и железом в случае арсенида галлия). Введение прозрачных согласующих (для кристаллографических параметров) слоев (рис. 9.65, *в*) является необходимым условием для получения качественных эпитаксиальных слоев, обеспечивающих близкое к единице теоретическое значение коэффициента передачи. Коэффициент передачи достигает 0,9, он ограничен внутренним квантовым выходом и геометрическими потерями. Использование усиления в фототранзисторе (рис. 9.65, *г*) позволяет также получить большой коэффициент передачи оптрона (за счет понижения быстродействия). В этой структуре, как и в предыдущих, изоляция источников и приемников осуществляется электрически пассивным высокоомным слоем полупроводника или диэлектрика.

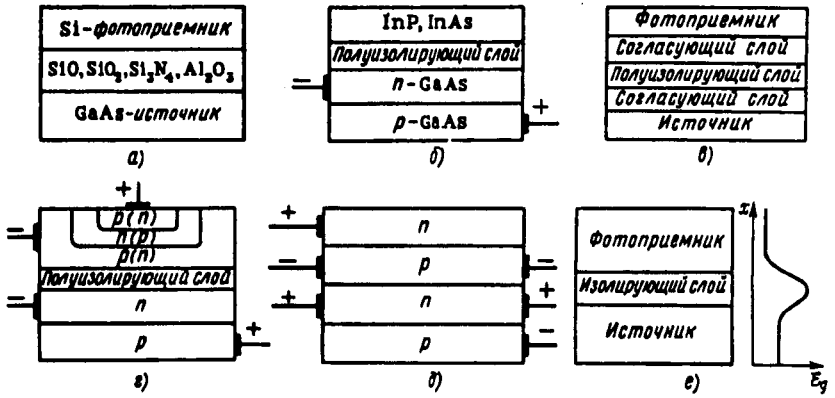


Рис. 9.65. Структуры интегральных оптронов с прямой оптической связью

Вместо изолирующего слоя для развязки входа и выхода можно применить обедненный слой обратносмещенного *p-n*-перехода (рис. 9.65, *д*) или слой с внутренним полем, возникающим в широкозонной области с высоким сопротивлением (рис. 9.65, *е*). Широкозонный слой с градиентом запрещенной зоны служит потенциальным барьером для носителей заряда, способных создавать утечку через этот слой: тепловые носители, генерированные в области источника и фотоприемника, под действием внутреннего поля как бы скатываются в более узкозонные области. В соответствии с принципом оптимальности структуры фотоприемник создается во всех вариантах оптронов на стороне пленки с меньшей шириной запрещенной зоны.

Интегральную матрицу регенеративных оптронов можно рассматривать как *универсальную однородную вычислительную среду*, осуществляющую все преобразования электрических и оптических сигналов. Многослойные полупроводниковые структуры, образующие интегральные оптоэлектронные элементы, позволяют устранить недостатки оптоэлектронных устройств на дискретных элементах: сложность системы, низкий коэффициент передачи и недостаточную надежность.

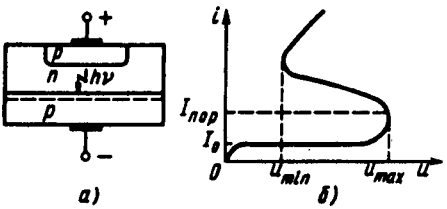


Рис. 9.66. Оптоэлектронный транзистор:

а — конструкция; *б* — вольт-амперная характеристика в лавинном режиме

К таким элементам относится оптоэлектронный транзистор в режиме генерации (рис. 9.66, *а*). Его излучающий *p-n*-

переход создается в эпитаксиальном слое германия или кремния, а коллекторный *p-n*-гетеропереход фотоприемника — в арсениде галлия или германия. При этом в коллекторе на расстоянии $d = 1$ мкм поглощается более 90% излучения, а в базе толщиной до нескольких десятков микрометров — менее 1%. Поскольку при надлежащей геометрии структуры на фотоприемник падает около половины излучаемого света, то максимальный коэффициент передачи может достигать 0,45. Прибор кроме высокой эффективности обладает большим быстродействием, поскольку передача сигнала от эмиттера к коллектору осуществляется оптически и быстродействие в основном определяется излучательной постоянной времени (в арсениде галлия $10^{-9}...10^{-10}$ с).

Регенерация происходит следующим образом: ток эмиттера вызывает в нем излучение. Это излучение,подаваемое на коллектор, создает фототок, в результате чего ток через прибор возрастает, что, в свою очередь, приводит к увеличению мощности излучения. Оптическая связь между эмиттером и коллектором возникает только в том случае, когда пороговый ток эмиттера $I_{пор}$, с которого начинается излучение, меньше коллекторного тока насыщения I_0. Однако в обычных приборах $I_{пор} > I_0$. Несогласованность токов устраняется в дискретных оптронах с помощью дополнительных усилителей. В интегральной конструкции несогласование может быть устранено путем применения механизма лавинного умножения (рис. 9.66, *б*).

Многоустойчивые оптроны можно получить либо схемным путем, либо использованием активных свойств фотоприемников, излучателей, световодов. Первый способ обладает большей гибкостью в выборе характеристик и числа устойчивых состояний оптрона, а второй — схемной простотой, большей надежностью и меньшей рассеиваемой мощностью, т. е. больше подходит для реализации в интегральном монолитном варианте. Например, многослойная фотодинисторная *p-n-p-n-p-n*-структура может работать как регенеративный оптрон, если первый *p-n*-переход является инжекционным излучателем, а следующие три перехода образуют фотодинистор. Его вольт-амперная характеристика может содержать три падающих участка, причем один из них обусловлен активными свойствами фотоприемника, а два других — действием положительной обратной связи (один из них связан с усилением фототока вследствие разделения фотоносителей на коллекторном переходе, а другой — с фотопроводимостью в базовых областях *p-n-p-n*-фотоприемника).

Оптические и электронные переходы в *полупроводниках с градиентом запрещенной зоны* позволяют их рассматривать как функциональные монолитные оптроны, являющиеся основой для различного рода сложных интегральных устройств. Их можно формировать в процессе про-

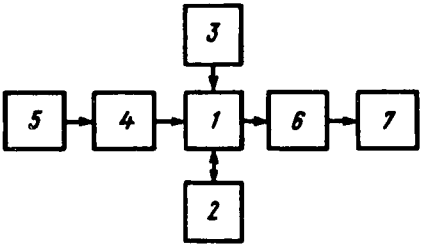


Рис. 9.67. Структурная схема звена матричного оптоэлектронного преобразователя:

1 — звено оптоэлектронного преобразователя; *2* — внутренняя память; *3* — источник энергии; *4* — формирователь управляющего сигнала; *5* — дешифратор; *6* — усилитель; *7* — преобразователь параметров выходного сигнала

граммного выращивания эпитаксиальных слоев по закону, заданному функцией преобразования сигнала.

Интегральные матричные оптоэлектронные преобразователи. Они реализуют наиболее существенное преимущество оптоэлектронных элементов — возможность беспроводной обработки и передачи информации одновременно по многим параллельным каналам (эффективность параллельной обработки информации в чисто электронных устройствах снижают одномерность электрических связей и электромагнитное взаимодействие соседних каналов). Данная возможность может быть осуществлена, если в соответствии с принципом информативности оптоэлектронного преобразования размножение каналов будет произведено в оптическом звене преобразователя, где информационный сигнал двухмерен.

Планарная технология интегральных схем позволяет создавать миниатюрные устройства с расщеплением излучения, сформированные вместе с электронными схемами управления. При этом ячейки матриц излучателей и фотоприемников могут обладать памятью. Элементы, в которых происходит оптоэлектронное преобразование, подсоединяются одним полюсом к вертикальным, а другим — к горизонтальным проводящим шинам в местах их пересечения. Отдельные звенья оптоэлектронного преобразователя (рис. 9.67) могут быть совмещены в одном элементе. Например, излучающие элементы, имеющие на вольт-амперной характеристике участок отрицательного сопротивления, позволяют создавать матрицы с внутренней памятью. Интегральная матрица фотоприемников может иметь внутреннюю дешифрацию, память и усиление сигнала. Монолитные матрицы, представляющие собой схемы большой степени интеграции, реализуются на кремнии, арсениде галлия по совмещенной технологии кремния и арсенида галлия или в виде многокристальных гибридных сборок.

Наиболее распространенными элементами матриц некогерентных источников света являются инжекционные светодиоды на арсениде и фосфиде галлия или твердых растворов арсенида—фосфида галлия, арсенида галлия—алюминия и т. д. (рис. 9.68). Число элементов в матрице может достигать 1024.

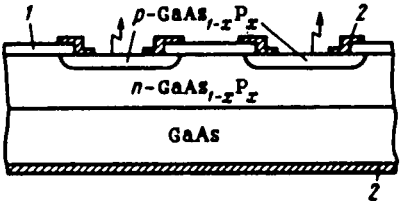


Рис. 9.68. Фрагмент планарной матрицы светодиодов:

1 — маска из оксида кремния; *2* — металлические контакты

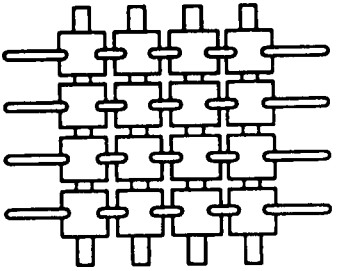


Рис. 9.69. Гибридная многокристальная матрица светодиодов с балочными выводами

Наиболее простым методом получения *матриц излучающих диодов* является метод, который совмещает планарную технологию и технологию балочных выводов, служащих контактами в *n*- и *p*-областях (рис. 9.69). Монолитные матрицы могут быть получены на изолирующем кристалле в виде системы ортогональных полос из материала *n*- и *p*-типа (светодиоды образуются в местах пересечения полос). Когда подложка для испускаемого излучения прозрачна, а выводы *p*- и *n*-областей выходят на одну и ту же поверхность пластины, сборку матриц, содержащих несколько кристаллов (чипов) , удобно производить методом обратного монтажа. Интеграция излучающих диодов в матрицу производится также с помощью различных стекол или компаундов.

Для интегральных *матриц фотоприемников* применяется преимущественно кремний. Они используются для регистрации изображения предметов (твердотельные видиконы), в системах оптической обработки информации, устройствах для определения положения движущихся предметов, для оптической многоканальной связи и локации, а также ввода и визуализации оптической информации. Созданы фотоприемные матрицы, содержащие до 800×800 элементов.

Объединение в одном кристалле большого числа ячеек обусловливает возникновение дополнительных трудностей по сравнению с дискретными фотоприемниками. К ним относятся обеспечение идентичности параметров элементов фотоматриц и сведение к минимуму гальванической и оптической связей между ними, которые возрастают при повышении степени интеграции.

Матрицы фотоприемников кроме светочувствительных содержат коммутирующие элементы, а иногда и элементы памяти. Простейшая ячейка включает фотодиод и последовательно подсоединенный конденсатор (рис. 9.70, *а*). Применение фототранзисторов упрощает структуру ячейки, поскольку коммутирующим элементом является эмиттер (рис.

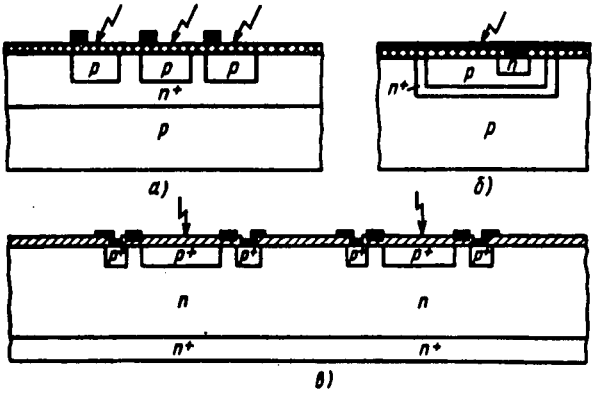


Рис. 9.70. Структура ячеек с внутренней коммутацией в матрицах с емкостью (*а*), на биполярных фототранзисторах (*б*), с МОП-транзисторами (*в*)

9.70, *б*). Высокое входное сопротивление двух МОП-транзисторов при использовании их для коммутации *p-i-n*-фотодиода (рис. 9.70, *в*) позволяет осуществить хорошую развязку по цепям коммутации и в большей мере реализовать пороговую чувствительность светочувствительных элементов.

В матрицах фотоприемников с усилением и памятью часто используются фототранзисторные структуры. В фототранзисторной МДП-структуре светочувствительным элементом является переход исток-подложка (рис. 9.71, *а*). Коммутирующим элементом служит канальный

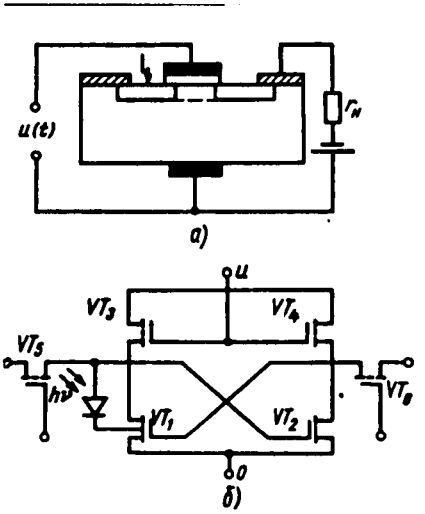


Рис. 9.71. Матрицы с усилением и памятью:

а — структура канального МДП-фототранзистора; *б* — прибор с зарядовой связью, *в* — схема фотоприемника с запоминающим триггером на МДП-транзисторах

транзистор с изолированным затвором, который при достаточном отрицательном потенциале считывания перезаряжает емкость перехода фотодиода. Фоточувствительный элемент может работать и в режиме накопления. Падающий на матрицу световой поток проходит через прозрачный оксидированный слой поверх истока транзистора, создавая в нем носители заряда, которые движутся к *p-n*-переходу и разряжают диод. За время освещения напряжение на диоде, образуемом истоком и подложкой, несколько возрастает, приближаясь к нулю. Поскольку в состоянии сброса напряжение на диоде равно напряжению источника питания, ток в нагрузочном резисторе и входное напряжение пропорциональны повышению напряжения на диоде в течение одного кадра. Следовательно, выходное напряжение оказывается пропорциональным световому потоку, падающему на исток транзистора. Изготовлена фототранзисторная матрица из 400×500 элементов с шагом 63,5 мкм по оси x и 50,6 мкм по оси y.

Матрицы с запоминанием и усилением можно выполнять и на ячейках, в которых фотодиод соединен с коммутирующим элементом и триггером на МДП-транзисторах. В схеме рис. 9.71, *б* триггер состоит из двух переключающих транзисторов VT_1 и VT_2, в цепях стоков которых в качестве резисторов включены транзисторы VT_3 и VT_4. Переключение триггера осуществляет фотодиод, подключенный параллельно транзистору VT_1. Коммутирующими элементами в схеме являются транзисторы VT_5 и VT_6. Такая матрица из 144×128 светочувствительных площадок диаметром 100 мкм была применена для считывания в постоянном голографическом запоминающем устройстве емкостью $2 \cdot 10^7$ бит.

В матричных преобразователях изображения и усилителях света может быть использована структура МДП-светодиода (рис. 9.72), в которой между прозрачными электродами *1* расположен изолирующий диэлектрический слой *2*, полупроводник (фоторезистор) *3* и излучающий *p-n*-переход *4, 5*. В накопительном режиме *p-n*-переход смещен в обратном направлении и его сопротивление составляет около 10^6 Ом. Воздействие светового потока I_1 вызывает фотогенерацию носителей, которые в сильном поле дрейфуют и накапливаются на границе раздела слоев. Выходной сигнал регистрируется в момент выключения импульса напряжения, когда в прямосмещенном *p-n*-переходе возникает ток разряда накопленного заряда и излучение I_2. Время накопления в кремниевой МДП-структуре $\tau_н \approx 10^{-3}$ с, время релаксации $\tau_{рел} \approx 10^{-7}$ с. В этом случае коэффициент трансформации тока определяет коэффициент усиления стационарного излучения $K = \tau_н / \tau_{рел} \approx 10^4$ вместо примерно 10

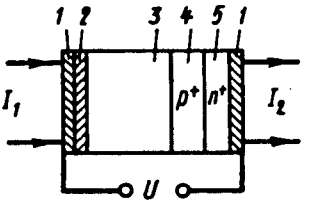


Рис. 9.72. Структура МДП-светодиода

при использовании электролюминофора (см. рис. 9.59, *б*).

Особенностью матриц на приборах с зарядовой связью (ПЗС) является возможность самосканирования. ПЗС называют МДП-приборы, в которых происходит передача заряда по поверхности диэлектрик-полупроводник. Если к затвору такой МДП-структуры (рис. 9.71, *в*) приложить отрицательное напряжение, превышающее пороговое, то у поверхности полупроводника образуется обедненная область. Попадая в эту область, инжектированные светом неосновные носители притягиваются к поверхности и могут храниться в этом узком локальном слое. Для перехода заряда под соседний электрод к нему прикладывается большее отрицательное напряжение, чем к предыдущему, в результате чего образуется продольное электрическое поле между электродами. Под действием этого напряжения и диффузии дырки переходят в более глубокую потенциальную яму под соседним электродом. После передачи заряда потенциал электрода снижается до напряжения, при котором он хранится в этой яме. Комбинации металлических и диэлектрических слоев позволяют создавать структуры с высокой степенью интеграции, на порядок превышающий уровень интеграции МДП-структур.

Приборы с ПЗС наиболее перспективны для построения интегральных фотоприемных устройств. Кроме более высокой степени интеграции, они имеют и более низкую потребляемую мощность по сравнению с другими матрицами. По сравнению с фотодиодными матрицами они обладают более высокой точностью и быстродействием. Разработаны матричные ПЗС с числом элементов 576×512. Матрицы ПЗС используются для построения телевизионных камер как в видимом, так и инфракрасном диапазоне (из силицида платины). По сравнению с телекамерами на ЭЛТ (видиконами) они более надежны, свободны от искажений, обладают лучшими массогабаритными параметрами, низким управляющим напряжением, жестким геометрическим растром, большим динамическим диапазоном, повышенным отношением сигнал/шум, позволяют определить размеры объектов с относительной погрешностью 0,1%. В действующих оптико-цифровых системах на основе ПЗС-фотоприемников реализация модуляционного метода приема обеспечила высокую чувствительность, необходимую для исследования космических источников излучения.

На основе ПЗС-фотоприемников создан новый класс систем —акустооптические устройства с интегрированием во времени. В них используется способность ПЗС накапливать сигнал, позволяя обрабатывать в реальном масштабе времени сигналы большой длительности с чрезвычайно высоким частотным разрешением. Коррелятор с ПЗС осуществляет двухмерную функцию корреляции за 10 с.

Многоэлементные ПЗС-фотоприемники, совмещающие в единой интегральной структуре как светочувствительные элементы, так и регистры сдвига сравнительно легко осуществляют параллельный прием большого

числа световых сигналов и последовательный ввод электрических сигналов с ЭВМ без уменьшения быстродействия оптических систем. Решив задачу их сопряжения с обладающими меньшим быстродействием ЭВМ, фотоприемники ПЗС открыли возможность создания гибридных оптико-цифровых устройств, сочетающих в себе преимущества как оптических процессоров, так и ЭВМ.

Применение матриц ПЗС позволяет также создать функционально богатый, компактный и простой в управлении электрически управляемый модулятор света (рис. 9.73), использующий электрооптические свойства жидкокристаллического слоя *4*. Предполагается, что считывающий свет *7* благодаря отражению от зеркала *3* дважды проходит жидкий кристалл, прозрачный электрод *5* и подложку *6*. Входной электрический сигнал *8* последовательно заполняет ячейки входного последовательного регистра *11* структуры ПЗС. Тактовые импульсы *9* обеспечивают последовательный сдвиг, формирующий строку. Тактовые импульсы *10* осуществляют параллельный сдвиг. После окончания формирования двухмерного распределения заряда в ПЗС-структуре с *12* каналами управления, расположенными в полупроводниковой пластине *1*, на электрод *5* подается управляющее напряжение и сформированный заряд переносится на границу электрооптического слоя (*2* — светопоглощающий слой). Тактовая частота последовательного ввода информации может достигать 100 МГц.

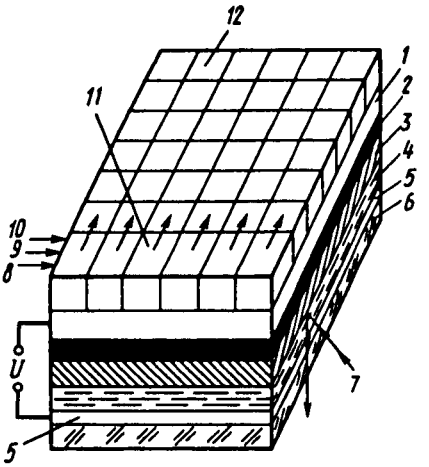


Рис. 9.73. Матрица ПЗС в электрически управляемом пространственном модуляторе света

Новым типом матричного фотоприемника является также линейная структура «Спрайт», предназначенная для считывания изображения с помощью одномерной непрерывной структуры в виде полоски полупроводника длиной $L + l$. При этом пространственное распределение интенсивности излучения $I(x)$, спроектированное на полоску L, может быть считано как электрический сигнал $u(t)$ на промежутке l, если распределение фотоносителей заряда перемещается со скоростью дрейфа во внешнем постоянном электрическом поле.

Интегральная оптика. Интегральная оптика является одним из разделов оптоэлектроники. Она позволяет осуществлять направленное распространение и преобразование электромагнитных волн в микрominиа-

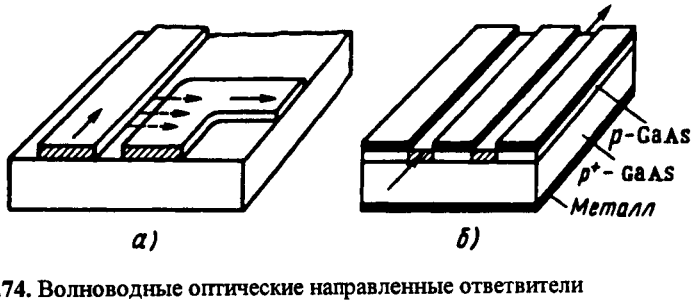


Рис. 9.74. Волноводные оптические направленные ответвители

тюрных интегральных устройствах (оптических интегральных схемах), основанных на использовании комбинации плоских диэлектрических волноводов и других тонкопленочных оптических и оптоэлектронных элементов. Ее основу составляют процессы распространения электромагнитных волн в *пленочном волноводе*. Необходимо отметить, что среди разнообразных материалов для их изготовления используются и полупроводники. Оптические потери в полупроводниковых волноводах в основном связаны с рассеянием на несовершенствах кристаллической решетки и поглощением на свободных носителях заряда. Так, в волноводе из арсенида галлия уменьшение суммарной концентрации донорных и акцепторных примесей с 10^{16} до $2 \cdot 10^{15}$ см$^{-3}$ привело к снижению потерь с 4,3 до 1,3 дБ/см (при длине волны 1,15 мкм).

Волноводная техника позволяет изготовлять основные виды пассивных оптических элементов. Так, эффект призмы (поворот светового луча) возникает от нанесенного на волновод дополнительного слоя того же вещества в форме треугольника (скорость света в волноводе зависит от его толщины). Эффект линзы возникнет, если происходит изменение показателя преломления по сечению волновода или на части волновода создается расширение полукруглой формы. Эффект зеркала (отражения) достигается путем образования на одной из граней световода дифракционной решетки, если расстояние между штрихами удовлетворяет условию Брэгга $l_{\text{Б}} = m\Lambda/2$, где m — порядок дифракции.

На основе диэлектрических волноводов возможна также реализация оптических аналогов большинства элементов техники СВЧ: направленных ответвителей, фильтров, переключателей, фазовращателей, гираторов, вентилей. В *направленном ответвителе* может быть использована связь волноводов, расположенных близко друг от друга (рис. 9.74, *а*).

Если в одном из волноводов возбуждена собственная мода $\dot{E}_1(z) = A_1 e^{-j\beta_1 z}$, то поле этой моды, экспоненциально затухающее во внешней области, достигает второго волновода и возбуждает в нем соб-

ственную моду $\dot{E}_2(z) = A_2 e^{-j\beta_2 z}$. Связь между волнами с достаточной точностью описывают уравнения

$$\frac{\partial \dot{E}_1}{\partial z} = -j\beta_1 \dot{E}_1 + jc_1 \dot{E}_2;$$

$$\frac{\partial \dot{E}_2}{\partial z} = -j\beta_2 \dot{E}_2 + jc_2 \dot{E}_1,$$

где c_1, c_2 — постоянные коэффициенты.

Для одинаковых волноводов ($\beta_1 = \beta_2 = \beta$) при $E_2(0) = 0$ решение системы уравнений связанных волн имеет вид

$$\dot{E}_1(z) = A_1 e^{-j\beta z} \cos kz;$$

$$\dot{E}_2(z) = jA_1 e^{-j\beta z} \sin kz,$$

где $k = \sqrt{c_1 c_2}$. Следовательно, происходит перекачка энергии излучения из одного волновода в другой с периодом $L = \pi/(2k)$. В частности, энергия полностью сосредотачивается во втором волноводе и отсутствует в первом в точках $kz = \pi/2 + m\pi$, где $m = 0, 1, 2, \ldots$.

Если коэффициенты распространения у обоих волноводов одинаковы, то возможна полная перекачка мощности из одного волновода в другой; если же эти постоянные неодинаковы, то происходит частичная перекачка. Управлять связью между волноводами можно и с помощью электрического поля. Например, при приложении обратного напряжения к барьеру Шотки со стороны одного из волноводов (рис. 9.74, *б*) осуществлено 95%-ное переключение мощности из одного канала в другой. Создание тонкопленочных оптических аналогов гиратора, циркулятора, вентиля и т. п. возможно на основе гиротропных сред с резко выраженными электрооптическими и магнитооптическими эффектами.

Ввод излучения в плоские оптические волноводы может быть обеспечен не только с помощью призмы (см. рис. 3.31), но и с помощью *дифракционного ввода* (рис. 9.75, *а*), конструктивно значительно более удобного. Для этой цели используется также клинообразный конец пленки (рис. 9.75, *б*), где происходит преобразование волноводной моды в моду подложки (и наоборот). Сопряжение многомодового оптического

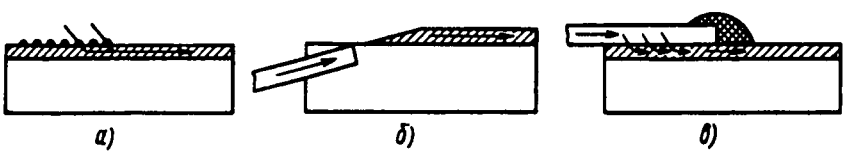


Рис. 9.75. Ввод излучения в оптический волновод

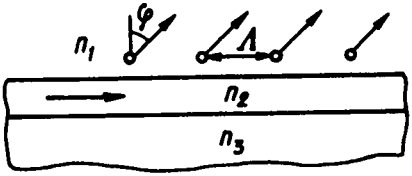


Рис. 9.76. Схема решеточного вывода излучения

волокна с многомодовым плоским волноводом с эффективностью до 70% достигается их соединением с помощью фоторезиста (рис. 9.75, *в*).

Решеточные и призменные устройства связи обладают обратимостью, т. е. могут быть применены не только для ввода, но и для вывода излучения из волновода. Решетку из фоторезистивного слоя или созданную периодическим изменением толщины волновода можно рассматривать как решетку излучателей с шагом Λ, находящуюся в поле собственной волны волновода (рис. 9.76). Максимумы диаграммы направленности формируются в направлении, в котором сигналы от всех излучателей находятся в фазе. Требование синфазности лучей от двух соседних излучателей приводит к условию

$$\frac{2\pi}{\lambda} n_1 \Lambda \sin\varphi - \frac{2\pi}{\lambda}\gamma\Lambda = 2\pi m,$$

где γ — коэффициент замедления для волноводной моды; $m = 0, \pm1, \pm2, \ldots$ — порядок дифракции. Отсюда можно определить значения резонансного угла φ_m заданной моды волновода. Например, при $n_1 = 1$

$$\sin\varphi_m = \gamma + m\frac{\lambda}{\Lambda}.$$

Зависимость коэффициента замедления от номера моды позволяет с помощью решеточного устройства связи осуществлять селекцию типов колебаний на выходе волновода или возбудить заданную моду на ее входе.

Волноводная техника используется во многих активных элементах оптоэлектронных интегральных схем: источниках излучения, фотоприемниках, модуляторах и т. д. В представленном на рис. 9.77, *а фотоприемнике* волновод выполнен на основе пленки фоторезиста, расположенной поверх оптически менее плотной пленки SiO_2, столь тонкой, что подложкой волновода становится кремний, показатель преломления которого значительно больше, чем у фоторезиста. Поэтому в активной области волноводные моды преобразуются в моды подложки, т. е. световая энергия из фоторезиста

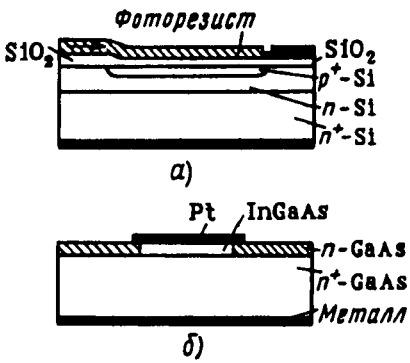


Рис. 9.77. Волноводные фотоприемники

переходит в кремний. При обратном смещении p^+-n-n^+-структуры обедненная область расширяется на всю толщу эпитаксиальной пленки, что обеспечивает эффективное собирание генерируемых носителей p^+-n-переходом. В фотоприемнике на рис. 9.77, *б* плоский световод из легированного GaAs непосредственно переходит в сильно поглощающий участок InGaAs. Собирание фотоносителей осуществляет барьер Шотки, образованный платиновым электродом, нанесенным на поверхность полупроводника.

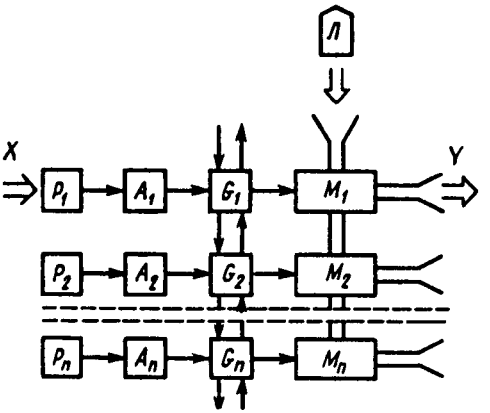


Рис. 9.78. Функциональная схема строки многоканального модулятора

В тонкопленочных *оптических модуляторах* используется большинство пригодных для этой цели физических эффектов. Из электрооптических материалов ниобат и танталат лития интересны тем, что на их основе в едином технологическом процессе можно изготовлять два вида элементов оптических интегральных схем — модуляторы и волноводы. Преимуществом тонкопленочной конструкции магнитооптических модуляторов является возможность расширения диапазона рабочих частот по сравнению с объемными приборами. Например, диапазон рабочих частот феррит-гранатовых модуляторов простирается до 10^8 Гц, при этом управляющие напряжения и потребляемая мощность на порядок меньше, чем у лучших образцов электрооптических модуляторов.

К перспективным методам модуляции в интегральной оптике относится использование линейного электрооптического эффекта в p-n-переходах. Он возникает в p-i-n-структурах и в области барьера Шотки при обратном напряжении, когда в области объемного заряда изменяется концентрация свободных носителей заряда, что ведет к модуляции диэлектрической проницаемости и показателя преломления.

Применение интегральной оптики для построения оптически управляемых пространственно-временны́х модуляторов света позволяет поднять их быстродействие, снизить рассеиваемую мощность и уменьшить габариты. Структурная схема многоканального модулятора (рис. 9.78) включает фотоприемники P, усилители A, логические элементы G и волноводные ответвители с модуляторами M. На вход поступают управляющие световые сигналы x. Считывающее излучение от внешнего источника $Л$ равномерно распределяется одномодовым волноводом элементов

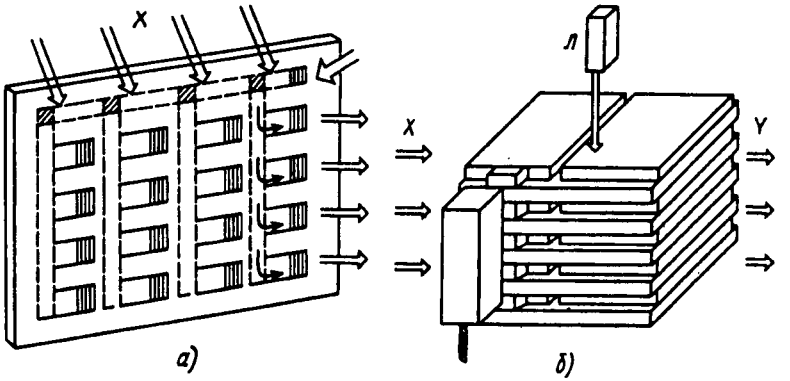


Рис. 9.79. Двухмерный интегрально-оптический модулятор в интегральном (*а*) и гибридном (*б*) вариантах

строки. Интегральный вариант модулятора предполагает размещение всех строк в одной плоскости подложки (рис. 9.79, *а*), с одной стороны которой размещены фотодиоды, а с другой — пленочные волноводы с модуляторами. Двухмерный модулятор можно образовать и набором строк (рис. 9.79, *б*). Здесь выходное изображение *Y* образуется путем вывода излучения через торцы волноводов, что удобно для сопряжения с входом другого модулятора без промежуточных оптических элементов.

Возможность изменения показателя преломления путем облучения квантами с энергией, превышающей ширину запрещенной зоны, позволяет создавать принципиально новые типы интегрально-оптических модуляторов и дефлекторов света, управляемых оптическим сигналом.

Изменение показателя преломления происходит в результате созданных светом электронно-дырочных пар и увеличения концентрации неравновесных носителей заряда. Существенное снижение управляющей мощности достигается использованием интерферометра Маха—Цендера на волноводах CdS_xSe_{1-x}. Предполагается, что использование светодиодов GaP в качестве источника модулирующего света приведет к созданию компактного устройства с малой массой.

В интегрально-оптической схеме (рис. 9.80) фотодетекторы *7* формируются в одномодовом оптическом волноводе *4*, отделенном от подложки *6* из кремния слоем оксида кремния *5*. Считывание информации с оптического диска *3* осуществляет луч полупроводникового лазера *12*. Вывод луча из планарного волновода производят с помощью решеточного элемента связи, образующего дифракционную линзу *2*. Отраженный луч коллимируется той же линзой в волновод и дифракционным расщепителем, образованным двумя дифракционными линзами *1*. Сигналы с отдельных секций фотодиодов после усилителей *11* направляются в ка-

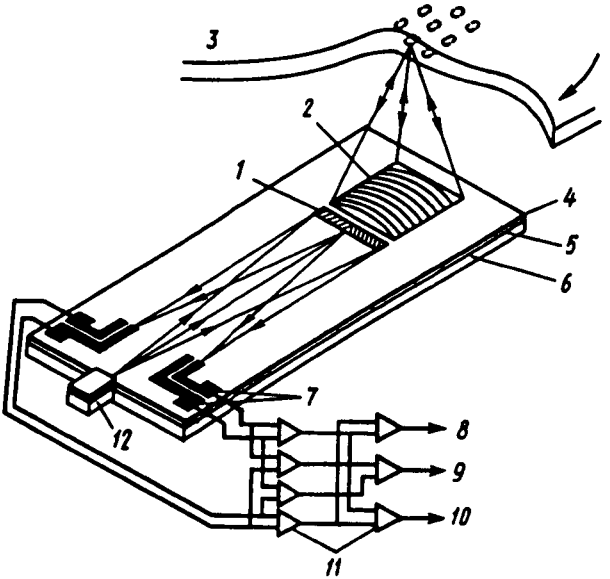


Рис. 9.80. Оптическая головка проигрывателя

налы демодуляции записанной информации *8*, системы автофокусировки *9* и автотренинга *10*.

В малогабаритном встречно-поточном *процессоре* (рис. 9.81) система электродов электрооптического модулятора-дефлектора *5* реализует умножение матрицы на вектор в планарном оптическом волноводе *1*. Элементы матрицы хранятся в цифровой памяти *8*, а компоненты вектора поступают в виде входных сигналов через регистр сдвига *9* и цифро-аналоговые преобразователи *7*. Выходной сигнал формируется инжекционным лазером *6* и, проходя через волноводные линзы *4*, модулятор *5*, диафрагму *3*, снимется с линейки фотодетекторов *2*.

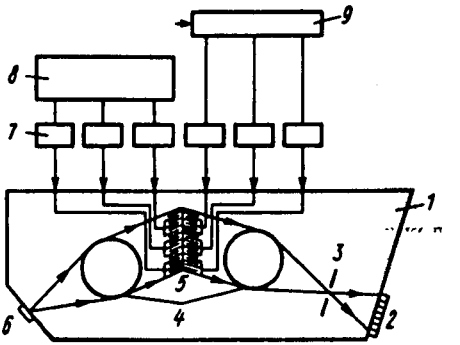


Рис. 9.81. Интегральная схема умножения матрицы на вектор

Анализ и синтез оптоэлектронных устройств. Анализ и синтез таких устройств на интегральных элементах можно свести к задачам обычных электронных цепей, используя теорему о том, что опто-

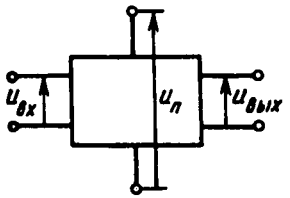


Рис. 9.82. Эквивалентная схема оптрона

электронная цепь эквивалентна изоморфной (имеющей ту же электрическую структуру) электронной цепи, в которой каждый светоизлучатель заменен его сопротивлением, а каждый фотоприемник — либо системой параллельно расположенных генераторов тока, зашунтированных сопротивлением фотоприемника, либо системой последовательно расположенных генераторов напряжения, нагруженных сопротивлением фотоприемника.

Анализ оптоэлектронных цепей сводится к построению эквивалентных электрических схем и их рассмотрению. При этом оптические характеристики цепей связаны через соответствующие коэффициенты квантовой эффективности (коэффициенты оптоэлектронного преобразования). Если эти коэффициенты постоянны во всем диапазоне изменения амплитуды и частотного спектра сигнала, элементарный оптрон можно рассматривать в общем случае как линейный шестиполюсник (рис. 9.82) и описывать его свойства Z-, Y-, или H- матрицами. Матрица проводимостей имеет вид

$$\begin{bmatrix} Y_{11} & 0 & Y_{13} & 0 & 0 & 0 \\ Y_{21} & Y_{22} & Y_{23} & Y_{24} & Y_{25} & Y_{26} \\ Y_{31} & 0 & Y_{33} & 0 & 0 & 0 \\ Y_{41} & Y_{42} & Y_{43} & Y_{44} & Y_{45} & Y_{46} \\ 0 & 0 & 0 & 0 & Y_{55} & Y_{56} \\ 0 & 0 & 0 & 0 & Y_{65} & Y_{66} \end{bmatrix},$$

или с учетом равенства токов $I_1 = -I_3$, $I_2 = -I_4$, $I_5 = -I_6$

$$\left[\begin{array}{cc|cc|cc} Y_{11} & -Y_{11} & 0 & 0 & 0 & 0 \\ -Y_{11} & Y_{11} & 0 & 0 & 0 & 0 \\ \hline Y_{21} & -Y_{22} & Y_{22} & -Y_{22} & Y_{25} & -Y_{25} \\ -Y_{21} & -Y_{21} & -Y_{22} & Y_{22} & -Y_{25} & Y_{25} \\ \hline 0 & 0 & 0 & 0 & Y_{55} & -Y_{55} \\ 0 & 0 & 0 & 0 & -Y_{55} & Y_{55} \end{array}\right].$$

Суммы строк и столбцов полученных подматриц тождественно равны нулю, что является прямым следствием гальванической развязки между входом и выходом оптрона с прямой оптической связью.

Если в оптроне канал управления по оптической передающей среде отсутствует, его эквивалентная схема сводится к трехполюснику, для которого матрица проводимости

$$\begin{bmatrix} Y_{11} & 0 \\ Y_{21} & Y_{22} \end{bmatrix},$$

а проводимость обратной связи $Y_{12} = 0$ обусловлена направленностью оптической связи от излучателя к фотоприемнику (отсутствие обратной реакции).

Анализ нелинейных оптоэлектронных цепей осуществляется переходом к эквивалентной электрической цепи путем исключения оптических характеристик из полной системы уравнений. Нахождение аппаратной функции системы по известным характеристикам ее элементов (анализ) или определение передаточных функций, функций элементов и связей (состава и конфигурации цепи) по данной аппаратной функции (синтез) решается методами теории нелинейных электрических цепей.

В нелинейных цепях широкое распространение получают переключательные элементы вычислительной техники. Для реализации оптоэлектронного преобразования сигналов в системах цифровых вычислительных машин оказывается достаточным рассмотренный набор интегральных базовых элементов: матрицы излучателей и фотоприемников, оптронный ключ, элемент цифровой индикации, матрица отображения, усилитель света, матрица фотоприемников с внутренней памятью. Они позволяют формировать сигналы, проводить логические операции и запоминать информацию.

При *синтезе* оптоэлектронной цепи сначала по заданной аппаратной функции синтезируется эквивалентная электронная схема (с учетом различных свойств оптических связей), а затем эквивалентные элементы заменяются соответствующими излучателями, световодами и фотоприемниками.

Электромагнитные явления в полупроводниках, рассмотренные в данной главе, относились к области молекулярной электроники и электролюминесценции на основе спонтанного излучения. Индуцированное излучение составляет основу квантовой электроники и рассматривается в гл. 11, посвященной нелинейным электромагнитным явлениям в активных средах квантовых усилителей и генераторов.

Глава 10
Электромагнитные явления в сверхпроводниках

Сверхпроводимость — состояние вещества, для которого характерно отсутствие сопротивления электрическому току и магнитного поля в толще сверхпроводника.

Сверхпроводимость возникает при уменьшении температуры ниже определенного для каждого материала значения, называемого *температурой (точкой) сверхпроводящего перехода* или *критической температурой* $T_{кр}$. Критическая температура зависит от напряженности внешнего магнитного поля, давления, структуры кристаллов и сплавов, технологии изготовления. Сверхпроводимость отсутствует в чистых металлах, имеющих атомный магнитный порядок (Co, Fe, Ni, Gd и др.); атомный магнетизм и сверхпроводимость являются исключающими друг друга явлениями. Сверхпроводимость обнаружена в полимере $(SN)_x$, в составе которого отсутствуют атомы металлов. Высокой критической температурой обладают сплавы и соединения на основе переходных металлов. В интерметаллическом соединении Nb_3Ge критическая температура составляет 23,2 К, в керамике Y-Ba-Cu-O — 93 К. Появление материалов, обладающих сверхпроводящими свойствами при температуре кипения жидкого азота должно привести к резкому увеличению использования сверхпроводников в технике.

При температуре выше критической сопротивление становится конечным, т. е. сверхпроводник переходит в *нормальное состояние*.

§ 10.1. Электрон-фононное взаимодействие и куперовские пары

Сверхпроводимость есть результат такого взаимодействия электрона с решеткой кристалла, при котором возникает *межэлектронное притяжение*. Формально это означает, что в законе Кулона для силы взаимодействия между двумя одноименными зарядами диэлектрическая проницаемость становится отрицательной. Аналогичное явление возникает в плазме при определенных частотах. В сверхпроводниках диэлектрическая проницаемость, связанная с колебаниями решетки (фононным газом), также оказывается отрицательной.

Притяжение между электронами наиболее сильно, если они обладают равными по значению и противоположными по направлению импульсами и спинами. Поэтому такая электронная система состоит из пар связанных электронов с суммарным импульсом и спином, равными нулю. Эти пары называют *куперовскими* по имени Л. Купера, показавшего

(1956 г.), что подобное спаривание электронов ведет к уменьшению общей энергии, а основное состояние нормального металла является неустойчивым относительно сколь угодно слабого притяжения на поверхности Ферми. Изъятие одной куперовской пары из общей совокупности сопровождается ее разрушением, для чего необходима определенная энергия — *энергия связи пары*.

В электрическом поле куперовские пары ускоряются, т. е. получают импульс, причем для всех пар одинаковый. Куперовские пары не могут отдавать энергию решетке до тех пор, пока сообщенная электронам энергия меньше энергии связи (иначе пара перешла бы в другое состояние, что исключается). Это означает беспрепятственный перенос заряда сквозь кристаллическую решетку (без сопротивления). Отсутствие сопротивления в сверхпроводниках обусловлено стабильностью квантовомеханического состояния куперовских пар.

Многочисленные экспериментальные данные свидетельствуют о том, что куперовские пары находятся в одинаковом квантовом состоянии, описываемом одной волновой функцией. Замечательные свойства сверхпроводников обусловлены именно таким *макроскопическим* заполнением единственного квантовомеханического состояния, т. е. все пары соответствуют друг другу по всем физическим параметрам.

Объединение электронов (фермионов) в куперовские пары означает появление новых частиц (бозонов) с суммарным спином, равным не половине, а нулю. Можно показать, что в соответствии с действующей для них статистикой Бозе—Эйнштейна бозоны имеют тенденцию к тем большему заполнению своего уровня, чем больше он заполнен («принцип анти-Паули»). Если уровень уже занят несколькими частицами, то это способствует переходу в это состояние других частиц и т. д.; произойдет нечто вроде конденсации в такое состояние. Сверхпроводящий конденсат электронов обладает в высшей степени упорядоченной структурой, соответствующей нулевой энтропии.

Равенство импульса и энергии всех куперовских пар свидетельствует также о равенстве длины волны всех квантовомеханических частиц $\lambda = h/p$ и частоты. Существенным следствием равенства волновых функций куперовских пар является то, что для них фиксированной оказывается и *фаза*. Фаза квантовомеханической волны, соответствующей определенному состоянию, может быть задана тем точнее, чем больше заполнено это состояние. Поскольку в одинаковом состоянии находится большое число куперовских пар, то такое состояние можно описать волновой функцией с вполне определенной фазой. Описание совокупности куперовских пар одной волновой функцией с жесткой фазовой корреляцией определяет их фазовую когерентность. Постоянство импульсов и волновых функций куперовских пар во всем объеме сверхпроводника, их

корреляция на больших расстояниях создают представление о сверхпроводнике как о гигантской молекуле или квантовой системе в макроскопическом масштабе.

Одним из следствий фазовой когерентности является квантование такой макроскопической величины, которой является магнитный поток.

Плотность сверхпроводящих электронов, объединенных в куперовские пары, максимальна при абсолютном нуле; по мере увеличения температуры она уменьшается, становясь равной нулю при критической температуре. При изменении температуры в этом интервале происходят разрушение или образование куперовских пар, обмен энергией между ними и неспаренными электронами.

Если два электрона в вырожденной электронной системе (электронный газ в металле), находящейся вблизи поверхности Ферми при температурах $T < T_{кр}$, притягиваются друг к другу, то их энергия *уменьшается* на некоторое значение Δ. При этом система переходит в сверхпроводящее состояние, для разрушения которого каждой паре электронов нужно сообщить энергию Δ. Энергия спаривания (связи) Δ является той минимальной энергией, которая требуется для разрушения куперовской пары, и может быть названа *энергетической щелью*.

Тепловое движение кристаллической решетки может разрушить сверхпроводящее состояние, если тепловая энергия превосходит энергию спаривания. При температуре, отличной от абсолютного нуля, тепловое движение приводит к всевозрастающему возбуждению электронной системы, ослабляющему притяжение между электронами. При температуре T энергетическая щель $\Delta(T) \approx \Delta(0)\sqrt{1-T/T_{кр}}$.

Разрушить сверхпроводящее состояние может также увеличение в электрическом поле импульса куперовских пар выше критического значения, при котором прирост кинетической энергии станет равным энергии связи Δ. При дальнейшем увеличении импульса, а, следовательно, и электрического тока, пара разрушается и вновь появляется диссипативное взаимодействие с металлом. Следовательно, существование энергетической щели и критического импульса для системы куперовских пар означает существование *критической плотности тока* для сверхпроводника в целом.

Предположение, что энергетическая щель определяет сверхпроводящие свойства металлов, было высказано впервые в 1935 г. Ф. Лондоном. Первым экспериментальным подтверждением гипотезы о существовании щели были опыты Г. Лондона (1940 г.), в которых было обнаружено, что при температурах ниже критических поглощение электромагнитного излучения отсутствует для частот ниже 10^9 Гц (раньше было известно, что в инфракрасном диапазоне нет различия в поглощении нормального и сверхпроводящего состояний). Исчезновение поглощения для

частот $\nu < \Delta/h$ позволяет изготовлять объемные резонаторы из сверхпроводников с добротностью, достигающей 10^{11} (для частот порядка 10 ГГц). С помощью резонатора, использованного для стабилизации мазера, при температуре 0,8 К была получена стабильность частоты 10^{-17} на временно́м интервале 100 с.

Энергетическая щель определяет разброс импульсов электронов, вовлекаемых в образование куперовских пар. Так, при нулевой температуре

$$\Delta = \delta\left(\frac{p^2}{2m}\right) = \frac{p}{m}\delta p = v_F \delta p,$$

где p, v_F — импульс и скорость электрона у поверхности Ферми (учтено, что всегда $\Delta \ll \mathscr{E}_F$). Поскольку $\delta p \approx \Delta/v_F$ и согласно соотношению неопределенностей $\delta p \delta x \geqslant \hbar$, то

$$\delta x \approx \hbar v_F/\Delta = \xi_0. \qquad (10.1)$$

Минимальная протяженность ξ_0 волновой функции и куперовской пары характеризует степень пространственной упорядоченности сверхпроводящего состояния и носит название *длины когерентности* чистого сверхпроводника.

В непереходных металлах скорость электронов на поверхности Ферми составляет около 10^8 см/с, а длина когерентности имеет порядок 10^{-4} см. В переходных металлах и интерметаллических соединениях типа Nb_3Sn и V_3Ga эффективная масса электрона очень велика, фермиевская скорость мала (порядка 10^6 см/с), как и длина когерентности (порядка $5\cdot10^{-7}$ см). В материалах с примесями и сверхпроводящих сплавах, в которых длина свободного пробега l электронов вследствие неупорядоченности структуры мала, длина когерентности ξ меньше ξ_0: $\xi^{-1} = \xi_0^{-1} + l^{-1}$.

Концентрация всех электронов проводимости $n \approx 10^{23}$ см$^{-3}$, энергия уровня Ферми $\mathscr{E}_F \approx mv_F^2/2 \approx 10^{-18}$ Дж, концентрация куперовских пар $n_s \approx n\Delta/\mathscr{E}_F \approx 10^{19}$ см$^{-3}$. Поэтому в пределах одной куперовской пары, т. е. в объеме $\xi_0^3 \approx 10^{-12}$ см^3, находится примерно $n_s\xi_0^3 \approx 10^7$ других куперовских пар. Поскольку в этом объеме волновая функция каждой пары отлична от нуля, то волновые функции различных куперовских пар сильно перекрыты.

В результате куперовские пары образуют как бы совокупность сильно взаимодействующих синхронизированных осцилляторов с одинаковыми частотами. Обусловленная этим высокая степень корреляции определяет сильно выраженную способность к «конденсации», а также то, что спаривание электронов является коллективным эффектом. Зависимость

длины когерентности от температуры для чистых сверхпроводников выражает эмпирическая формула

$$\xi(T) \approx \xi_0 (1 - T/T_{кр})^{-1/2}.$$

§ 10.2. Эффект Мейсснера

В 1911 г. Х. Камерлинг-Оннес обнаружил, что при некотором значении магнитного поля $H_{кр}$ сверхпроводник переходит в нормальное состояние. В 1917 г. Ф. Сильсби постулировал правило, согласно которому разрушение сверхпроводимости происходит при достижении магнитным полем *критического значения* $H_{кр}$ на поверхности сверхпроводника. Критическое поле зависит от температуры: $H_{кр} \approx H_0 (1 - T^2/T_{кр}^2)$, где H_0 — критическая напряженность поля при нулевой температуре.

В 1933 г. В. Мейсснер и Р. Оксенфельд установили, что при переходе образца в сверхпроводящее состояние магнитное поле из него выталкивается и магнитная индукция внутри массивного сверхпроводника равна нулю. Это явление получило название эффекта Мейсснера. Объяснение выталкивания магнитного поля на основе стационарного экранирующего тока было дано в 1935 г. в феноменологической теории братьев Ф. и Г. Лондонов.

Исходными в этой теории служат выражение для плотности электронного тока $\mathbf{J} = ne\mathbf{v}$ и уравнение движения электрона в электрическом поле $m\partial\mathbf{v}/\partial t = e\mathbf{E}$ ($\mathbf{v}$ — скорость электрона; n — число электронов в единице объема). В соответствии с этим уравнением

$$\frac{\partial \mathbf{J}}{\partial t} = ne\frac{\partial \mathbf{v}}{\partial t} = \frac{ne^2}{m}\mathbf{E}, \tag{10.2}$$

а по второму уравнению Максвелла

$$\frac{m}{ne^2}\operatorname{rot}\frac{\partial \mathbf{J}}{\partial t} + \frac{\partial \mathbf{B}}{\partial t} = 0. \tag{10.3}$$

Интегрируя уравнение (10.3) по времени и предполагая постоянную интегрирования равной нулю, получим

$$\frac{m}{ne^2}\operatorname{rot}\mathbf{J} + \mathbf{B} = \frac{m}{ne^2}\operatorname{rot}\mathbf{J} + \operatorname{rot}\mathbf{A} = 0. \tag{10.4}$$

Пренебрегая токами смещения и подставляя в первое уравнение Максвелла $\operatorname{rot}\mathbf{H} = \mathbf{J}$ уравнение (10.4), найдем

$$\mu_0 \operatorname{rot}\mathbf{J} = \operatorname{rot}\operatorname{rot}\mathbf{B} \equiv \operatorname{grad}\operatorname{div}\mathbf{B} - \nabla^2\mathbf{B} \equiv -\nabla^2\mathbf{B} = -\frac{\mu_0 ne^2}{m}\mathbf{B} \tag{10.5}$$

или

$$\mu_0 \operatorname{rot}\mathbf{H} = -\frac{m}{ne^2}\operatorname{rot}\operatorname{rot}\mathbf{J} \equiv$$

$$\equiv -\frac{m}{ne^2}(\operatorname{grad}\operatorname{div}\mathbf{J} - \nabla^2\mathbf{J}) \equiv \frac{m}{ne^2}\nabla^2\mathbf{J} = \mu_0\mathbf{J}, \qquad (10.6)$$

т. е.

$$\nabla^2\mathbf{B} = B/\lambda_L^2, \quad \nabla^2\mathbf{J} = \mathbf{J}/\lambda_L^2, \qquad (10.7)$$

где

$$\lambda_L = \sqrt{m/(\mu_0 ne^2)}. \qquad (10.8)$$

Первоначально в теории Ф. и Г. Лондонов предполагалось, что сверхпроводящий ток переносится отдельными электронами. Последующие эксперименты показали, что свойства сверхпроводящего состояния определяются куперовскими парами массой $m_s = 2m$, зарядом $q_s = 2e$ и концентрацией $n_s = n/2$. При этом величина λ_L не меняется: $m/(ne^2) = m_s/(n_s q_s^2)$. Для сверхпроводника, занимающего полупространство $x > 0$, решение уравнений (10.7) приводит к выражению

$$\mathbf{B}(x) = \mathbf{B}(0)e^{-x/\lambda_L}, \quad \mathbf{J}(x) = \mathbf{J}(0)e^{-x/\lambda_L}, \qquad (10.9)$$

т. е. индукция и плотность тока экспоненциально убывают в глубь сверхпроводника.

В выражении (10.8) λ_L имеет размерность длины и носит название *лондоновской глубины проникновения*. Для обычных плотностей электронов в сверхпроводниках она имеет порядок 10^{-6} см. Таким образом теория Лондонов объясняет, почему токи в сверхпроводниках могут проходить лишь в тонком *поверхностном слое*, а в толще сверхпроводника результирующее магнитное поле практически отсутствует (B = 0). Поле поверхностных токов компенсирует внутри сверхпроводника внешнее магнитное поле (намагниченность M = 0), а вне образца магнитное поле складывается из поля внешнего источника и поля, создаваемого поверхностными токами.

Эффект Мейсснера формально можно описать, исходя из отличной от нуля намагниченности и предположения, что поверхностные (размагничивающие) токи отсутствуют. Намагниченность $M = -H_i$ определяется из условия $B_i = \mu_0(H_i + M) = 0$. Таким образом, сверхпроводник здесь рассматривается как идеальный диамагнетик с магнитной восприимчивостью $\chi_м = -1$. При этом магнитное поле вне образца складывается из внешнего поля и поля, обусловленного намагниченностью образца.

В этом случае сравнительно просто могут быть получены некоторые термодинамические соотношения, если напряженность поля внутри образца H_i будет равна напряженности внешнего поля H. Например, из условия, что $H = H_{кр}$, переход из сверхпроводящего в нормальное состояние происходит при постоянных температуре и обобщенной силе, термодинамические потенциалы состояний (фаз) равны

$$g_s(H_{кр}, T) = g_n(H_{кр}, T), \tag{10.10}$$

причем в общем случае $\mathrm{d}g = -s\,\mathrm{d}T - \mu_0 M\,\mathrm{d}H$, где s — энтропия. При постоянной температуре $\mathrm{d}g = -\mu_0 M\,\mathrm{d}H$. В поле с напряженностью H термодинамический потенциал (или *свободная энергия Гиббса*) сверхпроводящей фазы

$$g_s(H,T) = g_s(H_{кр},T) - \mu_0 \int_{H_{кр}}^{H} M\,\mathrm{d}H = g_s(H_{кр},T) - \frac{\mu_0 H_{кр}^2}{2} + \frac{\mu_0 H^2}{2},$$

и с учетом (10.10) $g_n(H_{кр},T) - g_s(H,T) = \mu_0 H_{кр}^2/2 - \mu_0 H^2/2$.

Поскольку в нормальном состоянии намагниченность практически равна нулю, $\mathrm{d}g_n = 0$ и $g_n(H_{кр},T) = g_n(H,T)$, то из последнего соотношения

$$g_n(H,T) - g_s(H,T) = \mu_0 H_{кр}^2/2 - \mu_0 H^2/2,$$

а при $H = 0$

$$g_n(0,T) - g_s(0,T) = \frac{\mu_0 H_{кр}^2}{2}. \tag{10.11}$$

Следовательно, рассмотрение магнитных свойств сверхпроводников также приводит к выводу о том, что при температурах ниже критической сверхпроводящее состояние оказывается более энергетически выгодным и устойчивым, чем нормальное.

Описание магнитных свойств сверхпроводников на основе их намагниченности более удобно и чаще применяется, чем на основе размагничивающего действия поверхностных токов, несмотря на то, что последний способ позволяет иногда более детально исследовать физическую природу рассматриваемых явлений.

Экспериментальные данные показывают, что зависимость глубины проникновения от температуры хорошо согласуется с феноменологической формулой

$$\lambda_L(T) = \lambda_L(0)(1 - T^4/T_{кр}^4)^{-1/2}.$$

Возрастание глубины проникновения по мере приближения к критической температуре позволяет ожидать изменения свойств

сверхпроводящих образцов небольших размеров по сравнению с массивными. Если размеры образца (толщина пленки или диаметр проволоки) соизмеримы с глубиной проникновения, то область, в которую проникает магнитное поле, занимает относительно больший объем, чем в массивном. Поэтому интеграл $\int_{H_{кр}}^{H} M\,\mathrm{d}H$ в

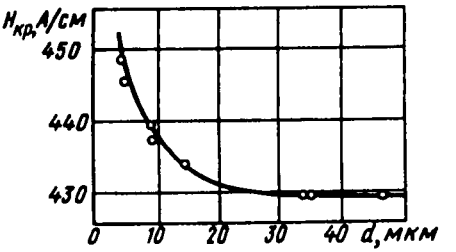


Рис. 10.1. Зависимость напряженности критического поля от диаметра свинцовой проволоки

сверхпроводящей области уменьшается и для обеспечения необходимой разности термодинамических потенциалов (10.11) требуются большие критические поля (рис. 10.1).

Равенство нулю постоянной интегрирования в уравнении (10.4) Ф. Лондон получил, исходя из допущения, что импульс сверхпроводящих электронов остается равным нулю и после включения магнитного поля («жесткость» волновой функции). Квантовая теория объясняет это тем, что сильная корреляция куперовских электронов с противоположными импульсами (суммарный импульс равен нулю) остается энергетически выгодной, пока внешнее поле не достигнет критического значения ($\mu_0 H_{кр}^2/2 \approx n_s \Delta$). Эффект Мейсснера и явление бесконечной проводимости оказываются взаимосвязанными.

Решение уравнения (10.4) не является единственным, так как условие равенства нулю суммарного импульса куперовской пары не может быть обосновано для неодносвязной области, например для сверхпроводящего кольца. В этом случае оно должно быть заменено обычным условием квантования Бора—Зоммерфельда

$$\oint \mathbf{p}_s\,\mathrm{d}\mathbf{l} = nh, \tag{10.12}$$

где n — целое число (1, 2, ...);

$$\mathbf{p}_s = m_s \mathbf{v}_s + q_s \mathbf{A} \tag{10.13}$$

— канонический импульс куперовской пары в магнитном поле $\mathbf{B} = \mathrm{rot}\,\mathbf{A}$*, а ее скорость

$$\mathbf{v}_s = \mathbf{J}_s/(n_s q_s). \tag{10.14}$$

---

* Магнитное поле действует перпендикулярно $\mathbf{v}_s$, изменяя только направление скорости, но не ее значение. Поэтому магнитное поле изменяет лишь импульс частицы, оставляя энергию постоянной.

Поэтому согласно условию (10.12)

$$\frac{nh}{q_s} = \frac{m_s}{n_s q_s^2} \oint \mathbf{J}_s \, d\mathbf{l} + \oint \mathbf{A} \, d\mathbf{l} = \frac{m}{ne^2} \oint \mathbf{J}_s \, d\mathbf{l} + \Phi = \Phi_F, \qquad (10.15)$$

где Φ_F — *флюксоид*. Токи в сверхпроводниках имеются только в очень тонком поверхностном слое, поэтому токовый член в выражении (10.15) необходимо учитывать при интегрировании по внутреннему краю кольца или для очень малых отверстий. Внутри массивного сверхпроводника $J_s = 0$ и выражение (10.15) превращается в условие квантования магнитного потока $\Phi = \oint \mathbf{A} d\mathbf{l} = \oint \mathbf{B} d\mathbf{S} = n\Phi_0$, где *квант магнитного потока* $\Phi_0 = h/2e = 2 \cdot 10^{-15}$ Вб иногда называют *флуксоном*.

Ограниченность теории Ф. и Г. Лондонов связана с пренебрежением зависимостью длины проникновения от длины свободного пробега электрона l и введенной в 1953 г. А. Пиппардом длины когерентности $\xi^{-1} = \xi_0^{-1} + l^{-1}$. Это следует из того, что равенство среднего импульса связанных электронов является точным лишь при усреднении по объему, занимаемому куперовской парой.

Феноменологическое приближение, полученное в нелокальной теории А. Пиппарда, не утратило до сих пор своего значения. Теория А. Пиппарда объяснила превышение получаемой в эксперименте глубины проникновения

$$\lambda = \frac{1}{B(0)} \int_0^\infty B(x) \, dx, \qquad (10.16)$$

глубины проникновения (10.8), а также неэкспоненциальный характер изменения магнитной индукции в зависимости от глубины проникновения x.

§ 10.3. Магнитные свойства сверхпроводников первого рода

Идеальными называют сверхпроводники, в которых переход из сверхпроводящего состояния в нормальное под воздействием магнитного поля обратим. Такие сверхпроводники подразделяют на две группы. К первой относятся *сверхпроводники первого рода* (или пиппардовские) с определенной температурой перехода в сверхпроводящее состояние. У них глубина поверхностного слоя λ меньше корреляционной длины ξ. К этой группе относят все чистые непереходные сверхпроводящие металлы ($\lambda_L \approx 3 \cdot 10^{-6}$ см). Признаками сверхпроводников первого рода являются также эффект Мейсснера и положительная энергия поверхности раздела между нормальной и сверхпроводящей фазами. Препятствуют их

использованию в качестве соленоидов для создания сильных магнитных полей низкие значения $H_{кр}$.

Ко второй группе относятся *сверхпроводники второго рода*, переходящие в сверхпроводящее состояние в некотором интервале температур. У них глубина проникновения λ больше корреляционной длины ξ. Подобными свойствами обладают переходные металлы и интерметаллические соединения типа Nb_3Ge, Nb_3Sn и V_3Ga ($\lambda \approx 2 \cdot 10^{-5}$ см). Образец из чистого непереходного металла можно превратить в сверхпроводник второй группы путем введения в него точечных примесей или использования в виде тонкой пленки. При этом корреляционная длина сокращается за счет значительного уменьшения длины свободного пробега электрона. Признаками сверхпроводников второго рода являются также частичное проникновение магнитного поля во всю толщу сверхпроводящего образца (в определенном интервале полей) и отрицательная поверхностная энергия.

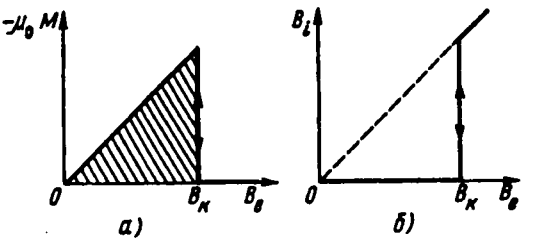


Рис. 10.2. Зависимость намагниченности и индукции внутреннего поля B_i внутри цилиндра от индукции в продольном внешнем поле B_e

Если в магнитном поле находится образец из *сверхпроводника первого рода* в виде эллипсоида с осью, параллельной направлению поля, то искажающее влияние образца на распределение поля внутри него можно выразить коэффициентом размагничивания.

Для длинного цилиндра, ось которого параллельна полю, размагничивающий фактор $N = 0$, напряженность поля внутри образца равна напряженности внешнего поля ($H_i = H_e = H$) и намагниченность образца в сверхпроводящем состоянии $M = -H_i = -H$. Для массивного цилиндра, в котором магнитный момент образца создают экранирующие токи в слое пренебрежимо малой толщины, намагниченность растет пропорционально внешнему полю до некоторого критического значения $B_e = \mu_0 H_{кр}$, при котором сверхпроводимость разрушается (рис. 10.2, *а*). До этого значения магнитная индукция $B_i = \mu_0 (H_i + M) = 0$, т. е. внутри цилиндра она полностью нейтрализована его намагниченностью (экранирующими токами). Для полей больше критического значения образец находится в нормальном состоянии: $B_i = B_e = B = \mu_0 H$ (рис. 10.2, *б*).

При других значениях коэффициента размагничивания продольная составляющая (вдоль внешнего поля) магнитной индукции внутри эллип-

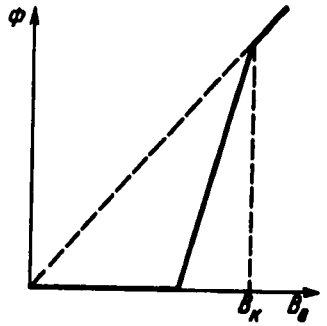


Рис. 10.3. Зависимость магнитного потока через цилиндрический стержень от магнитной индукции поперечного внешнего поля

соида $H_i' = H_e - NM$, причем при идеальном эффекте Мейсснера $M = -H_i$, т. е.

$$H_i' = H_e/(1-N). \qquad (10.17)$$

Из условия непрерывности тангенциальной составляющей напряженности магнитного поля на поверхности образца следует, что (10.17) позволяет получить выражение для максимального значения напряженности поля на поверхности эллипсоида вблизи «экватора». В частности, на поверхности проволоки круглого сечения в поле, перпендикулярном ее оси (N = 1/2), максимальная напряженность оказывается в два раза больше напряженности внешнего однородного поля ($H'_{max} = 2H_e$). Поэтому разрушение сверхпроводимости должно начаться во внешнем поле $H > H_{кр}/2$. Однако этот процесс не может распространиться на всю толщину проволоки, поскольку в нормальном ее состоянии поле проникло бы внутрь и стало бы меньше критического.

Оказывается, что во всем интервале $H_{кр}/2 < H < H_{кр}$ проволока расслаивается на нормальные и сверхпроводящие области, параллельные полю. Подобное состояние, когда сверхпроводник первого рода не является ни полностью сверхпроводящим, ни полностью нормальным, называют *промежуточным*. При возрастании внешнего поля в рассматриваемых пределах размеры нормальных областей внутри проволоки увеличиваются таким образом, что поле на экваторе поддерживается на постоянном уровне, соответствующем критическому режиму $H_{кр}$. Благодаря этому магнитный поток Φ, пронизывающий проволоку, монотонно увеличивается с ростом внешнего поля (рис. 10.3).

В тонкой металлической п л а с т и н е, расположенной перпендикулярно полю (N = 1), промежуточное состояние будет существовать во всей области полей $(1-N)H_{кр} = 0 < H < H_{кр}$. При этом на рис. 10.4 нормальные (n) и сверхпроводящие (s) области образуют слои, перпендикулярные плоскости чертежа, причем силовые линии проходят только через нормальные области. В условиях равновесия напряженность поля на плоскостях раздела равна $H_{кр}$. В областях n поле параллельно внешнему полю, т. е. $\operatorname{div}\mathbf{H} = 0$, $\operatorname{rot}\mathbf{H} = 0$. Поэтому в n-областях $\mathbf{H} = \mathbf{H}_{кр}$, а в s-областях $\mathbf{H} = 0$. Доля объема $\zeta = d_s/(d_s + d_n)$, занятого s-областями, может быть определена из условия сохранения магнитного потока Φ. Вдали от

пластины $\Phi = \mu_0 SH$, если S — площадь пластины. В самой пластине $\Phi = (1-\zeta)\mu_0 H_{кр} S$. Следовательно, $\zeta = 1 - H/H_{кр}$.

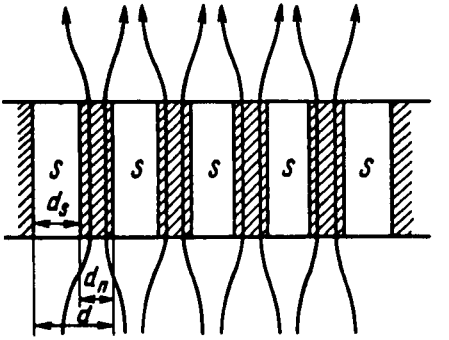


Рис. 10.4. Промежуточное состояние в пластине

В отсутствие внешнего магнитного поля промежуточное состояние может возникнуть при температурах ниже критических также под воздействием электрического тока. Например, это имеет место в достаточно длинном цилиндрическом проводнике радиусом a в определенном диапазоне значений постоянного тока I, создаваемого внешним генератором. Пока напряженность поля на поверхности $H_a = I/(2\pi a) < H_{кр}$, проводник находится полностью в сверхпроводящем состоянии. При токе $I > I_{кр} = 2\pi a H_{кр}$ проволока вблизи поверхности должна перейти в нормальное состояние только в ограниченном слое (с внутренним радиусом R). По всему сечению нормальное состояние невозможно, поскольку оно предполагает равномерное распределение тока, т. е. наличие в центральной части магнитного поля, меньшего критического $H_{кр}$. Поэтому ток во внутренней области $I_1 = 2\pi R H_{кр} = I_{кр} R/a < I_{кр} < I$. Ток во внешней нормальной области $I_2 = I - I_1$ создает электрическое поле напряженностью **E**, направленной вдоль оси проволоки. Если бы внутренняя область была полностью сверхпроводящей, то она накоротко замкнула бы внешний слой, что привело бы к дальнейшему расширению нормальной области. Таким образом, установившийся режим возможен лишь для промежуточного состояния внутренней области, когда n- и s-области расположены перпендикулярно оси цилиндра (рис. 10.5).

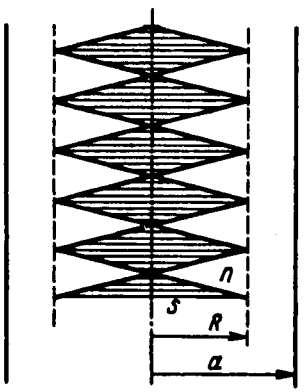


Рис. 10.5. Промежуточное состояние в проводе с током больше критического

Переходы между сверхпроводящим и нормальным состояниями могут быть использованы для создания бесконтактных управляющих устройств. В качестве примера можно привести криогенное устройство, предложенное для применения в вычислительной технике и названное

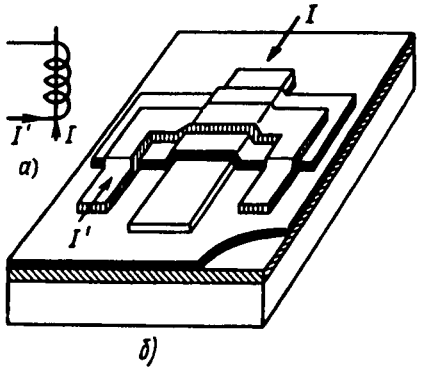


Рис. 10.6. Фазовый криотрон

криотроном. В первоначальном виде оно состояло из свинцовой или танталовой проволоки (клапан) с намотанной на нее управляющей катушкой из более тонкой ниобиевой проволоки (рис. 10.6, *а*). Если магнитное поле катушки меньше критического, то сопротивление клапана равно нулю, если поле больше критического, то сопротивление клапана конечно. Таким образом, током клапана можно управлять с помощью тока катушки. Индуктивность катушки значительно снижена и постоянная времени доведена до 10^{-9} с в криотроне, образованном двумя скрещенными металлическими пленками, разделенными изолирующим слоем толщиной около 100 нм (рис. 10.6, *б*).

До сих пор рассматривались магнитные свойства массивных образцов, для сверхпроводящего состояния которых было характерно отсутствие магнитного поля ($H = 0$), а свойства самого поверхностного слоя не играли существенной роли. Между тем поверхностный слой, являющийся границей сверхпроводящей области, обладает особыми свойствами. В частности, в нем $H \neq 0$.

Устойчивость сверхпроводящего состояния проявляется в том, что его термодинамический потенциал оказывается ниже, чем в нормальном состоянии. Разность их внутренних энергий, отнесенная к единице объема, составляет, согласно (10.11), величину $\mu_0 H_{кр}^2/2$. Поэтому можно считать, что связанная с выталкиванием магнитного поля из сверхпроводника объема V энергия $\mathscr{E}_в = \mu_0 H_{кр}^2 V/2$. Наличие магнитного поля в поверхностном слое на глубине порядка (10.16) означает, что в этой области происходит переход от сверхпроводящей к нормальной фазе и поэтому термодинамический потенциал постепенно повышается, приближаясь к своему значению в нормальной фазе. Возрастание термодинамического потенциала при образовании границы раздела двух фаз соответствует уменьшению энергии выталкивания на $\Delta\mathscr{E}_в = \lambda\mu_0 H_{кр}^2 S/2$, где S — площадь границы. Таким образом, проникновение магнитного поля в поверхностный слой связано с увеличением его энергии и определяет положительную составляющую *поверхностной энергии.*

Противоположную роль играет изменение плотности куперовских пар в поверхностном слое. Чем она выше, тем ниже термодинамический потенциал в соответствии с энергией $\mathscr{E}_s$, освобождаемой при

конденсации электронов в пары. Оценка поверхностной энергии с учетом квантовомеханических эффектов была впервые произведена в феноменологической теории В. Л. Гинзбурга и Л. Д. Ландау (1950 г.) и основывалась на рассмотрении образования сверхпроводящего состояния как фазового перехода второго рода. В качестве параметра упорядочения, плавно изменяющегося в процессе перехода, используются концентрация сверхпроводящих электронов и их волновая функция ψ^*. Обычно параметр упорядочения ω нормируется так, что в сверхпроводящей фазе $\omega = 1$, в нормальной фазе $\omega = 0$.

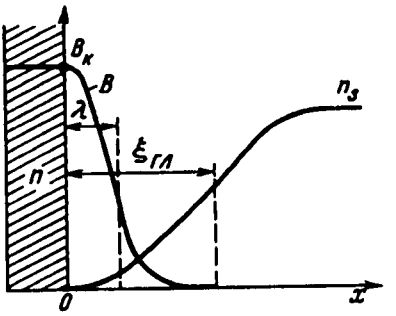


Рис. 10.7. Распределение плотности куперовских пар и магнитного поля вблизи границы между нормальной ($x < 0$) и сверхпроводящей ($x > 0$) областями

В отличие от теории Лондонов теория Гинзбурга—Ландау учитывает изменение волновой функции под воздействием магнитного поля и наличие ее пространственных градиентов в поверхностном слое. Если в теории Лондонов «жесткая» волновая функция электронов скачкообразно изменяется при переходе через границу фаз, то более строгий анализ Гинзбурга—Ландау приводит к плавному изменению в этой области концентрации сверхпроводящих электронов и параметра упорядочения. В связи с этим было получено уточненное значение *длины когерентности*

$$\xi_{гл} = \frac{1}{n_s(\infty)} \int_0^\infty n_s(x)\, dx \approx \frac{\xi_0}{\sqrt{1 - T/T_{кр}}},$$

большей среднего размера куперовской пары ξ_0. Эта длина характеризует минимальное расстояние, на котором может заметно изменяться плотность куперовских пар (рис. 10.7), отражая как бы меру их «жесткости». Связанная с этим зависимость энергии

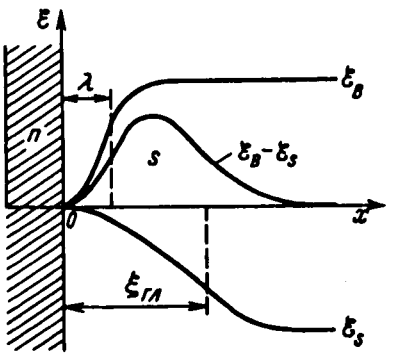


Рис. 10.8. Распределение энергий выталкивания и конденсации вблизи поверхностного слоя в сверхпроводниках первого рода

---

* Квадрат модуля волновой функции $|\psi|^2$ пропорционален плотности сверхпроводящих электронов.

конденсации $\mathscr{E}_s$, от расстояния до фазовой границы изображена на рис. 10.8.

Уменьшение плотности куперовских пар по сравнению с равновесным значением (при $x > \xi_{гл}$) приводит к понижению энергии конденсированного состояния на $\Delta\mathscr{E}_s = \xi_{гл}\mu_0 H_{кр}^2 S/2$. Энергия границы между фазами

$$\mathscr{E} = \int_0^\infty (\mathscr{E}_в - \mathscr{E}_s)\,\mathrm{d}x = \Delta\mathscr{E}_в - \Delta\mathscr{E}_s = (\xi_{гл} - \lambda)\frac{\mu_0 H_{кр}^2}{2} S,$$

т. е. *поверхностная плотность энергии*

$$\sigma_{гр} = (\xi_{гл} - \lambda)\mu_0 H_{кр}^2/2 = \Delta'\mu_0 H_{кр}/2, \qquad (10.18)$$

где $\Delta' = \xi_{гл} - \lambda$ — параметр свободной энергии поверхности.

В сверхпроводниках первого рода ($\xi_{гл} > \lambda$) поверхностная энергия (10.18) и параметр Δ' положительны. Следовательно, для создания границы между сверхпроводящей и нормальной фазами необходима затрата некоторой энергии. Это объясняет причину расслоения сверхпроводника в промежуточном состоянии только на конечное число сверхпроводящих областей (расслоение на большое число очень тонких областей потребовало бы затрат энергии на граничные слои, которые не могут компенсировать энергетический выигрыш от создания конденсированных состояний с повышенными критическими полями) и причину перехода сверхпроводника в промежуточное состояние на рис. 10.3 в полях, несколько превышающих $(1 - N)B_{кр}$.

§ 10.4. Магнитные свойства сверхпроводников второго рода

В сверхпроводниках второго рода глубина проникновения магнитного поля больше длины когерентности ($\lambda > \xi_{гл}$) и поверхностная энергия (10.18) отрицательна. Поэтому создание границы раздела между нормальной и сверхпроводящей фазами должно быть связано с выигрышем энергии и магнитное поле должно оказаться способным проникать в сверхпроводник при напряженностях, меньших критических. Количественное описание сверхпроводников можно дать, исходя из теории, развитой А. А. Абрикосовым на основе теории Гинзбурга—Ландау. Ее связь с микроскопической теорией была установлена Л. П. Горьковым. В связи с этим феноменологическая теория сверхпроводников второго рода получила название теории ГЛАГ. А. А. Абрикосовым, в частности, было показано (1957 г.), что проникновение магнитного поля внутрь сверхпроводника связано с образованием в нем нитевидной структуры. Когда

происходит частичное проникновение магнитного поля в сверхпроводящую область, под действием силы Лоренца электроны начинают двигаться по окружности, образуя вихри (*абрикосовские вихри*). Внутри вихря скорость электрона возрастает по мере приближения к оси, пока не происходит «срыв» сверхпроводимости. В результате этого сверхпроводящий образец оказывается пронизанным нитями из несверхпроводящих (нормальных) областей, ориентированных в направлении силовых линий магнитного поля. Магнитный поток, пронизывающий сечение вихря, одинаков для всех вихрей и равен кванту магнитного потока $\Phi_0 = h/(2e)$, что обеспечивает максимальное расслоение системы на фазы.

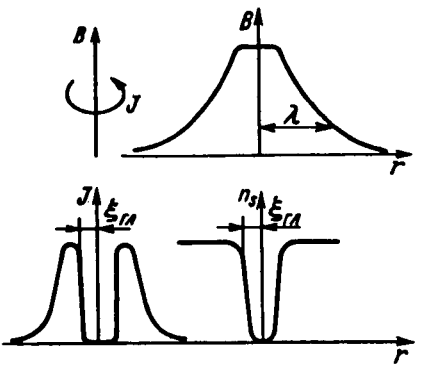


Рис. 10.9. Структура абрикосовского вихря

Вихревая нить представляет собой цилиндр радиусом порядка $\xi_{гл}$. Токи, охватывающие цилиндр, создают локальное магнитное поле, достигающее максимума в центре и спадающее по мере удаления от оси на расстояниях порядка λ (рис. 10.9). Рассмотрим структуру изолированной вихревой нити, предполагая $\lambda \gg \xi_{гл}$. При этом основной вклад в энергию нити дает магнитное поле в области $r > \xi_{гл}$, а ее электромагнитные свойства описываются уравнениями Лондонов. Например, для расстояний от центра нитей $r > \xi_{гл}$, согласно (10.5) $\mathbf{B} + \lambda^2 \operatorname{rot}\operatorname{rot}\mathbf{B} = 0$. Для области внутри нити ($r < \xi_{гл}$) необходимо воспользоваться более сложным уравнением, однако, поскольку радиус $\xi_{гл}$ сравнительно мал, можно имеющуюся особенность учесть двухмерной δ-функцией:

$$\mathbf{B} + \lambda^2 \operatorname{rot}\operatorname{rot}\mathbf{B} = \Phi_0 \delta(r). \qquad (10.19)$$

Убедиться в том, что Φ_0 представляет собой полный поток, переносимый нитью, можно, проинтегрировав обе части (10.19) по поверхности, охватываемой круговым контуром радиусом $r \gg \lambda$ с центром на оси нити. Тогда

$$\Phi_0 = \int \mathbf{B}\, d\mathbf{S} + \lambda^2 \oint \operatorname{rot}\mathbf{B}\, d\mathbf{l} = \int \mathbf{B}\, d\mathbf{S} + \mu_0 \lambda^2 \oint \mathbf{J}\, d\mathbf{l} = \int \mathbf{B}\, d\mathbf{S},$$

поскольку при $r \gg \lambda$ ток практически отсутствует и контурным интегралом можно пренебречь.

Из решения уравнений (10.19) и $\operatorname{div}\mathbf{B}=0$ следует, что

$$B=\frac{\Phi_0}{2\pi\lambda^2}K_0\left(\frac{r}{\lambda}\right), \tag{10.20}$$

где K_0 — функция Бесселя нулевого порядка от мнимого аргумента.

Асимптотические приближения (10.20) имеют вид

$$B=\frac{\Phi_0}{2\pi\lambda^2}\begin{cases}\ln(\lambda/r) \text{ при } r\ll\lambda; & (10.21\text{а})\\ \mathrm{e}^{-r/\lambda} \text{ при } r\gg\lambda. & (10.21\text{б})\end{cases}$$

При малой длине когерентности можно пренебречь изменением энергии конденсации сверхпроводящей фазы и считать, что энергия вихревой нити на единицу длины складывается из энергии магнитного поля и кинетической энергии сверхпроводящих электронов:

$$\mathscr{E}=\iint\limits_{r>\xi_{\text{гл}}}\left(\frac{BH}{2}+\frac{nmv^2}{2}\right)\mathrm{d}x\,\mathrm{d}y=$$

$$=\frac{1}{2\mu_0}\iint\limits_{r>\xi_{\text{гл}}}\left(B^2+\lambda^2\operatorname{rot}^2\mathbf{B}\right)\mathrm{d}x\,\mathrm{d}y, \tag{10.22}$$

где интегрирование ведется по плоскости, перпендикулярной направлению магнитного поля, и скорость выражена через ее значение из уравнения Максвелла $\mathbf{J}=n_s q_s\mathbf{v}=\operatorname{rot}\mathbf{H}$ с учетом (10.8). Подставив в (10.22) распределение поля (10.21 а) и учитывая, что $\operatorname{rot}_z\mathbf{H}=-\mathrm{d}H/\mathrm{d}r$, получим

$$\mathscr{E}\approx\frac{1}{\mu_0}\left(\frac{\Phi_0}{4\pi\lambda}\right)^2\ln\frac{\lambda}{\xi_{\text{гл}}}. \tag{10.23}$$

Рассматривая сверхпроводник второго рода как идеальный диамагнетик ($M=-H$) в виде длинного цилиндра с поперечным сечением S и длиной l в достаточно малом внешнем продольном поле, можно записать выражение для изменения потенциала:

$$\mathrm{d}g=-\mu_0 M\,\mathrm{d}H=\mu_0 H\,\mathrm{d}H=-\mu_0 H\,\mathrm{d}M.$$

Проникновение одной вихревой нити в образец означает изменение магнитного потока на Φ_0 и соответствующее изменение намагниченности, т. е. $S\,\mathrm{d}B=\mu_0 S\,\mathrm{d}M=\Phi_0$. Поэтому изменение термодинамического потенциала системы $Sl\,\mathrm{d}g=-H\Phi_0 l$.

Образование вихревой нити термодинамически выгодно, если

$$-Sl\,\mathrm{d}g=H\Phi_0 l>\frac{l}{\mu_0}\left(\frac{\Phi_0}{4\pi\lambda}\right)^2\ln\frac{\lambda}{\xi_{\text{гл}}}\approx\mathscr{E}l.$$

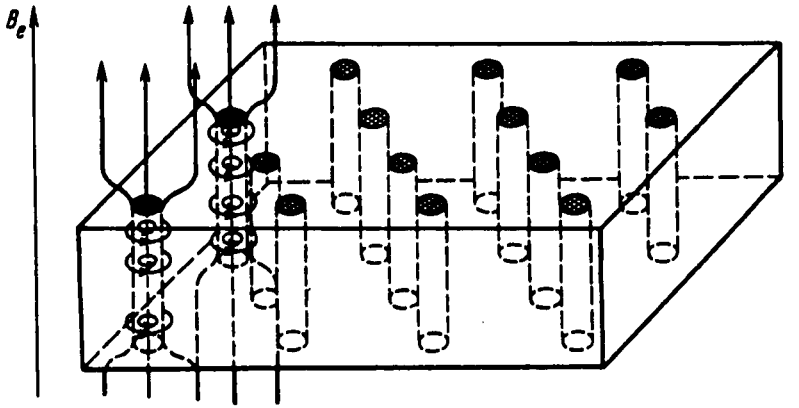


Рис. 10.10. Структура шубниковской фазы (магнитные силовые линии и круговые токи показаны только для двух нитей)

Следовательно, образование абрикосовского вихря возможно лишь в том случае, когда внешнее поле превышает пороговое значение

$$H_{к1} \approx \frac{\Phi_0}{16\pi^2\mu_0\lambda^2}\ln\frac{\lambda}{\xi_{гл}} \approx \frac{\mathscr{E}}{\Phi_0}, \qquad (10.24)$$

называемое *первым критическим полем* сверхпроводника второго рода. В полях напряженностью, меньшей $H_{кр1}$ в (10.24), его состояние мало отличается от сверхпроводника первого рода, т. е. он находится в мейсснеровской фазе.

По мере увеличения внешнего поля выше порогового значения (10.24) число вихревых нитей возрастает, а расстояние между ними — постепенно уменьшается. Такое состояние называют *смешанным* или *шубниковской фазой*, поскольку предположение о существовании сверхпроводников второго рода впервые было высказано Л. В. Шубниковым (в конце 30-х годов).

Эксперимент и расчет термодинамического потенциала показывают, что энергетически наиболее выгодной является вначале треугольная решетка вихревых нитей (рис. 10.10). При напряженности внешнего поля $H_1 = H_{кр1} + 0{,}0394\,H_{кр}/(\varkappa\sqrt{2})$ треугольная решетка переходит в квадратную ($\varkappa = \lambda/\xi_{гл}$).

При увеличении внешнего поля одновременно с уменьшением расстояния между нитями снижается средняя плотность куперовских пар. Однако в пространстве между нитями сохраняется сверхпроводящее состояние и образец в целом характеризуется идеальной электропроводностью. Когда расстояние между нитями станет сравнимым с $2\xi_{гл}$,

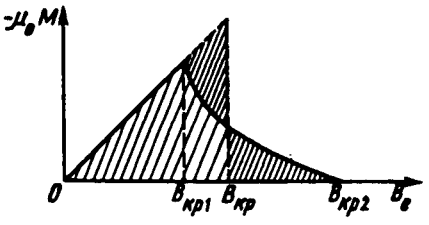


Рис. 10.11. Кривая намагничивания сверхпроводника второго рода

происходит столь сильное перекрытие токов, что понятие о сверхпроводящем круговом токе теряет смысл, плотность куперовских пар стремится к нулю и весь объем сверхпроводника переходит в нормальное состояние. При этом магнитное поле во всем объеме сравняется с внешним полем. Внешнее поле, при котором намагниченность становится равной нулю, называют *вторым критическим полем*. Его напряженность

$$H_{кр2} = \sqrt{2}\varkappa H_{кр}$$

просто выражается с помощью параметра Гинзбурга—Ландау $\varkappa = \lambda/\xi_{гл}$. Этот параметр различен для сверхпроводников первого и второго рода. В первом случае $\varkappa < 1/\sqrt{2}$, во втором $\varkappa > 1/\sqrt{2}$, и при этом всегда $H_{кр2} > H_{кр}$. Для $\varkappa >> 1/\sqrt{2}$ А. А. Абрикосовым было получено

$$H_{кр1} = H_{кр}(\ln\varkappa + 0{,}08)/(2\varkappa).$$

Следовательно, с ростом $\varkappa$ $H_{кр2}$ увеличивается, а $H_{кр1}$ уменьшается.

Кривая намагничивания стержня из сверхпроводника второго рода в продольном поле изображена на рис. 10.11. В полях, меньших первого критического (мейсснеровская фаза), происходит полное выталкивание магнитного поля из стержня. Пропорциональный рост намагниченности в этой области определяется необходимостью полной компенсации растущего внешнего поля. При $H_{кр1} < H < H_{кр2}$ (шубниковская фаза) возрастание внешнего магнитного поля сопровождается все большим его проникновением в сверхпроводник. В связи с этим намагниченность монотонно снижается вплоть до нуля.

При постоянных температуре и давлении, аналогично (10.11), разность термодинамических потенциалов

$$g_n - g_s = -\mu_0 \int_0^{H_{кр2}} M \, dH = \frac{\mu_0 H_{кр}}{2}, \qquad (10.25)$$

где $H_{кр}$ — критическое значение напряженности поля сверхпроводника первого рода, обладающего такой же разностью термодинамических потенциалов (график его намагничивания изображен на рис. 10.11 пунктирной линией). Согласно (10.25), площади под обеими кривыми должны быть одинаковыми, так же как и мелко заштрихованные площади.

Зависимость магнитного потока в стержне от напряженности магнитного поля (рис. 10.12) отражает практически полное выталкивание потока в мейсснеровской фазе и всевозрастающее его проникновение в сверхпроводник в шубниковской фазе. Сверхпроводник в шубниковской фазе способен выдерживать сильные поля, которые много больше критических полей сверхпроводников первого рода*. Например, для V_3Ga $H_{кр2} \approx 2{,}4 \cdot 10^5$ А/см. Однако для создания обмоток магнитных систем сверхпроводники второго рода непригодны, поскольку даже при очень слабом токе они переходят в нормальное состояние.

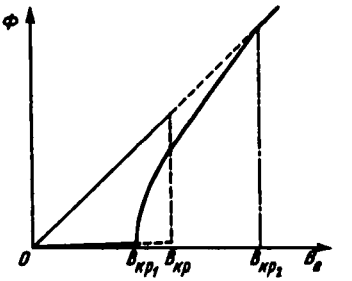


Рис. 10.12. Зависимость магнитного потока через цилиндрический стержень от индукции продольного внешнего поля

Одна из особенностей шубниковской фазы заключается в том, что при появлении транспортного тока в направлении поперек вихревых нитей он распределяется равномерно по всему сечению пластины, а не ограничен тонким слоем на поверхности сверхпроводника, как это происходит в мейсснеровской фазе. Важным следствием этого обстоятельства является взаимодействие между транспортным током и нитями магнитного потока. Это взаимодействие определяется силой

$$\mathbf{F} = \int \mathbf{J} \times \mathbf{B}\, dV,$$

где $\mathbf{J}$ — плотность только транспортного тока, поскольку вихревой ток не дает вклада в интеграл ввиду симметричного распределения; $\mathbf{B}$ — магнитная индукция внутри вихревой нити (интегрирование производят по объему рассматриваемой нити).

Если пренебречь изменением плотности транспортного тока на расстояниях, приблизительно равных размеру нити, то $\mathbf{F} = \Phi_0 \mathbf{J} \times \mathbf{e}_l l$ где $\mathbf{e}_l$ — единичный вектор по направлению нити; l — длина нити.

Сила взаимодействия, направленная перпендикулярно направлению тока и магнитного поля, приводит к перемещению нитей. Таким образом, смешанное состояние при наличии тока не может быть стационарным, а меняющееся магнитное поле приводит к возникновению ЭДС и электри-

---

* Качественно это следует из того, что с проникновением поля в сверхпроводник уменьшается энергия выталкивания. Поэтому для того, чтобы термодинамический потенциал сверхпроводящего состояния стал равен потенциалу нормального состояния, необходимы более сильные критические поля (как и в случае тонких сверхпроводников). Некоторое увеличение энергии конденсации при $\lambda \gg \xi_{гл}$ несущественно.

ческого поля, направленных по направлению тока. Таким образом возникают напряжение и диссипации энергии, т. е. превращение электрической энергии в теплоту.

Поскольку в идеальном сверхпроводнике сколь угодно малые транспортные токи ведут к смещению вихря и появлению сопротивления, то в шубниковской фазе это означает нулевое значение критического тока, разрушающего сверхпроводимость.

§ 10.5. Магнитные свойства сверхпроводников третьего рода

Сверхпроводники первого и второго рода представляют собой предельный случай, лишь приближенно отражающий свойства некоторых достаточно однородных реальных образцов. Техническое применение в качестве обмоток магнитных систем получили материалы, существенно отличающиеся от идеальных сверхпроводников дефектами кристаллической решетки и другими неоднородностями состава, хотя, как и в сверхпроводниках второго рода, в них возникает смешанное состояние при напряженностях полей в интервале между первым и вторым критическими значениями. В отличие от идеальных сверхпроводников процесс намагничивания и переход в нормальное состояние становятся необратимыми, что и послужило основанием для названия *неидеальные сверхпроводники второго рода*, «*жесткие*» или *сверхпроводники третьего рода.*

Характерной особенностью таких сверхпроводников является *пиннинг* (pinning — закрепление), т. е. притягивание вихревых нитей к неоднородностям и закрепление на них. К таким часто встречающимся неоднородностям относятся несверхпроводящие включения. Если вихревая нить, проходя через участок нормального металла, уменьшает свою длину в сверхпроводящей фазе на Δl (рис. 10.13), то в положении *а*) ее энергия будет на $\Delta \mathscr{E} = \mathscr{E} \Delta l$ меньше, чем в положении *б*). Эту энергию необходимо затратить для перемещения вихря из положения *а*) в положение *б*) и приложить соответствующую силу от *а*) к *б*). Когда в материале присутствует множество таких центров, вихревые нити займут энергетически наиболее выгодное положение. Аналогично можно объяснить действие и других центров пиннинга, в частности неоднородностей материала, неблагоприятных с точки зрения сверхпроводимости (областей с более низкой температурой фазового перехода, областей локальных сжа-

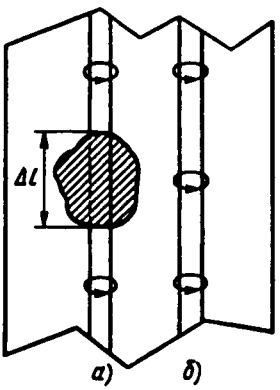


Рис. 10.13. Вихревые нити около несверхпроводящих включений

тий кристаллической решетки и т. д.).

Отсутствие свободной подвижности вихревых нитей, связь их с определенным участком сверхпроводника третьего рода затрудняют установление во внешнем магнитном поле намагниченности, соответствующей термодинамическому равновесию сверхпроводника второго рода. Это определяет совершенно иной вид кривых первого намагничивания и при циклических изменениях — петлю гистерезиса (рис.10.14).

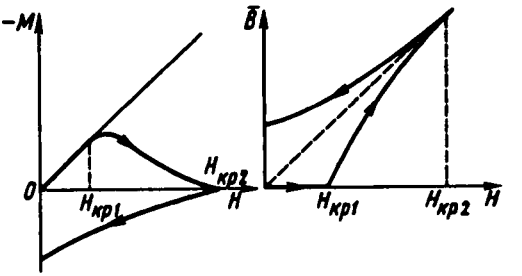


Рис. 10.14. Кривые статического намагничивания сверхпроводников третьего рода

До первого критического поля образец находится в мейсснеровской фазе, мало чувствительной к нарушениям структуры. Начиная с первого критического поля вихревые нити проникают в объем образца. Однако вследствие пиннинга они не могут равномерно распределиться по всему объему. Например, для сверхпроводника в полупространстве $x > 0$ средняя индукция $\overline{B}(x) \approx \sqrt{B^2(0) - \beta x}$, где $B(0)$ — индукция на поверхности; β — постоянный параметр. Проникновение вихревых нитей в приповерхностный слой увеличивает эффективную толщину экранирующего слоя и суммарный экранирующий ток, обеспечивающий диамагнитное поведение образца. При $H = H_{кр2}$ магнитное поле равномерно пронизывает весь проводник. Последующее понижение внешнего поля приведет к появлению вихревых нитей, с большим трудом покидающих сверхпроводник. Поэтому даже при исчезновении внешнего поля в сверхпроводнике останется некоторый «замороженный» магнитный поток, созданный вихревыми нитями и направленный вдоль внешнего поля. Полностью петлю гистерезиса можно снять, поменяв направление поля.

Существенным отличием сверхпроводников третьего рода от сверхпроводников второго рода является возможность пропускания по ним тока в присутствии поперечного магнитного поля. При этом некоторые сверхпроводники выдерживают очень большие плотности тока в достаточно сильном поле. Например, проволока из Nb_3Sn позволяет пропускать ток плотностью свыше 10^5 А/см^2 при напряженности поля, приблизительно равной 10^5 А/см. Сверхпроводники третьего рода, сохраняя преимущества сверхпроводников второго рода (способность выдерживать сильные магнитные поля), преодолевают их основной недостаток (нулевой критический ток).

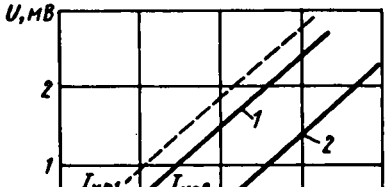


Рис. 10.15. Вольт-амперная характеристика образцов в резистивном состоянии (кривая *1* соответствует образцу с меньшими структурными дефектами)

Критический ток в данном магнитном поле соответствует предельному состоянию, при котором удерживающие центры еще могут обеспечить квазистатическую конфигурацию магнитного поля в сверхпроводнике. Связки нитей приходят в движение только в том случае, когда сила взаимодействия транспортного тока и нити магнитного потока оказывается достаточной для преодоления пиннинга и отрыва абрикосовских вихрей от неоднородности. После превышения транспортным током критического значения, срыва вихрей с неоднородностей и начала их движения возникает нагрев материала и сверхпроводимость исчезает. Величина критического тока зависит от масштаба неоднородностей, температуры и уровня внешнего магнитного поля.

В шубниковской фазе при токах выше критического сила Лоренца превосходит удерживающую силу пиннинга, возникает движение вихревых нитей, сопровождаемое потерями, нагревом материала и появлением напряжения. При этом образец переходит в особое состояние, которое отлично от сверхпроводящего и нормального. Оно получило название *резистивного состояния*.

Для ВАХ образцов в резистивном состоянии характерно то, что их дифференциальное сопротивление $R_d = \mathrm{d}U/\mathrm{d}I$ не зависит от тока при $I > I_{кр1}$ и меньше сопротивления R_n в нормальном состоянии, соответствующем пунктирной прямой на рис. 10.15 (сверхпроводник второго рода). Дифференциальное сопротивление не зависит от числа дефектов, хотя критический ток растет при их увеличении, причем в достаточно большом диапазоне полей $R_d/R_n \approx H/H_{кр2}(0)$, где $H_{кр2}(0)$ — напряженность второго критического поля при $T = 0$.

В резистивном состоянии напряжение на образце $U = R_d(I - I_{кр1})$ и мощность, выделяющаяся в нем

$$P = UI = R_d I(I - I_{кр1}) = R_n I(I - I_{кр1}) H/H_{кр2}(0). \qquad (10.26)$$

Выражение (10.26) свидетельствует о том, что вихревые нити под действием разности сил Лоренца $F_Л = \Phi_0 I$ и пиннинга $F_п = \Phi_0 I_к$ не ускоряются неограниченно, а движутся с постоянной скоростью v, характерной для вязкостного течения потока: $F_Л - F_п = (I - I_к)\Phi_0 = \eta v$, где η — коэффициент вязкости. Коэффициент вязкости можно выразить через

дифференциальное сопротивление, если вспомнить, что для n нитей потока на единицу площади индукция $B = n\Phi_0 = \mu_0 H$, а мощность в единице объема $p = ev F_Л = \mu_0 HI(I - I_{кр1})\Phi_0/\eta$. Следовательно, коэффициент вязкости $\eta = \mu_0 \Phi_0 H_{кр}(0)/\rho_n$ — действительно постоянная величина (ρ_n — удельное сопротивление в нормальном состоянии). Предполагают, что вязкостное сопротивление может быть связано с потерями энергии, обусловленными компонентой тока «нормальных» электронов, запаздыванием между процессами испарения и конденсации куперовских пар при перемещении вихревых нитей, а также возникновением дополнительного электрического поля при переходе электронов с круговых на циклоидные орбиты.

При *переменном токе* или пульсациях внешнего магнитного поля происходят циклические изменения магнитной индукции внутри сверхпроводника, приводящие к диссипации энергии и появлению некоторого сопротивления. Появление переменного тока в сверхпроводнике сопровождается возникновением электрического поля, ускоряющего куперовские пары с конечной массой. Потери в идеальных сверхпроводниках связаны с движением неспаренных электронов в этом электрическом поле. В неидеальных сверхпроводниках к ним добавляются потери гистерезисного характера (до частот порядка десятка килогерц). На частотах 10—100 Гц напряженность электрического поля мала, неспаренные электроны дают пренебрежимо малый вклад в общий ток и критическая плотность тока мало зависит от частоты, причем по порядку величин ее амплитудные значения соответствуют критическим значениям постоянного тока. С дальнейшим повышением частоты происходит увеличение сопротивления в сверхпроводящем состоянии, критическая плотность переменного тока падает и может стать более чем на два порядка ниже критического постоянного тока. На частотах, для которых энергия фотонов $h\nu$ превышает ширину энергетической щели Δ, последняя мало сказывается на процессах обмена энергией и поведение сверхпроводника все более приближается к поведению нормального металла. Для видимого диапазона световых колебаний энергия квантов, составляющая несколько электрон-вольт, значительно превышает ширину энергетической щели (около 10^{-3} эВ) и сверхпроводник ведет себя как нормальный металл.

Сверхпроводники третьего рода находят все более широкое применение в различных областях науки и техники. Одно из важнейших достижений сверхпроводниковой техники — создание сверхсильных магнитных полей, используемых в магнитогидродинамических генераторах, накопителях, ускорителях заряженных частиц и при исследовании термоядерных реакций. Современные сверхпроводящие соленоиды позволяют получать поля с индукцией около 20 Тл. Затраты энергии на охлаждение

(получение жидкого гелия) при этом могут быть в 1000 раз меньше энергии, требуемой для питания обычного электромагнита.

Применение сверхпроводящих обмоток в трансформаторах и электрических машинах открывает возможность почти полного устранения в них тепловых потерь, позволяет сделать их значительно более компактными и увеличить их мощности. Одной из причин этого является возможность отказа от стали в качестве магнитного материала, поскольку создаваемые сверхпроводниками магнитные поля намного превосходят их уровень в стальных материалах. Компактность и экономия массы особенно существенны при создании магнитных систем космических кораблей, в частности для защиты их от радиации.

Эффект механического отталкивания сверхпроводника, окруженного магнитной подушкой, используется для создания опор без трения и электрических вращающихся машин с КПД, практически равным 100%. Принцип *сверхпроводящего подвеса* может быть применен как в гироскопах, так и в сверхскоростных поездах. Применением нескольких вложенных друг в друга сверхпроводящих цилиндров можно добиться того, что во внутреннем цилиндре не будет содержаться ни одного кванта магнитного потока, т. е. можно получить идеальное магнитное экранирование. Ведется инженерная проработка сверхпроводящих волноводов и кабелей для линий передачи энергии на далекие расстояния.

Помимо передачи электрической энергии сверхпроводящая линия электропередачи (СП ЛЭП) должна обеспечить передачу больших потоков информации для управления самой ЛЭП. Для этого предполагается использовать в общей криогенной оболочке миниатюрные коаксиальные пары, позволяющие передавать мощные сигнальные потоки в СП ЛЭП. Другие сверхпроводящие пары внутри общей криогенной оболочки могут быть использованы для образования огромного числа каналов связи общего назначения. В этом случае СП ЛЭП превратится в комплексную энергоинформационную линию передачи. Высокочастотное сопротивление сверхпроводников на несколько порядков меньше, чем нормальных металлов, а потери в диэлектриках при криогенных температурах уменьшаются на один-два порядка. Это определяет высокое качество сверхпроводящих линий связи: небольшое затухание, полную экранировку, низкий уровень шумов, стабильность и широкополосность, недостижимые в других линиях передачи информации. Подобный экспериментальный образец успешно прошел испытания в нашей стране. В США проведено испытание сверхпроводящего кабеля связи длиной 1220 м для передачи информации при проведении подземных ядерных испытаний. Обычные кабели имели для этого недостаточную полосу пропускания.

Сверхпроводящие микрокоаксиальные кабели нашли практическое применение также в качестве линий задержки пикосекундных импульсов, рециркуляторов наносекундных импульсов и линий передачи СВЧ. В

подобных электронных устройствах сверхпроводящая пара используется в виде компактной катушки, помещаемой в криостат. Одна из линий задержки длиной 17 м имела полосу пропускания до 26 ГГц и время установления 16 пс при температуре 1,7 К. Применение сверхпроводящего рециркулятора позволило осуществить стробоскопическое осциллографирование однократного наносекундного импульса. Расчеты показывают, что затраты на эксплуатацию многоканальной телевизионной сверхпроводящей линии связи могут быть в несколько раз ниже расходов для обычных линий. Для сверхпроводящих микрокабелей с поперечным сечением 1—2 мм^2 при увеличении частоты от 20 до 100 ГГц коэффициент затухания возрастает от 0,04 до 0,8 дБ/м, причем основная часть затухания обусловлена потерями в диэлектрике.

§ 10.6. Слабосвязанные сверхпроводники

В высшей степени упорядоченная структура, соответствующая нулевой энтропии куперовских пар, достигается установлением во всем объеме строгих фазовых соотношений (фазовой когерентности). При этим *макроскопическое* сверхпроводящее состояние может быть описано общей волновой функцией $\Psi = \sqrt{n_s}\, e^{i\theta}$.

Фаза θ волновой функции не определяет стационарных свойств сверхпроводников, зависящих от плотности куперовских пар $n_s = \Psi^*\Psi$. Однако она играет существенную роль во всех кинетических эффектах. Например, из уравнений Шредингера для нестационарного ($j\hbar\dot{\Psi}/\partial t = \hat{\mathscr{H}}\Psi$) и стационарного ($\hat{\mathscr{H}}\Psi = \mathscr{E}\Psi$) состояний следует, что скорость изменения фазы волновой функции

$$\hbar\frac{\partial\theta}{\partial t} = -\mathscr{E} \tag{10.27}$$

пропорциональна энергии пары $\mathscr{E}$. Из квантовомеханического выражения для плотности тока

$$\mathbf{J} = j\frac{q_s\hbar}{2m_s}(\Psi\nabla\dot{\Psi} - \dot{\Psi}\nabla\Psi) - \frac{q_s^2}{m_s}\mathbf{A}\Psi\dot{\Psi} = \frac{q_s n_s}{m_s}(\hbar\nabla\theta - q_s\mathbf{A}) \tag{10.28}$$

следует, что сверхпроводящий ток и скорость пар пропорциональны градиенту фазы.

В квантовой механике фаза волновой функции и число электронов являются канонически сопряженными переменными, для которых, согласно соотношению неопределенностей $\delta n_s \delta\theta \geqslant 1$. Следовательно, в

образце сверхпроводника с фиксированным числом электронных пар абсолютное значение фазы волновой функции не определено. Поэтому о фазе отдельного сверхпроводника говорить не имеет смысла, хотя можно рассматривать соотношение фаз в двух разных сверхпроводниках.

Когда сверхпроводники находятся далеко друг от друга, соотношение (разность) фаз меняется хаотически (фазы меняются независимо друг от друга).

По мере сближения сверхпроводников в области зазора происходит все большее перекрытие волновых функций, быстро затухающих в вакууме по экспоненциальному закону. Когда зазор станет много меньше корреляционной длины, устанавливается стабильное соотношение фаз, обеспечивающее минимум энергии связи (при ширине зазора порядка 10...100 Å). Образуется система из двух сверхпроводников, обладающих когерентными состояниями и называемых *слабосвязанными*.

Одной из особенностей сверхпроводников, связанных изолирующим слоем, является скачкообразное изменение фазы при переходе через область связи (в пределах сверхпроводника возможно лишь непрерывное изменение фазы). Туннелирование через изолирующий барьер куперовских пар из основного состояния (с поверхности Ферми) одного сверхпроводника на поверхность Ферми другого не требует энергии на разрыв пар и не сопровождается изменением их энергии, т. е. сверхпроводящий туннельный ток может возникнуть и при равной нулю разности потенциалов между сверхпроводниками. В этом проявляется качественное отличие данного явления от туннелирования через потенциальный барьер между двумя сверхпроводниками нормальных электронов.

Связь между сверхпроводящим туннельным током и разностью фаз может быть получена путем распространения метода связанных мод на квантовые состояния куперовских пар.

Если две квантовые системы, описываемые волновыми функциями Ψ_1 и Ψ_2, полностью изолированы друг от друга (связь отсутствует), то скорость изменения во времени волновых функций, согласно уравнению Шредингера, определяется производными

$$\frac{\partial \Psi_1}{\partial t} = -\frac{j}{\hbar}\mathscr{E}_1\Psi_1, \qquad \frac{\partial \Psi_2}{\partial t} = -\frac{j}{\hbar}\mathscr{E}_2\Psi_2,$$

где $\mathscr{E}_1$, $\mathscr{E}_2$ — собственные энергии, являющиеся гамильтонианами отдельных сверхпроводников.

Если системы слабо связаны между собой, то скорость изменения волновой функции на одной стороне перехода зависит от мгновенных значений волновых функций по обе стороны перехода:

$$j\hbar\frac{\partial\Psi_1}{\partial t}=\mathscr{E}_1\Psi_1+K\Psi_2;$$
$$j\hbar\frac{\partial\Psi_2}{\partial t}=\mathscr{E}_1\Psi_2+K\Psi_1, \qquad (10.29)$$

где K — матричный элемент, обеспечивающий связь между волновыми функциями системы и определяющий интенсивность обмена куперовскими парами между сверхпроводниками.

Особенностью волновых функций двух слабо связанных сверхпроводников является то, что они приближенно описывают состояния с макроскопическим заполнением: $\Psi_1=\sqrt{n_{s1}}\,e^{j\theta_1}$, $\Psi_2=\sqrt{n_{s2}}\,e^{j\theta_2}$. В этом случае из (10.29)

$$\frac{1}{2\sqrt{n_{s1}}}\frac{\partial n_{s1}}{\partial t}e^{j\theta_1}+j\sqrt{n_{s1}}\frac{\partial\theta_1}{\partial t}e^{j\theta_1}=-\frac{j}{\hbar}\left(\mathscr{E}_1\sqrt{n_{s1}}\,e^{j\theta_1}+K\sqrt{n_{s2}}\,e^{j\theta_2}\right);$$
$$\frac{1}{2\sqrt{n_{s2}}}\frac{\partial n_{s2}}{\partial t}e^{j\theta_2}+j\sqrt{n_{s2}}\frac{\partial\theta_2}{\partial t}e^{j\theta_2}=-\frac{j}{\hbar}\left(\mathscr{E}_2\sqrt{n_{s2}}\,e^{j\theta_2}+K\sqrt{n_{s1}}\,e^{j\theta_1}\right).$$

Разделение и сравнение вещественных и мнимых частей позволяет получить

$$\frac{1}{2\sqrt{n_{s1}}}\frac{\partial n_{s1}}{\partial t}=\frac{K}{\hbar}\sqrt{n_{s2}}\sin(\theta_2-\theta_1);$$
$$\frac{1}{2\sqrt{n_{s2}}}\frac{\partial n_{s2}}{\partial t}=\frac{K}{\hbar}\sqrt{n_{s1}}\sin(\theta_1-\theta_2);$$
$$-\hbar\sqrt{n_{s1}}\frac{\partial\theta_1}{\partial t}=\mathscr{E}_1\sqrt{n_{s1}}+K\sqrt{n_{s2}}\cos(\theta_2-\theta_1);$$
$$-\hbar\sqrt{n_{s2}}\frac{\partial\theta_2}{\partial t}=\mathscr{E}_2\sqrt{n_{s2}}+K\sqrt{n_{s1}}\cos(\theta_1-\theta_2).$$

Разность фаз $\varphi=\theta_2-\theta_1$ и напряжение между сверхпроводниками $u=(\mathscr{E}_1-\mathscr{E}_2)/(2e)$. Поэтому предыдущие выражения в случае двух одинаковых сверхпроводников ($n_{s1}=n_{s2}=n_s$) можно переписать в виде

$$\frac{\partial n_{s1}}{\partial t}=-\frac{\partial n_{s2}}{\partial t}=\frac{2K}{\hbar}n_s\sin\varphi; \qquad (10.30)$$
$$\frac{\partial\varphi}{\partial t}=\frac{2e}{\hbar}u, \qquad (10.31)$$

причем при постоянном $u=U_0$

$$\varphi = U_0 \frac{2e}{\hbar} t + \varphi_0. \qquad (10.32)$$

При этом ток между сверхпроводниками объемом V составит

$$i_s = 2eV \frac{\partial n_{s1}}{\partial t} = I_m \sin\varphi = I_m \sin\left(\frac{2e}{\hbar} U_0 t + \varphi_0\right), \qquad (10.33)$$

если $I_m = 4en_s KV/\hbar$ и оба сверхпроводника связаны с источником тока, поддерживающим постоянное значение концентрации куперовских пар.

Из выражения (10.33) следует, что, действительно, постоянный туннельный ток связанных пар

$$I_s = I_m \sin\varphi_0 \qquad (10.34)$$

может возникнуть и при отсутствии напряжения на изолирующем барьере ($U_0 = 0$). Это явление называют *стационарным эффектом Джозефсона.*

Из выражения (10.33) также следует, что при постоянном напряжении на контакте $U_0 \neq 0$ в нем возникает переменный ток, частота которого $\omega_0 = 2eU_0/\hbar$. Это явление называют *нестационарным эффектом Джозефсона.* Благодаря этому явлению туннельный переход может служить источником, генератором электромагнитного излучения с частотой, управляемой в широких пределах постоянным напряжением U_0, причем $f_0/U_0 = 2e/h = \Phi_0^{-1} = 483$ МГц/мкВ.

Нестационарное туннелирование возникает также и в том случае, когда напряжение на переходе помимо постоянной содержит переменную составляющую: $u = U_0 + U_1 \cos\omega t$. Тогда из (10.31) разность фаз $\varphi = [2eU_1/(\hbar\omega)]\sin\omega t + (2eU_0/\hbar)t + \varphi_0$ и с учетом (10.30) и (10.33) ток

$$i_s = I_m \sin\{[2eU_1/(\hbar\omega)]\sin\omega t + \omega_0 t + \varphi_0\} =$$

$$= I_m \sum_{n=-\infty}^{\infty} J_n\{[2eU_1/(\hbar\omega)]\sin[(\omega_0 + n\omega)t + \varphi_0]\}, \qquad (10.35)$$

где J_n — функция Бесселя n-го порядка; предполагается, что $u << \Delta/e$ (при невыполнении последнего неравенства необходимо учитывать зависимость I_m от напряжения).

Если в напряжении на переходе имеется только переменная составляющая ($U_0 = 0$), то в сверхтоке (10.35) будет присутствовать постоянная составляющая (при $n = 0$)

$$I_0 = I_c J_0[2eU_1/(\hbar\omega)], \qquad (10.36)$$

зависящая от переменного сигнала. Это позволяет использовать туннельный переход в качестве детектора излучения миллиметрового и дальнего инфракрасного диапазонов.

При напряжениях с переменной составляющей $U_1 = n\hbar\omega/(2e)$ ($n = 1, 2, 3, ...$) появляются всплески тока или так называемые токовые ступеньки. При этом на барьере возникает дополнительная разность энергий куперовских пар, равная $2eU_1$. Поэтому если переменная компонента напряжения связана с внешним облучением и $\omega_0 = \pm n\omega$, то происходит туннелирование куперовских пар, сопровождаемое поглощением и излучением фотонов с частотой ω, определяющее появление компоненты постоянного тока:

$$i_n = I_c J_n\left(n\frac{U_1}{U_0}\right)\sin\varphi_0.$$

Туннелированию пар с поглощением фотонов соответствует положительная ступенька тока, а излучению фотонов — отрицательная ступенька. Максимальная амплитуда n-й ступеньки тока

$$|I_n(U_1)| = I_c\left|J_n\left(n\frac{U_1}{U_0}\right)\right|$$

зависит от амплитуды U_1 напряжения СВЧ и, следовательно, от квадратного корня из значения мощности поля СВЧ. Разность фаз φ_0 между джозефсоновским излучением тока $I_c = \sin\omega_0 t$ и внешним сигналом СВЧ определяется и регулируется внешним генератором постоянного тока (комплексное сопротивление джозефсоновского прибора обычно значительно меньше комплексного сопротивления источника).

Кроме рассмотренных ранее токов куперовских пар («сверхтоков») в слабосвязанных сверхпроводниках при $u > 0$ может возникнуть «нормальный» ток неспаренных электронов, обусловленный разрывом некоторого числа куперовских пар и некоторым снижением энергии связи 2Δ. Эти неспаренные электроны обычно называют *квазичастицами*, так как присутствие сверхпроводящего конденсата делает их свойства отличными от свойств электронов нормального металла. Квазичастичный или нормальный ток $i_N = Gu$ можно характеризовать «нормальной» проводимостью $G = 1/R$, которая при температуре, близкой к критической и напряжении, близком к щелевому ($U_g = 2\Delta/e$), близка к проводимости перехода $G = 1/R$ в нормальном состоянии. В общем случае эта проводимость имеет нелинейный характер.

Емкость C между электродами джозефсоновского перехода приводит к появлению тока смещения $i_D = C\dfrac{\mathrm{d}u}{\mathrm{d}t}$, если напряжение меняется со временем.

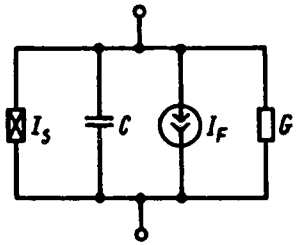


Рис. 10.16. Эквивалентная схема джозефсоновского перехода

Флуктуационный ток i_F связан с флуктуациями различной природы: тепловой (со спектральной плотностью $2kT/\pi R_m$), дробовой, от наводок и избыточной типа $1/f$.

Выражению для полного тока слабо связанных сверхпроводников

$$i = i_s + i_N + i_D + i_F = I_c \sin\varphi + Gu + C\frac{\mathrm{d}u}{\mathrm{d}t} + i_F =$$

$$= I_c \sin\varphi + G\frac{\hbar}{2e}\frac{\partial\varphi}{\partial t} + C\frac{\hbar}{2e}\frac{\partial^2\varphi}{\partial t^2} + i_F$$

соответствует эквивалентная схема, изображенная на рис. 10.16.

До сих пор рассматривалась связь в туннельной системе сверхпроводник—изолятор—сверхпроводник (рис. 10.17, *а*). Подобная система представляет собой только один из многих типов слабосвязанных сверхпроводников, общим признаком которых является наличие неоднородной сверхпроводящей структуры, содержащей в направлении тока короткую область с подавленной сверхпроводимостью. Например, в точечном контакте (рис. 10.17, *б*), тонкопленочном мостике (рис. 10.17, *в*) и мостике переменной толщины (рис. 10.17, *г*) подавление сверхпроводимости в центре мостика или контакта происходит в результате превышения локальной плотностью тока своего критического значения. В пленке сверхпроводника с нанесенной узкой полоской нормального металла *1* (рис. 10.17, *д*) подавление сверхпроводимости обусловлено эффектом близости. За счет этого же явления в двух близко расположенных сверхпроводящих планках с нанесенной на них пленкой из нормального металла или

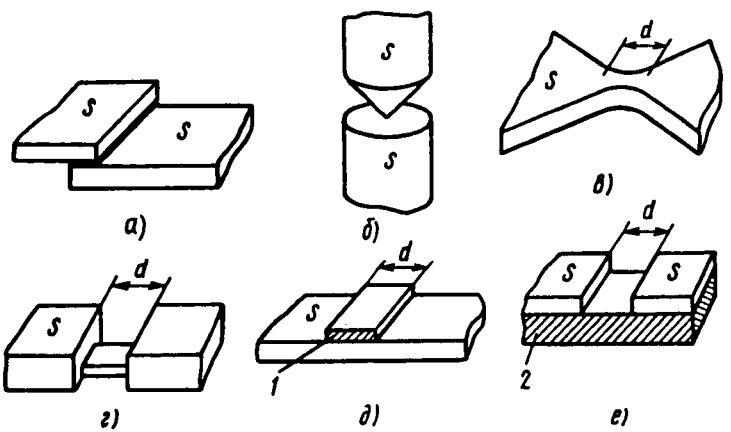


Рис. 10.17. Схемы слабосвязанных сверхпроводников

сильно легированного полупроводника *2* (рис. 10.17, *е*) происходит не только подавление, но и наведение сверхпроводимости в поверхностном слое перемычки (с пониженной относительно барьеров концентрацией куперовских пар). Несмотря на различие в параметрах различных типов, контактов слабосвязанных сверхпроводников, в них возникают эффекты Джозефсона, если размеры области подавления d не превышают длину когерентности ξ.

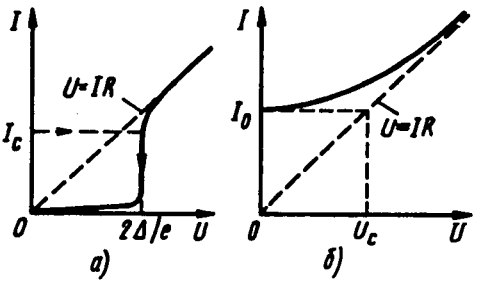


Рис. 10.18. Вольт-амперные характеристики контактов с прослойками из изолятора (*а*) и нормального металла (*б*)

Вольт-амперные характеристики зависят от типа структуры. Контакты с изолятором (см. рис. 10.17, *а*) обладают относительно большой емкостью. Поэтому их ВАХ, как правило, имеют гистерезис (рис. 10.18, *а*). В контактах с прослойкой из нормального металла (см. рис. 10.17, *г*) емкость существенно меньше и их ВАХ не имеют петли гистерезиса.

§ 10.7. Туннелирование сверхпроводящих электронов в магнитном поле

Зависимость фазовых соотношений между системами куперовских пар от магнитного поля может быть получена из выражения для канонического импульса (10.13) после исключения из него скорости (10.14)

$$\mathbf{p}_s = \frac{m_s}{n_s q_s}\mathbf{J}_s + q_s\mathbf{A} \qquad (10.37)$$

и выражения для разности фаз между точками i и j, расположенными вдоль направления $\mathbf{dn}$ распространения канонического импульса и волны де Бройля,

$$\varphi_{ij} = \theta_j - \theta_i = 2\pi\int_i^j \frac{\mathrm{d}n}{\lambda} = \frac{2\pi}{h}\int_i^j p\,\mathrm{d}n = \frac{2\pi}{h}\int_i^j \mathbf{p}\,\mathrm{d}\mathbf{l}\ ,$$

где последний интеграл остается в силе при любом расположении точек, между которыми определяется разность фаз.

Рассмотрим разность фаз в изолирующем промежутке, расположенном во внешнем постоянном магнитном поле $B_y = -B$, направленном параллельно оси y (рис. 10.19). Предположим, что в изолирующем

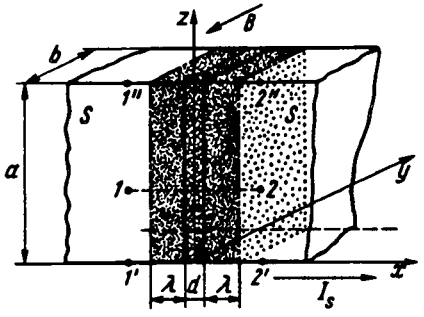


Рис. 10.19. Туннельный переход в магнитном поле

промежутке d магнитное поле однородно, а в прилегающих сверхпроводящих областях оно вследствие эффекта экранирования быстро спадает. Таким образом, без большой ошибки можно считать, что при $|x| < \lambda + d/2$ индукция $B_y = -B$, а при $|x| > \lambda + d/2$ $B = 0$. В области, занятой магнитным полем, векторный потенциал имеет единственную компоненту $A = A_z = -B_y x$, а в области $|x| > \lambda + d/2$ потенциал сохраняет постоянное значение $|A| = |A_z| = B(\lambda + d/2)$.

Разность фаз между двумя точками *1* и *2*, достаточно удаленными от туннельного контакта

$$\varphi = \theta_2 - \theta_1 = \frac{2\pi}{h}\int_1^2 \mathbf{p}\,\mathrm{d}\mathbf{l} = \frac{q_s}{\hbar}\int_1^2 \left(\frac{m_s}{n_s q_s^2}\mathbf{J}_s + \mathbf{A}\right)\mathrm{d}\mathbf{l} = \varphi_{11'} + \varphi_{1'2'} + \varphi_{2'2},$$

где

$$\varphi_{1'2'} = \theta_{2'} - \theta_{1'} = \frac{q_s}{\hbar}\int_{1'}^{2'} \left(\frac{m_s}{n_s q_s^2} J'_{sx} + A_x\right)\mathrm{d}x = \frac{m_s}{\hbar n_s q_s}\int_{1'}^{2'} J'_{sx}\,\mathrm{d}x = \varphi_0,$$

$$\varphi_{11'} = \theta_{1'} - \theta_1 = \frac{q_s}{\hbar}\int_1^{1'} \left(\frac{m_s}{n_s q_s^2} I_{sz} + A_z\right)\mathrm{d}z = \frac{q_s}{\hbar} B(\lambda + d/2)z,$$

$$\varphi_{2'2} = \theta_2 - \theta_{2'} = \frac{q_s}{\hbar}\int_{2'}^{2} \left(\frac{m_s}{n_s q_s^2} J_{sz} + A_z\right)\mathrm{d}z = \frac{q_s}{\hbar} B(\lambda + d/2)z,$$

поскольку плотность тока вследствие эффекта Мейсснера быстро спадает

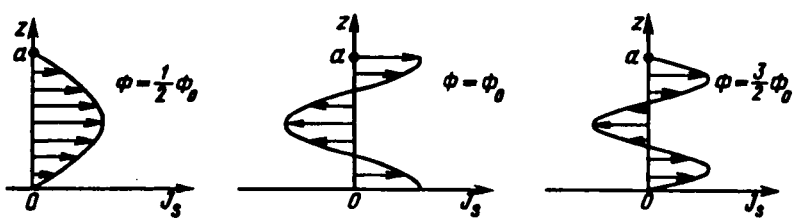


Рис. 10.20. Распределение сверхпроводящего туннельного тока в магнитном поле

при удалении от барьера и практически равна нулю при $|x_{1,2}| \gg \lambda + d/2$. Таким образом, $\varphi = \varphi_0 + \Delta\varphi$, где зависящая от магнитного поля разность фаз

$$\Delta\varphi = \frac{2e}{\hbar}(d + 2\lambda)Bz = \frac{2e}{\hbar}\Phi(z) = 2\pi\frac{\Phi(z)}{\Phi_0}, \qquad (10.38)$$

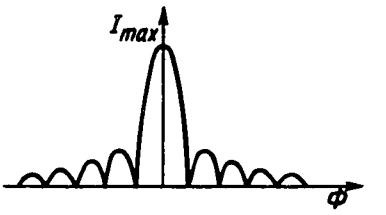


Рис. 10.21. Зависимость максимального сверхпроводящего туннельного тока от внешнего магнитного поля

а $\Phi(z) = (d + 2\lambda)Bz$ — магнитный поток, проходящий через плоскость контакта в пределах контура 11′2′2. Этот магнитный поток, как и разность фаз (10.38), линейно зависит от z. Максимальное значение $\Delta\varphi_{max} = 2\pi\Phi(a)/\Phi_0$ определяется шириной контакта a. Следовательно, плотность тока

$$J_s = 2e\frac{\partial n_{s1}}{\partial t} = J_m \sin\varphi = J_m \sin[\varphi_0 + 2\pi\Phi(z)/\Phi_0]$$

оказывается также зависящей от координаты z (рис. 10.20).

Полный ток через туннельный переход

$$I_s = \int_0^a\int_0^b I_s \, dy \, dz = \frac{ab\Phi_0}{\pi\Phi(a)} J_m \sin\left[\pi\frac{\Phi(a)}{\Phi_0} + \frac{\varphi_0}{2}\right]\sin\pi\frac{\Phi(a)}{\Phi_0}.$$

Изменяя напряжение U_0, можно изменить фазу φ_0 и, в частности, получить максимальное значение тока

$$I_{max} = I_m \frac{\sin[\pi\Phi(a)/\Phi_0]}{\pi\Phi(a)/\Phi_0}, \qquad (10.39)$$

которое по мере увеличения магнитного потока $\Phi(a)$ осциллирует с постепенно уменьшающейся амплитудой (рис. 10.21). Нулевое значение тока соответствует условию $\Phi(a) = n\Phi_0$, когда по высоте контакта a фаза меняется на $2\pi n$ ($n = 1, 2, 3, ...$) и вклады участков контакта с противоположными направлениями плотности тока оказываются одинаковыми.

§ 10.8. Макроскопическая квантовая интерференция

Сверхпроводящий туннельный переход может быть датчиком магнитного поля, поскольку его ток зависит от магнитного поля. Однако использование единичного перехода малоэффективно, так как площадь проникновения магнитного потока в диэлектрический барьер

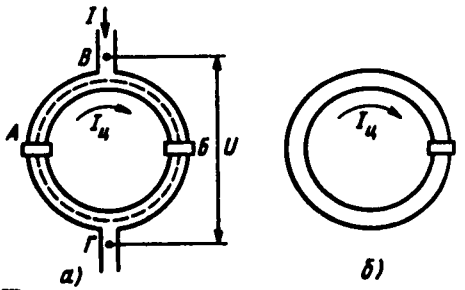


Рис. 10.22. Интерференционный магнитометр

мала (10^{-5}—10^{-6} см^2) и для изменения потока на один квант необходимы напряженности поля 10—100 мА/см. Поэтому развитие прецизионных магнитометров пошло по пути включения одного или двух джозефсоновских переходов в сверхпроводящее кольцо с площадью проникновения потока порядка 10 мм^2. Такое кольцо получило название СКВИД — сверхпроводящий квантовый интерференционный датчик.

СКВИД *на постоянном токе* образован сверхпроводящим кольцом с двумя туннельными переходами (рис. 10.22, *а*). Под действием внешнего магнитного поля, пронизывающего кольцо, в нем начнется циркуляция сверхпроводящего тока $I_ц$. Если выбрать внутри кольца замкнутый контур АВБГА так, чтобы он был удален от поверхности сверхпроводника на расстояние, большее λ, то плотность тока вдоль него везде, кроме областей переходов, будет равна нулю, а изменение фазы волновой функции вдоль этого контура

$$\varphi = \frac{1}{\hbar}\oint \mathbf{p}\,d\mathbf{l} = \frac{q_s}{\hbar}\oint\left(\frac{m_s}{n_s q_s^2}\mathbf{J}_ц + \mathbf{A}\right)d\mathbf{l} \approx$$

$$\approx \frac{q_s}{\hbar}\Phi + \varphi_A + \varphi_Б = 2\pi\frac{\Phi}{\Phi_0} + \varphi_A + \varphi_Б = 2\pi n,$$

где $\Phi = \oint \mathbf{B}\,d\mathbf{S}$ — магнитный поток в кольце от внешнего магнитного поля; $\varphi_A, \varphi_Б$ — разности фаз на переходах.

При вычислении потока Φ для удобства расчета сделано допущение, что погрешность от включения в путь интегрирования участков с очень короткими переходами пренебрежимо мала. Расчет предполагал также столь малую индуктивность кольца, что можно было пренебречь магнитным потоком, создаваемым в кольце током $I_ц$.

Для каждого из двух одинаковых переходов интерференционного датчика разность фаз

$$\Delta\varphi = \varphi_A = \varphi_Б = \pi(n - \Phi/\Phi_0) \qquad (10.40)$$

и, если внешний ток отсутствует ($I = 0$), циркулирующий ток

$$I_ц = I_m \sin\Delta\varphi = I_m \sin\left[\pi\left(n - \frac{\Phi}{\Phi_0}\right)\right] = I_m \cos\left(\pi\frac{\Phi}{\Phi_0}\right).$$

После включения постоянного внешнего тока I в переходе $А$ ток уменьшается на $I/2$, а в переходе $Б$ — увеличивается на $I/2$ и соответственно на $\pm\alpha$ изменяются разности фаз на переходах:

$$I_А = I_ц - I/2 = I_c \sin[\pi(n - \Phi/\Phi_0) - \alpha];$$
$$I_Б = I_ц + I/2 = I_c \sin[\pi(n - \Phi/\Phi_0) + \alpha],$$

откуда $I = 2I_c \cos[\pi(n - \Phi/\Phi_0)]\sin\alpha$.

Максимальное значение тока (при $\sin\alpha = 1$)

$$I_{max} = 2I_c \cos[\pi(n - \Phi/\Phi_0)] = 2I_c \cos(\pi\Phi/\Phi_0) \qquad (10.41)$$

изменяется периодически с максимумами при $\Phi = n\Phi_0$ и нулевыми значениями при $\Phi = (2n+1)\Phi_0/2$, т. е. с периодом, равным кванту потока. Если учесть изменение разности фаз в отдельных контактах под действием проникающего в них магнитного потока $\Phi_к$, то на зависимость (10.41) необходимо наложить осцилляции тока отдельных контактов (10.39). Тогда

$$I_{max} = 2I\frac{\sin(\pi\Phi_к/\Phi_0)}{\pi\Phi_к/\Phi_0}\cos\left(\pi\frac{\Phi}{\Phi_0}\right). \qquad (10.42)$$

Полученное выражение характеризует результат наложения двух интерференционных картин (аналогичная картина возникает в поле излучения двух изолированных антенн). Площадь, охватываемая контуром АВБГА, обычно много больше площади каждого из переходов. Поэтому второй сомножитель в (10.42) изменяется гораздо быстрее первого.

Высокочастотный СКВИД образован сверхпроводящей петлей с одним контактом. Если сверхпроводящее кольцо содержит один переход (рис. 10.22, *б*), то аналогично (10.40) разность фаз на туннельном переходе $\Delta\varphi = 2\pi(n - \Phi/\Phi_0)$, причем в общем случае полный поток через кольцо $\Phi = \Phi_e - LI_ц$, где Φ_e — магнитный поток внешнего поля; L — индуктивность кольца; $I_ц = I_c \sin[2\pi(n - \Phi/\Phi_0)]$ — наведенный в кольце циркулирующий ток. Зависимость полного потока

$$\Phi = \Phi_e + LI_c \sin[2\pi(\Phi/\Phi_0 - n)] \qquad (10.43)$$

от внешнего потока приведена на рис 10.23, *а* для трех значений параметра $l = 2\pi LI_c/\Phi_0$: 0,3; 1; 3. При $l > 1$ зависимость содержит неустойчивые участки (показаны пунктиром), благодаря чему возникают скачки полного потока и гистерезис. Эффективная индуктивность системы кольцо—переход

$$L_{эф} = -\frac{d\Phi_e}{dI} = L - \frac{1}{dI/d\Phi} = L + \frac{\Phi_0}{2\pi I_c \cos(2\pi\Phi/\Phi_0)} = L + L_J,$$

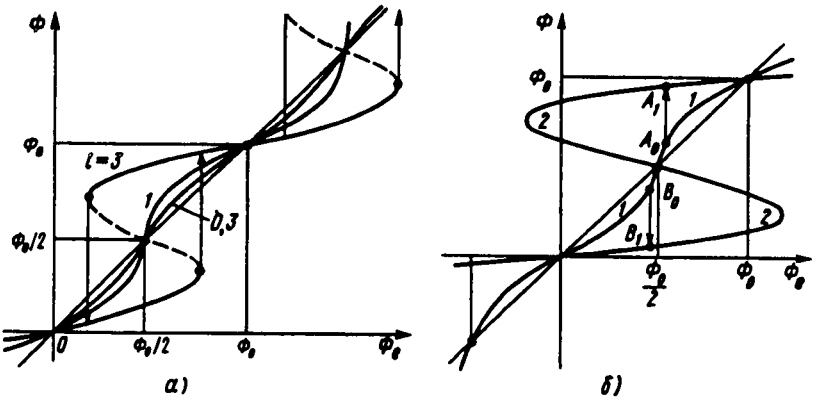


Рис. 10.23. Зависимости полного потока от внешнего потока в простом (*а*) и параметрическом (*б*) квантронах

где $L_J = \dfrac{L}{l\cos(2\pi\Phi/\Phi_0)}$ — эквивалентная индуктивность контакта.

В интерферометре с одним контактом последний оказывается замкнутым сверхпроводником, поэтому выходной сигнал снимают с помощью индуктивно связанного с ним резонансного контура с высокочастотным сигналом. Именно по этой причине такого рода устройство обычно называют *высокочастотным интерферометром* или *ВЧ-СКВИДом* (рис. 10.24, *а*). Используя его эквивалентную схему (рис.10.24, *б*) нетрудно получить индуктивность связанного со СКВИД резонансного контура

$$L_{эк} = L_к\left(1 - k^2\frac{L}{L_J + L}\right),$$

где $k^2 = M^2/(LL_к)$ и предполагается, что сопротивление перехода R_J достаточно мало и не шунтирует емкость перехода ($\omega_J CR_J \ll 1$). При $l \ll 1$

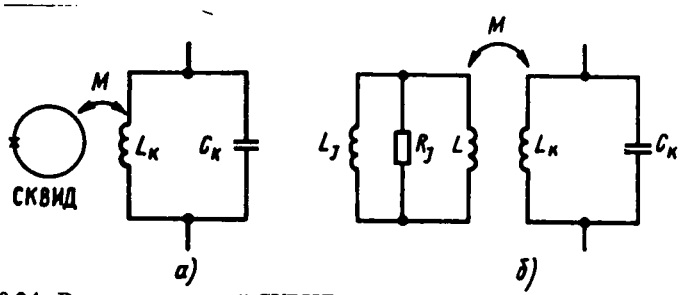


Рис. 10.24. Высокочастотный СКВИД с резонансным контуром (*а*) и его эквивалентная схема (*б*)

$$L_{эк} \approx L_к\left(1 - k^2 l \cos 2\pi \frac{\Phi_e}{\Phi_0}\right)$$

— периодическая функция магнитного потока через кольцо СКВИДа. Ее можно наблюдать, измеряя изменение эффективной резонансной частоты $\omega_{эк} = (L_{эк}C_к)^{-1/2}$ в зависимости от внешнего магнитного поля. При $k^2 l \ll 1$

$$\frac{\Delta f_{эк}}{f} \cong -\frac{1}{2}\frac{\Delta L_{эк}}{L_к} \approx \frac{1}{2}k^2 l\left(\cos 2\pi \frac{\Phi_e}{\Phi} - 1\right).$$

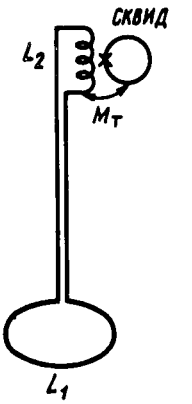


Рис. 10.25. Трансформатор магнитного потока

Для тонкопленочного СКВИДа в микроэлектронном исполнении площадь приемного контура примерно составляет 100 мкм^2, что соответствует чувствительности по напряженности магнитного поля 10^{-6} А/см. Для увеличения чувствительности и изолирования СКВИДа от области исследуемого источника поля, а также в случае настолько больших размеров образца, что он не может быть помещен непосредственно в СКВИД, применяют сверхпроводящие *трансформаторы магнитного потока.* Трансформатор магнитного потока обычно изготовляется из тонкой проволоки и представляет собой сверхпроводящую петлю, индуктивность которой складывается из индуктивности L_1 приемной катушки, связанной с внешним полем, и индуктивности L_2 вторичной катушки, связанной со СКВИДом (рис. 10.25). Появление в приемной катушке потока $\Delta\Phi$ приводит к циркуляции в трансформаторе экранирующего тока $I_т = N\Delta\Phi/(L_1 + L_2)$, поскольку в связи с квантованием потока полный поток в замкнутом сверхпроводящем контуре должен оставаться неизменным (N — число витков в приемной катушке). Попадающий в СКВИД поток

$$\Delta\Phi_e = M_т I_т = \frac{NM_т}{L_1 + L_2}\Delta\Phi.$$

Максимальное значение коэффициента преобразования

$$G = \frac{\Delta\Phi_e}{\Delta\Phi} = \frac{NM_т}{2L_2}$$

обеспечивает условие $L_1 = L_2$. Для реализации оптимального значения L_2 иногда используют во вторичной катушке ферромагнитный сердечник. Трансформаторы магнитного потока позволяют повысить предельную чувствительность СКВИДов по напряженности поля до 10^{-11} А/м.

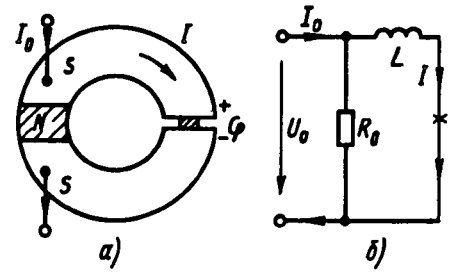


Рис. 10.26. Резистивный СКВИД (*а*) и его эквивалентная схема в отсутствие флуктуации (*б*)

К разновидностям СКВИДов с одним контактом относится резистивный интерферометр, в котором небольшая часть сверхпроводящего кольца заменена нормальным металлом (рис. 10.26). Такие устройства называют также *резистивными СКВИДами* или *РЕ-СКВИДами*. При напряжении смещения $U_0 = R_0 I_0$ согласно второму закону Кирхгофа

$$U_0 - R_0 i - L\frac{\mathrm{d}i}{\mathrm{d}t} - u = 0,$$

где i, u — ток и напряжение в джозефсоновском переходе.

Тогда

$$(1 + l\cos\varphi)\frac{\mathrm{d}\varphi}{\mathrm{d}t} - \omega_0\left(1 - \frac{I_c}{I_0}\sin\varphi\right) = 0,$$

где $\omega_0 = 2\pi U_0/\Phi_0$; $I_c/I_0 < 1$ (лишь в этом случае можно пренебречь квазичастичной проводимостью).

Если индуктивность L мала ($l \ll 1$), то для тока смещения I_0 система ведет себя как единый джозефсоновский переход с квазичастичным током. В отсутствии смещения ($I_0 = 0$) и $\Phi_e(t) = \mathrm{const}$ циркулирующий ток i затухает до нуля. Для высокочастотных составляющих Φ_e и малых сопротивлений R_0

$$U_0 - R_0 i = R_0(I_0 - i) \approx 0$$

и интерференция происходит как в обычном сверхпроводящем кольце. В этих условиях введение вставки из нормального металла, уничтожая интерференцию для постоянных значений фазы φ, сохраняет возможность измерения ее быстрых приращений. При $I_c > I_0 > 0$ джозефсоновский переход находится в резистивном состоянии с $\overline{u} \neq 0$, т. е. возникает джозефсоновская генерация с высокой стабильностью частоты $\omega_J = 2eI_c R_0/\hbar$.

В интерферометре с двумя контактами сверхпроводник не шунтирует переходы. Переключение под действием тока смещения, превышающем критический, формирует выходной сигнал в виде постоянного электрического напряжения, зависящего от внешнего магнитного поля. Устройство этого типа называют ПТ-СКВИДами. Спектральная плотность шума ВЧ-СКВИДов определяется шумами ВЧ-предусилителя. В ПТ-

СКВИДах основным источником шума является шум в сопротивлении, шунтирующем джозефсоновский контакт. Поэтому они имеют более высокий коэффициент преобразования и более предпочтительны при проведении прецизионных измерений. Их широкое применение затрудняют технологические сложности изготовления стабильных планарных квантовых интерферометров с идентичными параметрами обоих контактов. Использование ПТ-СКВИДов позволило получить энергетическое разрешение на уровне 10^{-33} Дж/Гц, что соответствует нескольким постоянным Планка. Предел их чувствительности на уровне $\hbar/2 = 5 \cdot 10^{-35}$ Дж/Гц определяется нулевыми квантовыми флуктуациями (в области предельно малых температур).

Наибольшее распространение СКВИДы получили в качестве магнитометров. Использование трансформатора потока позволило в серийных магнитометрах достигнуть чувствительности 10^{-14} Тл/Гц$^{1/2}$; применение несерийных СКВИДов увеличивает чувствительность на два-три порядка. До появления СКВИДов не удавалось проводить детальное исследование биомагнитных полей. СКВИД-магнитометры позволили восстановить трехмерную карту распределения магнитного поля отдельных органов и всего организма человека в целом и детально проследить изменения этих распределений во времени бесконтактным способом. Это особенно существенно для измерения магнитных сигналов мозга. СКВИДы используются также для магнитного картирования Земли с помощью спутников, исследования электромагнитных полей околоземного пространства, изучения солнечного ветра, реализации инфранизкочастотной связи. СКВИД представляет собой уникальную измерительную систему, позволившую впервые достичь квантовой области чувствительности на частотах, значительно меньших, чем kT/h. Интерференционные измерители магнитной восприимчивости обладают чувствительностью на один-два порядка лучшей чувствительности обычных измерительных устройств. Амперметры на основе СКВИДов измеряют токи с точностью около 10^{-14} А в полосе частот 1 Гц. Предельная чувствительность вольтметров на основе СКВИДов составляет примерно 10^{-14} В/Гц$^{1/2}$. РЕ-СКВИДы используются в шумовой термометрии для измерений температуры в диапазоне 10^{-5} К.

Применения СКВИДов в криоэлектронных системах логики и памяти рассматриваются в § 10.10.

§ 10.9. Электродинамика джозефсоновского контакта

Рассмотрим одномерный распределенный контакт (переход), пренебрегая конечностью его размеров в плоскости yz (см. рис. 10.19), дис-

сипацией (рассеянием) энергии в переходе и флуктуациями. Тогда плотность тока в переходе

$$J_x = J_s \sin\varphi + \frac{\partial D_x}{\partial t} = J_s \sin\varphi + C\frac{\mathrm{d}u}{\mathrm{d}t} = J_s \sin\varphi + \frac{\hbar C}{2e}\frac{\partial^2\varphi}{\partial t^2},$$

где $C = \varepsilon/d$ — емкость контакта на единицу его площади. Согласно (10.38),

$$\frac{\partial\varphi}{\partial z} = -\frac{2e\mu_0}{\hbar} H_y \delta.$$

Подставив эти выражения в уравнение Максвелла

$$\mathrm{rot}\,\mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t};\quad \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = J_x + \frac{\partial D_x}{\partial t},$$

получим

$$\frac{\hbar}{2e\mu_0\delta}\frac{\partial^2\varphi}{\partial z^2} = J_s \sin\varphi + \frac{\hbar C}{2e}\frac{\partial^2\varphi}{\partial t^2},$$

или

$$\lambda_J^2 \frac{\partial^2\varphi}{\partial z^2} - \frac{1}{\omega_J^2}\frac{\partial^2\varphi}{\partial t^2} = \sin\varphi; \qquad (10.44а)$$

$$\frac{\partial^2\varphi}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2\varphi}{\partial t^2} = \frac{1}{\lambda_J^2}\sin\varphi, \qquad (10.44б)$$

где $\lambda_J^2 = \hbar/(2e\mu_0\delta J_s)$; $\omega_J^2 = 2eJ_s/(\hbar C) = c^2 d/(\varepsilon_r \delta\lambda_J^2) = v^2\lambda_J^{-2}$.

В уравнениях (10.44 а) и (10.44 б) под φ можно подразумевать не только разность фаз, но и, как это следует из (10.38), нормированный магнитный поток Φ/Φ_0.

Параметр λ_J служит мерой ширины области по краям контакта, в которой имеется туннельный ток, его называют *джозефсоновской глубиной проникновения*. Этот параметр характеризует экранирование тока магнитным полем, созданным этим сверхтоком. Это явление аналогично эффекту Мейсснера, однако в то время, как лондоновская глубина проникновения λ_L составляет десятки нанометров, джозефсоновская глубина проникновения оказывается порядка сотни микрометров, что характерно для слабой сверхпроводимости барьера.

В частном случае, когда фаза φ не зависит от координаты z и мала ($\sin\varphi \approx \varphi$), решению уравнения $\mathrm{d}^2\varphi/\mathrm{d}t^2 + \omega_J^2\varphi = 0$ соответствуют колебания с частотой ω_J. В этих условиях фаза на протяжении всего барьера постоянна, магнитное поле отсутствует, а электрическое направлено перпендикулярно плоскости барьера, т. е. возникает структура продольных

плазменных волн. Они сопровождаются периодическим переходом энергии, сосредоточенной в барьере, в электростатическую энергию ($Cu^2/2$) и обратно. Подобные коллективные колебания называют *джозефсоновскими плазмонами*, а ω_J — *плазменной частотой*. Относительно низкие значения этой частоты ($\nu_J = 10^9$—10^{11} Гц) обусловлены относительно малой плотностью носителей заряда в барьере, характерной для слабой сверхпроводимости. Колебания ограничены областью барьера.

Для малых отклонений от значения $\varphi = 0$ и плоской волны вида $\varphi(z,t) = \varphi(0,0)\exp(\omega t - kz)$ по (10.44б) может быть получено дисперсионное уравнение $\omega^2 = \omega_J^2 + k^2 v^2$. Следовательно, плазменная частота ω_J представляет ту предельную наименьшую частоту, с которой электромагнитная волна может распространяться внутри контакта со скоростью v (в пределе нулевого джозефсоновского тока). Из выражения для резонансной частоты $\omega = 1/\sqrt{LC}$ согласно теории электрических цепей, эквивалентная дифференциальная индуктивность джозефсоновского контакта (в пределе нулевого тока)

$$L = 1/(C\omega_J^2) = \hbar/(2eJ_s) = \Phi_0/(2\pi J_s).$$

Индуктивность может принимать отрицательные значения, что обуславливает характерное отличие свойств джозефсоновского перехода от свойств других нелинейных реактивных элементов. Одним из его проявлений является рассмотренное ранее явление джозефсоновской генерации колебаний СВЧ с джозефсоновской частотой $\omega_J = 2eU_0/\hbar = 2\pi U_0/\Phi_0$.

Вернемся к общему решению уравнения (10.44а).

Электродинамику протяженного джозефсоновского контакта ($a \gg \lambda_J$) удобнее исследовать после перехода к нормированным переменным $z_н = z/\lambda_J$, $t_н = \omega_J t$ и $\dot{\Phi} = 2\pi\Phi/\Phi_0 = \varphi$. Тогда уравнение (10.44а) приобретет более простой вид

$$\frac{\partial^2\dot{\Phi}}{\partial z_н^2} - \frac{\partial^2\dot{\Phi}}{\partial t_н^2} = \sin\dot{\Phi}, \qquad (10.45)$$

известный под названием *уравнение синус—Гордона* (УСГ)*.

Волновое уравнение (10.45) оказывается инвариантным по отношению к преобразованию Лоренца

---

* В дальнейшем знак «•» над нормированными значениями Ф и индексы «н» опускаются.

$$z \to z' = \frac{z - ut}{\sqrt{1-u^2}}, \quad t \to t' = \frac{t - uz}{\sqrt{1-u^2}},$$

что приводит к сокращению пространства и времени в $\sqrt{1-u^2}$ раз так же, как в специальной теории относительности. В волноводе, образованном джозефсоновским переходом, ненормированная предельная скорость $v_0 = c\sqrt{d/(\varepsilon_r \delta)}$ приблизительно в 20 раз меньше скорости света для типичной геометрии перехода.

Механической моделью УСГ и полосковой линии, образованной тунельным переходом, может служить цепочка маятников, колеблющихся вокруг одной оси и связанных с соседними маятниками пружиной (навитой на ось колебаний). В этой модели член φ_{zz} соответствует нормированному вращающему моменту пружины, член φ_{tt} — нормированному моменту инерции маятника, а член $\sin\varphi$ — моменту силы тяжести маятника.

Рассмотрим класс частных решений УСГ в виде бегущих волн неизменной формы, распространяющихся в бесконечной линии с постоянной скоростью и имеющих вид $\Phi = \Phi(z - v\,t) = \Phi(\xi)$. Для этого случая уравнение (10.45) принимает вид

$$\frac{d^2\Phi}{d\xi^2} = \frac{\sin\Phi}{1-v^2},$$

совпадающий по форме с уравнением движения обычного плоского маятника. Умножив полученное уравнение на $d\Phi/d\xi$, найдем

$$\frac{d\Phi}{d\xi} = \sqrt{2\frac{E - \cos\Phi}{1-v^2}},$$

где E — постоянная интегрирования. Если постоянная E не равна единице, то решение можно записать в неявном виде

$$\sqrt{1-v^2}\int_0^{\Phi} \frac{d\Phi}{\sqrt{2(E-\cos\Phi)}} = \xi. \qquad (10.46)$$

При $E > 1$ и $v < 1$ решение является монотонно возрастающей функцией ξ. При $-1 < E < 1$ и $v > 1$

$$\Phi = 2\arcsin\left[k\,\mathrm{sn}\left(\frac{\xi}{\sqrt{v^2-1}}; k\right)\right],$$

где sn — эллиптическая функция Якоби с модулем $0 < k < 1$. Это решение описывает периодическую плазменную волну большой амплитуды, осциллирующую около значения $\Phi = 0$. В плазменной волне фаза (10.32) и

ток (10.33) меняются также периодически с нулевым средним значением. При $E = 1$ и $v < 1$ решение

$$\Phi = 4\,\mathrm{arctg}\left[\exp\left(\pm\frac{\xi-\xi_0}{\sqrt{1-v^2}}\right)\right] \qquad (10.47)$$

описывает уединенную волну, называемую *кинком* (ξ_0 — произвольная постоянная, зависящая от начала отсчета переменных z и t). Импульсное решение (10.47) предусматривает один поворот фазы на угол 2π (рис. 10.27, *б*). Индексы $\pm$ указывают направление вращения и соответствуют кинку или антикинку (рис. 10.27, *в*). Кинки и антикинки могут сталкиваться без взаимного разрушения, т. е. являются солитонами и антисолитонами.

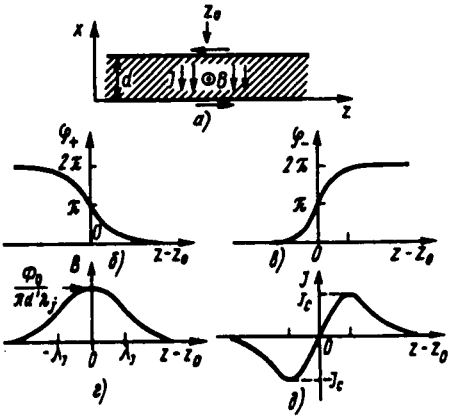


Рис. 10.27. Структура флуксона

§ 10.10. Динамика флуксона

Уединенная волна (10.47) представляет собой математическое описание *джозефсоновского вихря* (рис. 10.27, *а*), т. е. области размерами порядка $4\lambda_J$, в которую проникает магнитное поле (рис. 10.27, *г*) и вокруг которой циркулирует незатухающий сверхток (рис. 10.27, *д*). Интегрированием выражения $B = -\dfrac{\hbar}{2e\delta}\dfrac{\partial\varphi}{\partial z} = -\dfrac{\Phi_0}{4\pi^2\delta}\dfrac{\partial\Phi}{\partial\xi}$ по поперечному сечению магнитного потока можно убедиться, что каждый солитон (антисолитон) содержит ровно один положительный (отрицательный) квант потока Φ_0. Следовательно, решение (10.47) соответствует распространению в джозефсоновском контакте одиночного изолированного флуксона (антифлуксона). Свободная энергия такого вихря положительна и поэтому для его образования необходимы внешние источники энергии (внешние токи).

При $E > 1$ и $v^2 < 1$ решению (10.46) в виде

$$\Phi = 2\arcsin\left[\pm\,\mathrm{cn}\left(\frac{\xi-\xi_0}{k\sqrt{1-v^2}};k\right)\right]$$

соответствует решетка движущихся флуксонов (антифлуксонов) с периодом $2k(1-v^2)K(k)$, где $K(k)$ — полный эллиптический интеграл первого рода (cn — эллиптическая функция Якоби с модулем $0 < k < 1$).

Решению уравнения (10.45) в виде

$$\Phi = 4\operatorname{arctg}\{\operatorname{tg}\theta \sin[(t-t_0)\cos\theta \operatorname{sch}(z\sin\theta)]\} \qquad (10.48)$$

при $0 < \theta < \pi/2$ соответствует осциллирующее состояние уединенной волны (рис. 10.28), называемое *бризером* (breather — дышащее решение). Бризер иногда возникает при столкновении флуксона и антифлуксона. Уравнение (10.48) приведено в системе отсчета, связанной с их центром масс.

Общее решение УСГ состоит из N солитонов, L бризеров и обычного излучения. Солитоны и бризеры составляют компоненты волны, соответствующие флуксонам. Их число и скорость определяются спектральным преобразованием начального возмущения.

Рассмотренные ранее джозефсоновские вихри описывают распространение флуксонов в джозефсоновских контактах неограниченной протяженности. Для анализа явлений в контактах ограниченной длины l необходимо найти решение УГС, удовлетворяющее граничным условиям. В простейшем случае джозефсоновский контакт можно рассматривать как полосковую линию, открытую на обоих концах. Ее волновое сопротивление $\sqrt{L/C}$, как правило, намного меньше волнового сопротивления $\sqrt{\mu_0/\varepsilon_0}$ открытого пространства. Поэтому излучение из концов очень мало по интенсивности, в первом приближении им можно пренебречь, положив на концах линии ток равным нулю, т. е.

$$\Phi_z(0) = \Phi_z(l) = 0. \qquad (10.49)$$

Решением уравнения (10.45) с граничными условиями (10.49) является выражение

$$\Phi = 4\operatorname{arctg}[f(z)g(t)],$$

где в общем случае f, g — эллиптические функции, удовлетворяющие уравнениям $f_z^2 = af^4 + (1-b)f^2 - c$; $g_t^2 = cg^4 + bg^2 - a$ с произвольными константами a, b и c. Нелинейные стоячие волны в прямоугольных контактах, получаемые из этих решений, можно разделить на три основных класса: бризерные, флуксонные и плазменные.

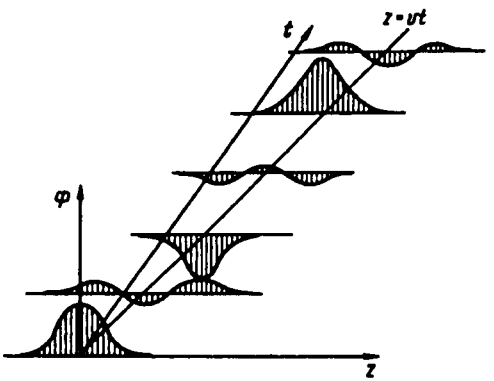


Рис. 10.28. Бризер уравнения синус-Гордона

Бризерное колебание соответствует либо флуксон-антифлуксонной паре в центре контакта, либо флуксону (антифлуксону), связанному с антифлуксоном

(флуксоном) на краях контакта. Из анализа этих колебаний следует, что, во-первых, бризерные колебания возможны лишь при превышении амплитудой некоторого минимального значения и, во-вторых, существует обмен энергией между бризерными и плазменными колебаниями, причем плазменные колебания могут иметь любые амплитуды.

Флуксонное колебание описывается решением в виде

$$\Phi = 4\,\mathrm{arctg}\{A\,\mathrm{dn}[\beta(z-z_0), k_f]\,\mathrm{tn}[\Omega(t-t_0), k_g]\},$$

где $k_f^2 = 1 - \dfrac{\beta^2(A^2-1)/A^2-1}{\beta^2(A^2-1)}$; $k_g^2 = 1 - \dfrac{\Omega^2(A^2-1)-1}{\Omega^2(A^2-1)}A^2$, а дисперсионное уравнение имеет вид $\beta = A\Omega$. Данное решение соответствует распространяющемуся одиночному флуксону или кинку, который при каждом отражении от конца линии превращается в антифлуксон (антикинк). Решение может описывать также симметричный случай мультифлуксонного распространения, поскольку согласно граничным условиям $\beta_n = (n/l)K(k_f)$. Для флуксонных колебаний существует максимальное значение амплитуды A, задаваемое уравнением

$$A_{\max}^2/(A_{\max}^2-1) = (n/l)K(\sqrt{1-A_{\max}^{-4}}).$$

Флуксонные колебания сопровождаются монотонным увеличением Φ со временем при любом значении z. Поэтому ток смещения может служить источником энергии для колебаний, поддерживая их даже при наличии потерь.

Уравнение синус-Гордона (10.45) было получено в предположении, что диссипация и другие факторы настолько малы, что ими можно пренебречь. Несмотря на их малость, их учет иногда необходим для рассмотрения реальных джозефсоновских переходов. При этом в левую часть уравнения (10.45) вводят дополнительный член, учитывающий *структурное возмущение*. В общем случае его можно представить в виде

$$\varepsilon f = -\alpha\Phi_t + \beta\Phi_{zzt} - \gamma - \sum \mu_i \delta(z-a_i)\sin\Phi,$$

где первое слагаемое описывает диссипацию вследствие квазичастичного туннелирования, второе — диссипацию, обусловленную движением нормальных электронов вдоль барьера с учетом скин-эффекта, γ — распределенный нормированный ток внешнего смещения контакта, последний член правой части учитывает локальные области сильного сверхтока (микроскопические закоротки, локальные утоньшения барьера и т. д.).

Члены с α, $\beta > 0$ описывают отбор энергии у волны. Ток смещения при $\gamma > 0$ увеличивает энергию, т. е. ускоряет флуксон в направлении $+z$, а антифлуксон в направлении $-z$. Закоротки отбирают энергию волны (энергия запасается в закоротке). Поэтому при приближении к закоротке

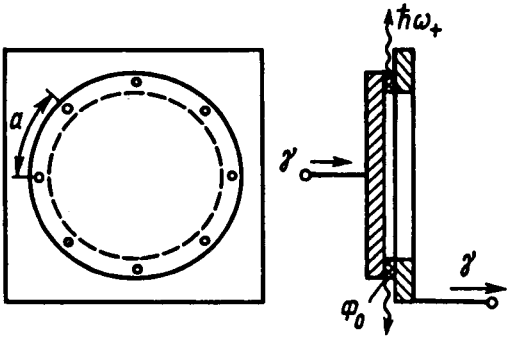


Рис. 10.29. Генератор на флуксонах

происходит замедление как флуксона, так и антифлуксона. Вдали от закоротки стационарная скорость $v_\infty = \pm[1+(4\alpha/\pi\gamma)^2]^{-1/2}$ определяется равновесием между тормозящим действием диссипации и ускоряющим влиянием распределенных токов смещения перехода.

Если ток γ и скорость v_∞ достаточно велики, то флуксон вблизи микрозакоротки замедлится, но не остановится. При малых токах γ он теряет свою кинетическую энергию не доходя до закоротки; в этом случае он «отражается» и постепенно приближается, осциллируя, к состоянию равновесия в точке «закрепления». Пороговое значение, т. е. максимальный ток, при котором возможно закрепление, $\gamma_0 = (\alpha/\pi)\sqrt{8\mu+\mu^2}$. Поэтому полосковая линия, периодически нагружаемая дискретными джозефсоновскими контактами, может выполнять функции «челнока квантов потока», в котором флуксоны могут «закрепляться» в промежутках дискретными контактами. Подача на линию импульса тока смещения подходящей длительности и амплитуды смещает каждый флуксон на одно звено. На этом принципе выполнен *сдвиговый регистр*. В отсутствии диссипации энергии система не может выполнять эту функцию, поскольку исчезнет «закрепление».

Если челноку квантов потока придать кольцеобразную форму (рис. 10.29), то он может выполнять функцию *генератора* излучения. При достаточном токе смещения γ флуксоны не закрепляются между микрозакоротками, а двигаются со средней скоростью v_m и совершают на фоне этого движения колебания с частотой $\omega_\phi = 2\pi v_m/a$, если не учитывать эффект Доплера (a — расстояние между микрозакоротками). Если учесть доплеровский сдвиг, то бóльшая частота

$$\omega_+ = \omega_\phi + k_+ v_m = \omega_\phi \frac{1+v_m\sqrt{1-(1-v_m^2)/\omega_\phi^2}}{1-v_m^2}$$

может быть значительно больше ω_ϕ (при $v_m \to 1$). При $a = \lambda_J = 5$ мкм и средней нормированной скорости $v_m = 0{,}5$ возникает излучение с длиной волны около 75 мкм.

§ 10.11. Микроэлектронные устройства на сверхпроводниках для обработки, запоминания и преобразования информации

В криотроне, рассмотренном в § 10.3, использован фазовый переход сверхпроводника в нормальное состояние. Типичные криотроны на фазовом переходе имеют критические токи около 100 мА, нормальное сопротивление 10^{-3} Ом, а напряжение сигнала 0,1 мВ. Это затрудняет их согласование с внешними электрическими цепями. Время переключения лимитируется индуктивностью ячеек. В схеме среднего уровня интеграции оно имеет значение приблизительно равное 30 нс, т. е. тактовая частота не может превышать 10 МГц. В стационарном состоянии в криотронных схемах энергия не выделяется, но при переключении рассеивается энергия около $3 \cdot 10^{-11}$ Дж. Кроме того, фазовые криотроны требуют термостабилизации с точностью до сотых долей градуса. Их электрические параметры сравнимы с параметрами полупроводниковых интегральных схем, работающих при комнатной температуре. Поэтому разработка схем на фазовых криотронах была прекращена.

В криотроне могут быть использованы также переход от туннелирования куперовских пар к туннелированию нормальных электродов и связанный с этим скачок напряжения (при критическом токе), как и другие эффекты, рассмотренные в § 10.7. Для таких криотронов реальные критические токи порядка 1 мА, нормальные сопротивления около 1 Ом, время переключения 30 пс и рассеиваемая энергия 10^{-17} Дж (предельные значения значительно ниже: 1 пс и 10^{-20} Дж). Параметры криотрона при изменении температур на несколько градусов не меняются (при $T < T_к/2$). Все это делает перспективным их применение для построения логических и запоминающих элементов ЭВМ.

Криотроны на основе слабосвязанных сверхпроводников. Первым и простейшим криотроном на основе слабосвязанных сверхпроводников был *продольный криотрон* на структуре типа «сэндвич» (рис. 10.30, *а*). Шина управления *1* расположена над переходом. Магнитное поле тока управления I_y в изолирующей прослойке *3* направлено вдоль оси *y*. Под действием этого поля возникает градиент разности фаз φ, направленный по оси *x* и определяющий изменение критического туннельного (вентильного) тока $I_{кр}$.

Если сверхпроводящий туннельный ток превышает это предельное значение, то на переходе появляется напряжение, определяемое энергетической щелью используемого сверхпроводника. Например, для олова оно составляет около 1 мВ, а для свинца — около 2,5 мВ, что на порядок больше напряжения на фазовых криотронах.

Зависимость максимального сверхпроводящего тока от тока управления может быть получена, если учесть, что магнитное поле в переходе

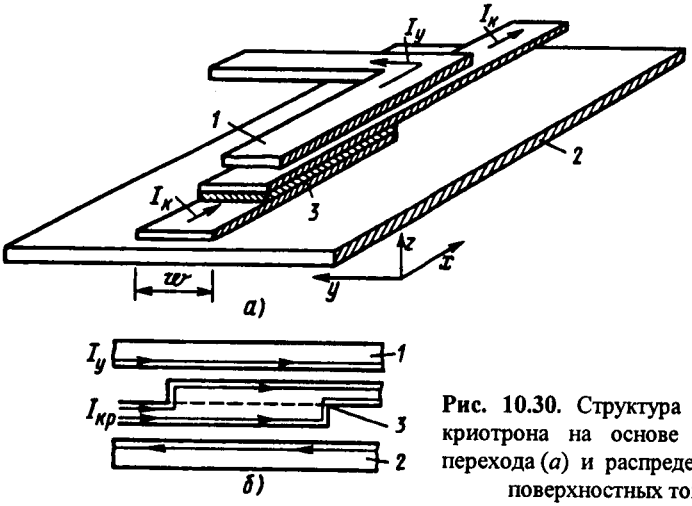


Рис. 10.30. Структура продольного криотрона на основе туннельного перехода (*а*) и распределение в нем поверхностных токов (*б*)

определяется совместным действием обоих токов и ток переключения криотрона

$$I_{кр} \approx I_0(1-|H|/H_0), \qquad (10.50)$$

где I_0 — максимальный критический ток (при $H=0$); H — напряженность результирующего магнитного поля; H_0 — магнитное поле, при котором $I_{кр}=0$.

При этом необходимо также учитывать, что при толщинах сверхпроводящих слоев, бóльших 2λ, линейные плотности поверхностных токов на обращенных друг к другу поверхностях соседних пленок всегда равны по значению и противоположны по направлению (рис. 10.30, *б*). В частности, туннельный ток возвращается через сверхпроводящий экран и ток в экране *2* почти не влияет на магнитное поле в переходе:

$$H=(I_{кр}/2+I_y)/w,^{*} \qquad (10.51)$$

где w — ширина пленок.

Подставив (10.51) в (10.50) и предполагая $I_{кр} \geqslant 0$, получим токи переключения

$$I_{кр}=I_0\frac{1+I_y/(H_0/w)}{1-I_0/(2H_0w)} \text{ при } I_y \leqslant -\frac{I_0}{2},$$

---

* Множитель 1/2 возникает из-за того, что в области туннельного перехода ток распределяется между двумя шинами. При этом нижний сверхпроводник, обращенный к экрану, не влияет на магнитное поле в области перехода. Поля над верхней шиной и шиной управления приблизительно равны нулю.

$$I_{кр} = I_0 \frac{1 - I_y/(H_0 w)}{1 + I_0/(2H_0 w)} \text{ при } I_y \geqslant -\frac{I_0}{2},$$

зависимость которых от управляющего тока приведена на рис. 10.31.

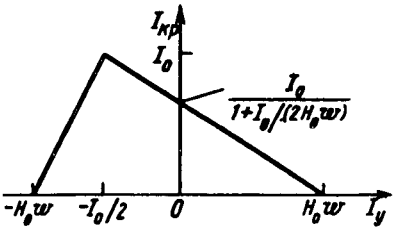


Рис. 10.31. Зависимость тока переключения продольного криотрона от управляющего тока

Реализация двух устойчивых состояний, переключаемых током управления, делает криотрон на слабосвязанных сверхпроводниках потенциальным запоминающим элементом. Скорость переключения ограничивается практически только временем заряда емкости перехода до барьерного напряжения Δ/e и индуктивностью цепей управления. Применение сверхпроводящего экрана позволяет уменьшить индуктивность контура и увеличить скорость переключения.

Возможность управления током криотрона лежит в основе его использования для обработки и запоминания информации.

Принципы *мажоритарной логики* могут быть осуществлены в криотроне с несколькими входами. В качестве примера на рис. 10.32, *а* показана схема продольного криотрона (помеченного на рисунке знаком «×») с тремя независимыми параллельными шинами управления *A*, *B*, *C* и током перехода *I*, несколько меньшим $I_0/[1 + I_0/(2H_0 w)]$. Уровень единичного сигнала управления $I_1 < I_0/2$.

Операция ИЛИ реализуется при совпадении направлений тока в шине управления (управляющего тока) и вентильного тока. В этом случае, если возникает ток в любой из шин управления, вентиль переключается в резистивное состояние (точка *b* на рис. 10.32, *б*).

Операция И реализуется, когда все управляющие токи направлены противоположно току перехода, поскольку при входных сигналах в одной или двух шинах управления вентиль остается в открытом состоянии (точки *c* и *d*). Вентиль переключается в резистивное состояние только в

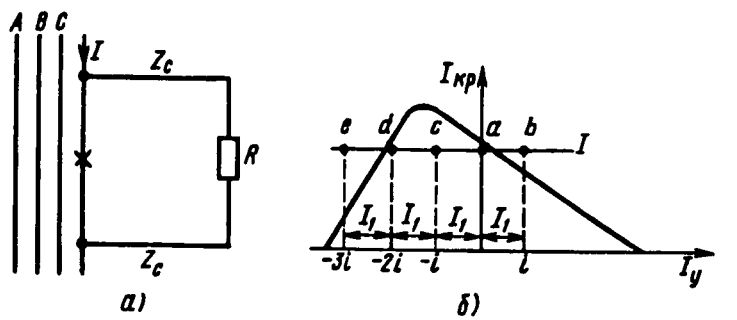


Рис. 10.32. Мажоритарный криотрон с тремя входами и его характеристика

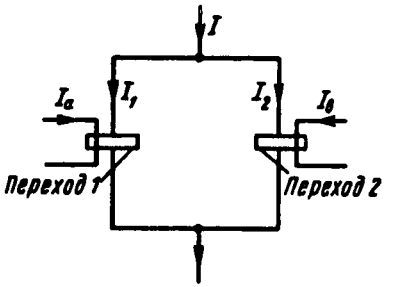


Рис. 10.33. Триггер на двух туннельных переходах

том случае, когда во всех трех управляющих проводниках будут токи I_1 (точка *e*).

Параллельно туннельному переходу присоединена нагрузка *R*, согласованная с волновым сопротивлением Z_c полосковой линии. В нагрузке после переключения вентиля в резистивное состояние возникает ток $I = 2\Delta/(eR)$. Он может быть обнаружен с помощью считывающего проводника (на рисунке не показан).

Таким образом, сигнал на выходе возникает в том случае, когда алгебраическая сумма входных сигналов либо больше нуля, либо меньше $-2I_1$. Подобный мажоритарный элемент позволяет построить полную систему логических функций, т. е. любое логическое или запоминающее устройство.

Схема *триггера* на двух симметричных параллельных туннельных переходах, подключенных к источнику постоянного тока, изображена на рис. 10.33. Триггер представляет собой бистабильный логический элемент, в котором ток может переключаться из одной ветви сверхпроводящего контура в другую путем изменения соответствующих управляющих токов. Предполагается, что сначала подводимый к устройству ток делится поровну между джозефсоновскими контактами ($I_1 = I_2 = I/2$). Если пропустить через управляющий электрод контакта *1* импульс тока I_a, магнитное поле которого уменьшит критический ток до значения, меньшего $I/2$, то произойдет переключение контакта *1* из состояния $u_1 = 0$, $i_1 = I/2$ в состояние $u_1 = 2\Delta/e$, $i_1 = I/2$ при некотором конечном сопротивлении *R* (для квазичастичного тока). Поэтому после прохождения импульса управления этот контакт перейдет в сверхпроводящее состояние $u_1 = 0$ с остаточным током $i_1 = I_r < I/2$ (по гистерезисной петле ВАХ), а контакт *2* — в состояние $u_2 = 0$, $i_2 = I - I_r$. Стабильность этого нового состояния будет обеспечена постоянством магнитного потока в сверхпроводящем контуре. Аналогичную операцию можно провести подачей импульса тока I_b.

Квантование потока в сверхпроводящем кольце триггера позволяет применить его и как элемент памяти, если для хранения «1» и «0» использовать направление циркулирующего тока по или против часовой стрелки. В первом случае $i_1 = -I_s$, $i_2 = I_s$, во втором $i_1 = I_s$, $i_2 = -I_s$.

Продольные криотроны с туннельным переходом обладают ВАХ с гистерезисом. Поэтому элементы с туннельной логикой возвращаются в

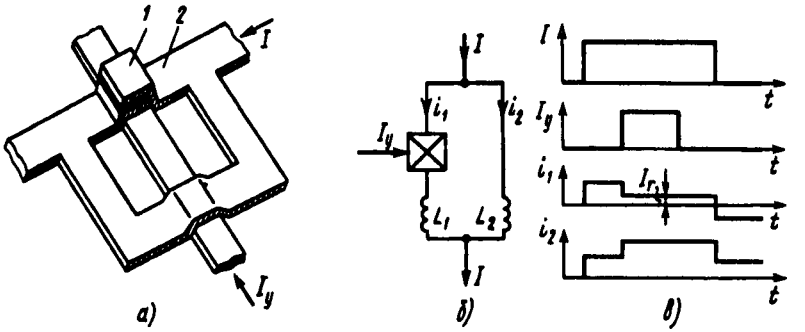


Рис. 10.34. Запоминающая ячейка (*а*), ее эквивалентная схема (*б*) и режим записи (*в*)

исходное состояние только при уменьшении тока питания до нуля, т. е. при импульсном питании.

Квантрон. Высокое быстродействие криотронов в рамках традиционных интегральных схем не приводит к заметному увеличению тактовых частот, поскольку последние ограничены временем распространения сигналов между процессором и оперативным запоминающим устройством. Более полное использование быстродействия когерентных эффектов осуществляется в сверхпроводящих устройствах, в которых информация за такт передается между соседними ячейками малых размеров, и оптимальными являются конвейерные принципы обработки информации. При этом оказываются эффективными управляющие устройства, в которых информация записывается одиночными квантами магнитного потока. Эти устройства называют *квантронами.*

Простой квантрон образован из сверхпроводящего кольца с одним переходом. Эта система имеет несколько устойчивых стационарных состояний, соответствующих числу квантов в магнитном потоке кольца.

Основой элемента сверхбыстродействующей памяти является сверхпроводящий контур с циркулирующим незатухающим током, который и является носителем информации. Информация может кодироваться либо изменением направления циркулирующего тока, либо его наличием или отсутствием. Запись циркулирующего тока в *ячейку памяти* (рис. 10.34, *а*) осуществляется с помощью включенного в контур *2* управляемого вентиля *1*. В отсутствии управляющего тока I_y вентиль открыт, ток I распределяется обратно пропорционально индуктивностям L_1 и L_2 (рис. 10.34, *б*). После появления тока управления I_y вентиль запирается и ток i_1 в нем уменьшается до уровня возвратного тока I_r (рис. 10.34, *в*). После этого вентиль открывается и переходный процесс прекращается при $i_1' = I_r$, $i_2' = I - I_r$. После последовательного отключения токов I_y и I в контуре

останется циркулирующий ток, направление которого определяется направлением тока I. Циркулирующий ток можно найти, исходя из того, что выключение тока I эквивалентно включению такого же тока в обратном направлении. Новые значения токов:

$$i_1'' = -\frac{L_2}{L_1 + L_2} I; \quad i_2'' = -\frac{L_1}{L_1 + L_2} I.$$

Циркулирующий ток

$$I_ц = i_2' + i_2'' = -(i_1' + i_1'') = \frac{L_2}{L_1 + L_2} I.$$

На рассмотренных элементах возможно создание запоминающих устройств, в которых квант потока может неограниченно долго храниться без потребления энергии в виде незатухающего тока в сверхпроводящем контуре.

Запоминающее кольцо с двумя туннельными переходами A_1 и A_2 изображено на рис. 10.35. Состоянию «0» соответствует циркуляция в кольце тока $I_ц$ против часовой стрелки. Продольные переходы A_1 и A_2 управляются с помощью общего разрядного провода *а*. Для записи «1» необходимо включить токи в питающем (I) и управляющих ($I_ц$) проводах, причем в силу симметрии системы ток I делится пополам и накладывается на циркулирующий ток: $I_1 = I/2 + I_ц$, $I_2 = I/2 - I_ц$. Ток записи подбирают так, чтобы при совпадении направлений токов $I_ц$ и I переход переключался (закрывался). Поэтому вентиль A_1 переходит в резистивное состояние. Ток I_2 возрастает до тех пор, пока ток через переход A_1 не снижается до малого значения I_r, при котором этот переход переключится обратно (откроется). Теперь ток $I_2 > I_1$ и при выключении тока I возникнет (для сохранения магнитного потока) ток $I_ц$, циркулирующий по часовой стрелке (состояние «1»). Для последующей записи «0» необходимо при сохранении прежнего тока питания изменить направление тока записи.

Считывание информации без разрушения может быть произведено с помощью туннельного перехода S при одновременном включении токов питания I и считывания $I_{сч}$. Переход S закрывается и переходит в резистивное состояние, поставляя необходимую информацию, если циркуляция тока в кольце происходит только по часовой стрелке и ток $I_2 = I/2 + I_ц$. При этом ток $I_ц$ не меняется.

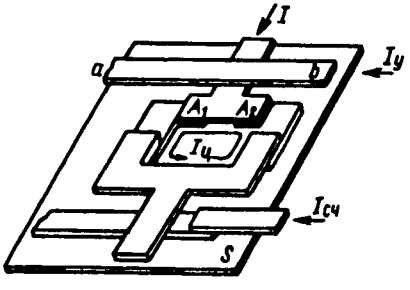


Рис. 10.35. Адресное запоминающее кольцо с нестирающим считыванием

Наряду с магнитным управлением на СКВИД можно воздействовать инжекцией тока

непосредственно в джозефсоновский контакт. Примером элемента *инжекционной логики* может служить схема рис. 10.36, где для развязки входных сигналов использованы резистивные связи. Если входные сигналы отсутствуют, то контакты открыты и ток в нагрузку R не поступает. При появлении хотя бы одного из сигналов x, y, z сначала закрывается контакт *2*, что вытесняет ток I в левую ветвь контура ($R \gg R_2$), приводит к запиранию контакта *1* и вытеснению тока I почти полностью в нагрузку. Контакт *1* использован для исключения попадания выходного сигнала во входные цепи.

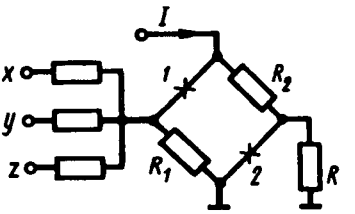


Рис. 10.36. Резистивная схема ИЛИ

Элементы *резистивной логики* менее трудоемки в изготовлении (отсутствует управляющий слой) по сравнению с интерференционными. Они имеют в четыре раза меньшие размеры и за счет меньшей индуктивности — максимальное быстродействие. В некоторых элементах время переключения достигает 10 пс. Достоинством логических элементов на резистивных СКВИДах с инжекцией являются также более широкие рабочие области параметров.

На основе различных систем джозефсоновских логических элементов созданы запоминающие устройства, сумматоры, дешифраторы, регистры, счетчики, тестовые схемы. Основу двоичного *счетчика* на рис. 10.37 составляют двухконтактные СКВИДы. Двоичный режим работы СКВИД достигает за счет одновременной подачи тока $i_1 = \Phi_0/(2M)$ и короткого импульса тока питания I_{g1}. Этот ток распределяется между двумя контактами, суммируясь в контакте D_2 с циркулирующим током $I_ц$, созданным током i_1. Под действием суммарного тока контакт D_2 пе-

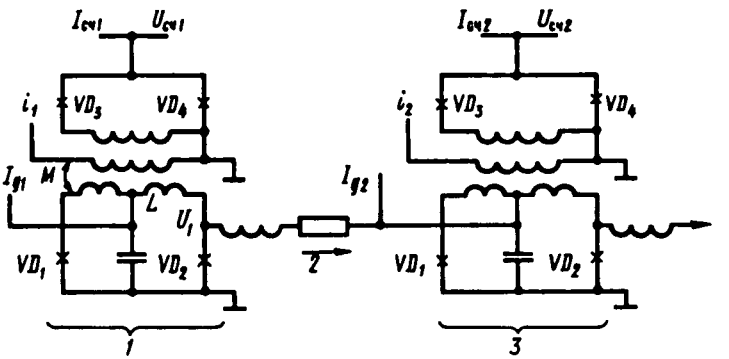


Рис. 10.37. Счетчик:

1 — первый разряд; *2* — импульс переноса; *3* — второй разряд

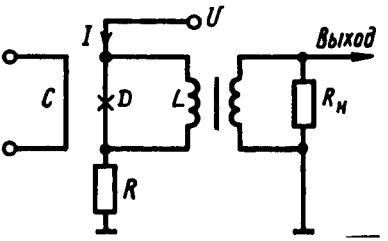


Рис. 10.38. Аналого-цифровой квантовый преобразователь

реходит в резистивное состояние. Возникающий при этом на индуктивности L импульс напряжения меняет направление тока $I_ц$. Поэтому следующий импульс I_{g1} запирает контакт D_1. Напряжение U_1, формирующее импульс переноса во второй каскад, возникает в два раза реже, чем происходит переключение в первом каскаде. СКВИД $D_3 - D_4$ служит для считывания состояния счетчика. Счетчик предназначен для работы на частотах порядка 100 ГГц и восприятия сигналов с энергией 10^{-18} Дж.

Квантование магнитного потока в сверхпроводящем контуре лежит в основе *аналого-цифрового преобразователя* последовательного типа (рис. 10.38). Аналоговый сигнал в виде напряжения U определяет ток контура $I = U/R$. Ток смещения в шине C позволяет управлять критическим током контакта I_c, который определяет требуемое разрешение преобразователя $N = I/I_c$ и удовлетворяет условию $L + L_J \approx \Phi/I_c$. При постепенном возрастании аналогового сигнала большая часть тока будет проходить через переход D ($L_J \ll L$), пока не достигнет значения порогового тока I_c. Тогда произойдет переход в нестационарное состояние и ток сосредоточится в сверхпроводящей петле L. Дальнейшее возрастание напряжения U приводит снова к возрастанию тока через переход до I_c, после чего переход опять перейдет в нестационарное состояние и второй квант магнитного потока войдет в контур. Этот процесс будет повторяться, пока аналоговый сигнал не достигнет своего предельного значения. Изменение потока в кольце может быть обнаружено дифференцированием тока с помощью трансформатора и резистора $R_н$. Для убывающего аналогового сигнала импульсы на нагрузочном резисторе будут иметь противоположную полярность, определяемую квантами потока в контуре. АЦП параллельного действия обладают большим быстродействием, но требуют по одному интерферометру на каждый разряд.

Рассмотренные устройства на основе простого квантрона могут быть использованы только в обычных адресных системах. Они не обеспечивают надежного близкодействия (управления от соседнего элемента), хотя рассеиваемая при переключении энергия может быть доведена до $6 \cdot 10^{-20}$ Дж.

Незначительной модификацией простой квантрон может быть превращен в параметрический квантрон, способный выполнять все операции, необходимые для близкодействующих схем. Для этого достаточно задать в кольце постоянный подмагничивающий поток

$\Phi_{e0} = \Phi_0/2$, например, постоянным внешним током в шине подмагничивания. Тогда состоянием квантрона можно управлять малым внешним полем $\Delta\Phi_e$ от соседних элементов.

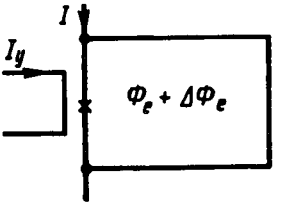


Рис. 10.39. Параметрический квантрон

В параметрическом квантроне (рис. 10.39) используется деформация основной диаграммы состояний (10.43) при изменении критического тока перехода I_c и наличии внешнего сигнала $\Delta\Phi_e$. При включенном управляющем токе I_y криотрона, малом критическом токе ($i < 1$) и $\Delta\Phi_e > 0$ состояние квантрона на диаграмме рис. 10.23, *б* характеризуется точкой A_0. После выключения управляющего тока и соответствующего возрастания критического тока ($i > 1$) состояние квантрона определяет точка A_1, соответствующая замороженному в квантроне одному кванту потока. Если в момент выключения управляющего тока входной сигнал $\Delta\Phi_e$ был отрицателен (точка B_0), то квантрон перейдет в состояние B_1, в котором результирующий поток Φ близок к нулю.

Таким образом, изменяя током управления критический ток контакта, можно ячейку делать либо квантующей ($i > 1$), либо неквантующей ($i < 1$). При этом перевод квантрона из одного устойчивого состояния в другое может осуществлять даже слабый $|\Delta\Phi_e| \ll \Phi_0$ внешний сигнал от соседних ячеек. Необходимая для этого энергия поступает от других источников: тока управления и поля подмагничивания. Ток смещения I смещает исходное состояние равновесия в точку $\Phi_0/2$.

Преимущества параметрического квантрона в полной мере проявляются не столько в том случае, когда он является простым элементом памяти, сколько в возможностях индуктивной передачи информации слабыми сигналами между соседними ячейками.

Передача записанной в нем информации возможна одновременно в несколько ячеек, слабо связанных с ним индуктивно (рис. 10.40, *а*). Если в ячейках *2* в момент такой записи информации критический ток был подавлен ($i < 1$) своим током управления I_2, то после его прекращения состояние этих ячеек определяется знаком $\Delta\Phi_e$, т. е. состоянием квантрона *1*. В схеме трехтактного *регистра сдвига* (рис. 10.40, *б*) в шины управления элементов *1...3* подаются импульсы тактовых токов I_1, I_2, I_3, сдвинутые друг относительно друга (рис. 10.40, *в*). Моменты записи информации в ячейки с соответствующими номерами показаны стрелками в предположении, что элемент с записанной информацией ($I_y = 0$, $i > 1$) был первоначально окружен выключенными элементами ($I_y \neq 0$, $i < 1$), а после перезаписи в соседний элемент исходный элемент выключается то-

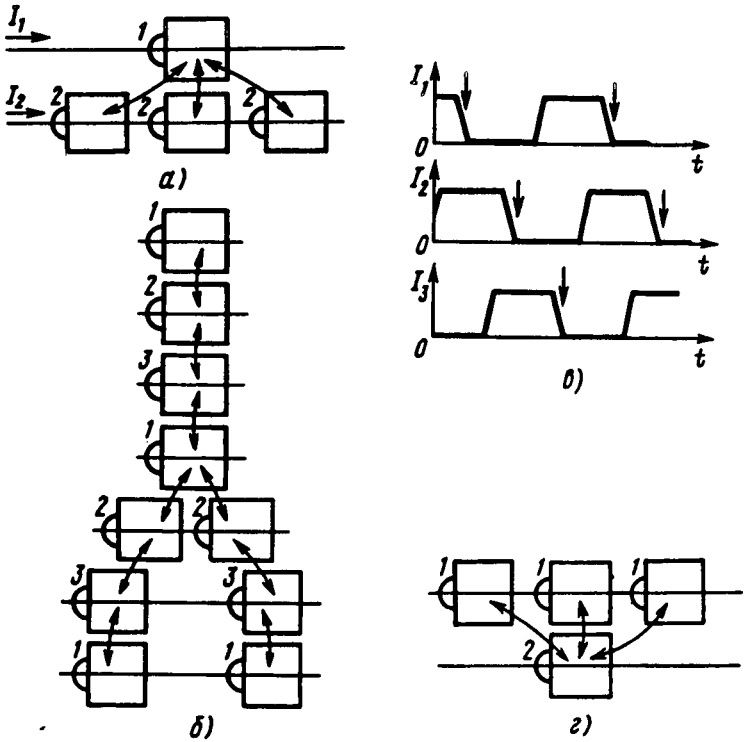


Рис. 10.40. Схемы на связанных квантронах

ком управления. В этом случае осуществляется последовательный сдвиг информации в соседние строки, а также, если нужно, ее размножение.

Операции *мажоритарной логики* могут быть произведены базисной ячейкой (рис. 10.40, *г*), если состояние одного квантрона *2* определяется состоянием заданного числа индуктивно связанных с ним квантронов *1*. Если они включены, а приемный квантрон сначала выключен, то после снятия с него тока управления номер его нового состояния определяется знаком суммы $\Phi_e = \sum k_m \Phi_m$, где k_m — коэффициенты связи; Φ_m — магнитные потоки в исходных ячейках.

Одной из особенностей параметрического квантрона является возможность устойчивой работы при рассеиваемых за одно переключение энергиях, меньших энергии тепловых флуктуаций $k_Б T \approx 10^{-22}$ Дж. Устойчивость квантрона ограничивается лишь квантовыми эффектами. Предельные энергетические и временны́е характеристики квантрона и криотрона формально получаются близкими, т. е. передача информации

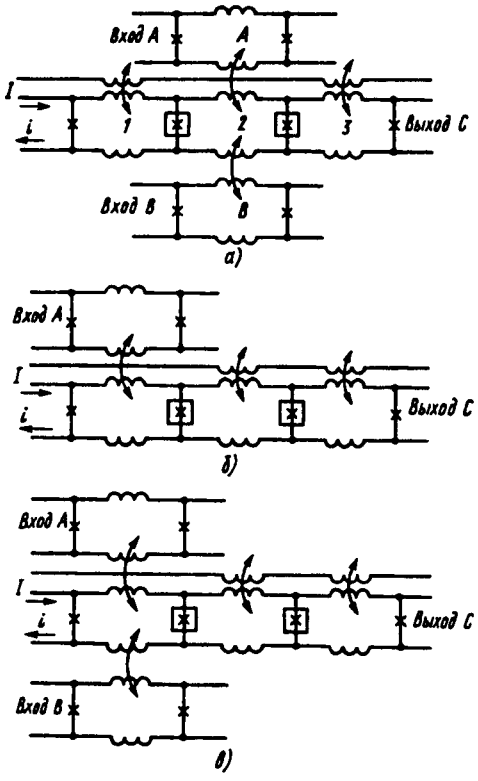


Рис. 10.41. Логические схемы И (*а*), НЕ (*б*) и НЕ-И (*в*) на параметрических квантронах

из одной части схемы в другую через цепочки квантронных и криотронных элементов происходит примерно за одинаковое время и с тем же рассеянием энергии. Преимуществом квантронных схем является возможность попутно с передачей информации производить ее обработку (при надлежащей организации вычислительной структуры). Эти особенности могут быть реализованы лишь в специальных вычислительных структурах, осуществляющих «конвейерные» принципы обработки информации.

Поскольку параметрический квантрон может реализовать операции мажоритарной логики, то его можно рассматривать так же, как базисный элемент однородной вычислительной структуры на одиночных квантах магнитного потока. В таких структурах отсутствует необходимость преобразования, усиления и формирования сигналов, передаваемых от одних функциональных узлов к другим. Сигналы на информационных выходах подобных структур формируются квантами магнитного потока (их наличием или отсутствием).

Примером подобных структур могут служить двухвходовые *логические элементы*, изображенные на рис. 10.41, *а*...*в*. Входная информация в логическом элементе, реализующем функцию И, формируется квантами магнитного потока в верхней и нижней ячейках (Вход *А* и Вход *В*) (их наличием или отсутствием). Функции выходного сигнала (Выход *С*) выполняет квант потока в контуре *3*. Шины токов управления критическими токами контактов, заключенных в рамку, на схеме не показаны. Постоянный ток смещения *I* создает в контурах *1* и *3* магнитные потоки Φ_1 и Φ_3. Амплитуда тактового тока *i* обеспечивает зарождение кванта магнитного потока в ячейке *1*. Реализация функции И предполагает выполнение ус-

ловий $\Phi_1 < \Phi_A + \Phi_B < \Phi_3$, $\Phi_{A,B} < \Phi_1$, где Φ_A, Φ_B — магнитные потоки в контуре *2*, наводящиеся от ячеек *A* и *B*. Если уменьшением тока смещения обеспечить условия

$$\Phi_1 < \Phi_{A,B} < \Phi_3, \quad \Phi_A + \Phi_B < \Phi_3,$$

то та же самая схема будет реализовать функцию ИЛИ. Аналогичным образом логическая схема НЕ-И при соответствующем увеличении тока смещения будет выполнять функцию схемы НЕ-ИЛИ.

Различные сочетания подобных схем образуют функционально полные системы логических устройств на одиночных квантах магнитного потока. Конструктивная однородность и возможность настройки схем позволяет формировать из них *однородные вычислительные среды*. Настройка среды на осуществление той или иной логической функции требует той же самой информации, что и для выполнения этой функции в групповой операции над словами. Как и при групповой обработке информации в ассоциативной памяти, структура ячеек вычислительного устройства на основе однородной вычислительной среды обеспечивает чередование двух основных операций — опроса и записи. Последовательность подобных ячеек образует универсальную перестраиваемую комбинационную схему, которая производит параллельные преобразования, происходящие в обычном ассоциативном процессоре последовательно во времени. Предполагается, что процессор на основе однородной вычислительной среды из криотронов может обеспечить производительность, соответствующую частоте операций до 10 ГГц при выполнении операции сдвига и 1 ГГц при параллельном сложении. Использование параллельной обработки данных приводит к дополнительному увеличению производительности системы.

В элементной базе на основе распределенных джозефсоновских переходов в качестве носителей информации применяют *солитоны*, содержащие квант магнитного потока (джозефсоновские вихри). Системы управляемой генерации, аннигиляции, распространения и переключения вихрей получены с помощью отрезков квазиодномерных переходов с активными и пассивными краями, сверхпроводящими и резистивными соединениями электродов. Эти принципы лежат в основе некоторых систем логических устройств. В десятиступенчатом древовидном дешифраторе при ширине структуры продвижения 5 мкм и плотности критического тока 250 А/см^2 были получены задержка 17 пс на одну ступень и рассеиваемая энергия 10^{-19} Дж. За счет повышения плотности критического тока до 1000 А/см^2 задержка может быть снижена до 5 пс . Экспериментально наблюдаемые скорости продвижения вихрей находились в пределах 10^6—10^7 м/с.

Генерация вихрей в распределенном джозефсоновском переходе возможна двумя способами: 1) прямой инжекцией тока в структуру;

2) пропусканием тока по шине управления, индуктивно связанной со структурой продвижения.

В распределенную структуру длинного джозефсоновского контакта, используемого в *сдвиговом регистре*, вводят неоднородность с периодической зависимостью ее параметров (локальной индуктивности электродов и плотности критических токов контактов) вдоль канала *3*, продвижения квантов (рис. 10.42). Локальные неоднородности создают для флуксонов периодически расположенные вдоль регистра места устойчивого положения равновесия, разделенные потенциальными барьерами. Канал продвижения *7* образуют два вытянутых пленочных электрода *1* и *2*, соединяемые периодически расположенными джозефсоновскими контактами *12*. Индуктивности участков электродов вместе с соседними джозефсоновскими контактами образуют последовательно расположенные одинаковые сверхпроводящие контуры, параметры которых удовлетворяют условию $LI_c \approx \Phi_0$. Управляющие цепи *13* вентилей включены в основную шину *3*, а цепи *15* — в дополнительную шину *11*. Одинаковые шины *3* и *11* сделаны в виде сверхпроводящих пленочных линий, причем их индуктивная связь с электродами *1* и *2* периодически изменяется (на участках *4*, *10* и *6*, *9*) путем изменения расстояния между ними. Индуктивная связь справа от контактов *12* (участки *14* и *16*) больше, чем с левой стороны от них.

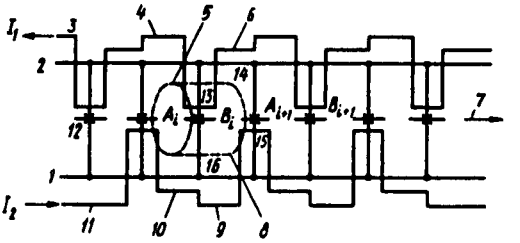


Рис. 10.42. Сдвиговый регистр с параметрическим управлением

Если в начальный момент в отсутствии токов I_1 и I_2 квант потока находится в контуре A_i, а в соседних контурах B_{i-1} и B_i квантов нет, то соответствующий кольцевой ток *5* замыкается через два соседних контакта. При увеличении тока криотроны, образованные контактами и управляющими шинами *13*, закрываются (уменьшается их критический ток). При этом ток в управляемом контакте уменьшается и циркулирующий ток *8* вытесняется в соседний контакт, так как охватываемый им магнитный поток сохраняется. Вследствие большей индуктивной связи шины *3* с контуром B_i возникает смещающий магнитный поток и квант перемещается в контур B_i, если циркулирующий ток направлен по часовой стрелке. Для дальнейшего продвижения кванта направо необходимо подать импульс тока I_2. При этом криотрон в правой ветви контура B_i закрывается и квант перемещается в контур A_{i+1}. Таким образом, продвижение квантов обеспечивает двухтактная система импульсных токов I_1 и I_2. Реверсивный сдвиг информации в регистре достигается изменением полярности токов

I_1 и I_2. Время передвижения кванта потока на один период измеряется десятком пикосекунд.

Для создания сверхпроводниковых БИС и крио-ЭВМ требуется разработка сложной технологии изготовления (отличной от полупроводниковой), компоновки микросхем со сверхвысоким быстродействием и сверхнизкими температурами и защита от сильных тепловых ударов при неизбежных отогревах. Однако работа по созданию крио-ЭВМ продолжается, поскольку полупроводниковые СБИС не всегда могут удовлетворить новым требованиям к вычислительной технике, в первую очередь, по быстродействию и энергопотреблению. Эту проблему позволяют решить элементы, основанные на эффекте Джозефсона. Время их срабатывания достигает 7 пс; есть резервы для его дальнейшего уменьшения. Джозефсоновские элементы расходуют на переключение энергию менее 10^{-16} Дж, что на три-четыре порядка меньше затрат энергии в полупроводниковых элементах. Кроме того, в сверхпроводящих микрополосковых линиях связи между элементами на частотах до 10^{12} Гц практически отсутствует, как и затухание и искажение сигналов несмотря на малые доли поперечного сечения (доли микрометра). К достоинствам элементной базы крио-ЭВМ относятся: хранение информации без затрат энергии, низкий уровень тепловых шумов, малые взаимные помехи и перекрестные наводки, идеальное магнитное экранирование от внешних воздействий, относительно низкие требования к чистоте исходных материалов и высокая радиационная стойкость.

Глава 11
Электромагнитные явления в активных средах

Квантовая электроника — одно из новых направлений современной электроники. Если в основе классической электроники лежит применение электронных потоков в вакууме и газах, то в молекулярной электронике используются свойства *свободных* носителей заряда, молекул и атомов монолитных интегральных схем. Квантовая электроника основана на взаимодействии *связанных* электронов, атомов и молекул с электромагнитным полем.

Взаимодействие электромагнитного поля с веществом в квантовой электронике сопровождается изменением *внутренней* энергии атомов и молекул. Для усиления и генерации электромагнитных колебаний в квантовых приборах используется та высвобождающаяся внутренняя энергия микрочастиц, которая приобретена была ими от отдельного источника. В основе работы квантовых усилителей и генераторов лежит эффект *индуцированного излучения*. Он заключается в том, что под воздействием внешнего электромагнитного поля квантовая система излучает фотоны и переходит в состояние с меньшей энергией. Фотоны, появившиеся в результате этого явления, имеют одну и ту же частоту, фазу, поляризацию и распространяются в одном направлении. Поэтому индуцированное или вынужденное излучение является когерентным.

В квантовых приборах излучение непосредственно не связано с электронным потоком, вследствие чего эти приборы не имеют источников собственных шумов, связанных с изменением плотности электронных потоков. Они могут работать практически на сколь угодно высоких частотах, так как отсутствуют ограничения, вызываемые инерционностью электронов. Квантовая электроника позволяет снять те ограничения, которые накладывает классическая электроника на верхний уровень используемых частот и нижний уровень шумов.

§ 11.1. Квантовый осциллятор

Ряд явлений, наблюдаемых при взаимодействии электромагнитного поля с веществом (дисперсия, резонансное поглощение, индуцированное излучение и т. п.), не может быть изучен и достаточно точно объяснен на основе макроскопической теории поля и классической механики, поскольку такое взаимодействие является взаимодействием двух квантовых систем. Взаимодействие фотонов и элементарных частиц может быть достаточно полно и точно описано только с помощью законов квантовой механики. Использование макроскопической теории поля и классических

методов допустимо лишь в частных предельных случаях при выполнении определенных условий, устанавливаемых квантовой теорией.

Принципиальное различие представлений квантовой и классической механики находит свое выражение в принципе неопределенности, который исключает, в частности, понятие траектории элементарного объекта. Вместо этого квантовая механика утверждает, что состояние любой системы описывает некоторая функция Ψ независимых переменных (например, координат и времени или импульса и времени), которая определяет *распределение вероятностей* измерения соответствующих параметров системы. Функцию Ψ, выражающую единство волновых и корпускулярных свойств материи, называют функцией состояния или *волновой функцией* системы. Волновая функция в нерелятивистском случае должна удовлетворять уравнению Шредингера

$$\hat{\mathcal{H}}\Psi - j\hbar\frac{\partial\Psi}{\partial t} = 0, \tag{11.1}$$

где $\hat{\mathcal{H}}$ — гамильтониан системы.

В качестве примера рассмотрим линейный гармонический осциллятор, который является квантовым аналогом частицы, совершающей малые линейные колебания около положения равновесия. Если частица с массой m удерживается около положения равновесия упругой силой $F = -m\omega^2 x = -\mathrm{d}U/\mathrm{d}x$, то потенциальная энергия частицы $U = m\omega^2 x^2/2$, функция Гамильтона

$$\mathcal{H} = \frac{p^2}{2m} + \frac{m\omega^2}{2}x^2, \tag{11.2}$$

и решение уравнения движения классической физики $m\mathrm{d}^2 x/\mathrm{d}t^2 = -m\omega^2 x$ имеет вид $x = x_0\cos(\omega t + \varphi)$.

В квантовой механике движение частицы описывает плоская волна вероятности в виде $\Psi(x,t) = \psi(x)\mathrm{e}^{j\mathscr{E}t/\hbar}$ и гамильтониан

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + U = \frac{\hat{p}^2}{2m} + \frac{m\omega^2}{2}x^2. \tag{11.3}$$

Поэтому стационарное (не зависящее от времени) уравнение Шредингера приобретает вид

$$\frac{\mathrm{d}^2\psi(x)}{\mathrm{d}x^2} + \frac{2m}{\hbar^2}\left(\mathscr{E} - \frac{m\omega^2}{2}x^2\right)\psi(x) = 0$$

или

$$\hat{\mathcal{H}}\psi = \mathscr{E}\psi. \tag{11.4}$$

Исследование уравнения (11.4) показывает, что его решение (*собственная функция* ψ оператора $\hat{\mathscr{H}}$) будет конечно и однозначно, если полная энергия осциллятора (*собственное значение* оператора $\hat{\mathscr{H}}$)

$$\mathscr{E}_n = \hbar\omega(n + 1/2), \tag{11.5}$$

где $n = 0, 1, 2...$ Следовательно, энергия квантового осциллятора может принимать только *дискретные* значения, причем уровни энергии расположены друг от друга на одинаковых расстояниях, равных $\hbar\omega$. В этом заключается его главное отличие от классического осциллятора.

Наименьшее возможное значение энергии осциллятора ($\mathscr{E}_0 = \hbar\omega/2$) отлично от нуля и по порядку значений соответствует соотношению неопределенностей $\Delta p \Delta x \geqslant \hbar$*. При абсолютном нуле кристаллы находятся в низшем (*основном*) энергетическом состоянии ($n = 0$). Колебания с энергией $\mathscr{E}_0$ объясняют наблюдаемое экспериментально рассеяние света и взаимодействие электронов (сверхпроводимость) при температурах, близких к абсолютному нулю.

В заключение отметим, что в общей теории колебаний пользуются выражением для функции Гамильтона (11.2) классического одномерного гармонического осциллятора *единичной массы* в виде

$$\mathscr{H} = \frac{1}{2}(p^2 + \omega^2 q^2) = \frac{1}{2}(\dot{q}^2 + \omega^2 q^2), \tag{11.6}$$

поскольку его уравнение движения $\mathrm{d}^2 x/\mathrm{d}t^2 + \omega^2 x = 0$ можно представить и в классической гамильтоновской форме:

$$\frac{\partial \mathscr{H}}{\partial p} = \frac{\partial q}{\partial t} = \dot{q} = p, \qquad \frac{\partial \mathscr{H}}{\partial q} = -\frac{\partial p}{\partial t} = -\dot{p} = \omega^2 q,$$

где p — сообщенный импульс осциллятора единичной массы; q — обобщенная координата (в предыдущем рассмотрении $q = x$).

В общем случае тепловые колебания в кристалле и электромагнитное поле можно также представить как совокупность колебаний линейных гармонических осцилляторов.

---

* Вблизи наименьшего значения энергия осциллятора $\mathscr{E} \geqslant \frac{\Delta p^2}{2m} + \frac{m\omega^2}{2}\Delta x^2 \geqslant$ $\geqslant \frac{\Delta p^2}{2m} + \frac{m\omega^2}{2}\left(\frac{\hbar}{\Delta p}\right)^2 \geqslant \hbar\omega$, поскольку правая часть неравенства минимальна при $\Delta p = \sqrt{m\omega\hbar}$.

§ 11.2. Квантование поля

Наиболее общее рассмотрение оптических явлений опирается на квантовые представления, согласно которым электромагнитное излучение есть поток элементарных частиц, называемых фотонами. Напомним основные свойства фотона:

1) фотон не имеет электрического заряда и массы покоя;

2) энергия фотона

$$\mathscr{E}_\nu = h\nu = \hbar\omega = \hbar c_0 k; \tag{11.7}$$

3) импульс фотона

$$\mathbf{p} = \hbar\mathbf{k} = \hbar\omega\mathbf{e}_n / c_0, \tag{11.8}$$

где $\mathbf{k}$ — волновой вектор; $\mathbf{e}_n$ — единичный вектор в направлении распространения волны;

4) спиновое число фотона равно единице;

5) фотон может иметь две независимые поляризации;

6) состояние фотона как микрочастицы описано полностью, если заданы четыре величины, например три проекции импульса ($\hbar k_x$, $\hbar k_y$, $\hbar k_z$) и поляризация (α), индекс α принимает значения 1 и 2; в этом случае говорят о $\mathbf{k}\alpha$-состоянии фотона;

7) локализация фотона в пространстве возможна лишь с точностью до длины волны, поэтому не существует волновой функции фотона в обычном, координатном представлении, хотя квантовая механика и исходит из корпускулярно-волнового дуализма свойств фотона. Дуализм проявляется в соотношениях (11.7) и (11.8), в которых взаимно связаны корпускулярные ($\mathscr{E}_\nu$, $\mathbf{p}$) и волновые (ω, $\mathbf{k}$) характеристики, а также при прохождении одного фотона через интерферометр. Лишь с помощью квантовой теории удалось удовлетворительным образом объяснить эти, казалось бы, противоречивые свойства электромагнитного поля.

Электромагнитное поле при квантовом рассмотрении — это фотонный коллектив, в котором объемная плотность энергии

$$\rho = \sum_\alpha \sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}} N_{\mathbf{k}\alpha}$$

и импульс единицы объема

$$p = \sum_\alpha \sum_{\mathbf{k}} \hbar\mathbf{k} N_{\mathbf{k}\alpha},$$

где $N_{\mathbf{k}\alpha}$ — число фотонов в $\mathbf{k}\alpha$-состоянии в единице объема (с точностью до длины волны).

Переход от электромагнитных волн к фотонным коллективам предполагает разложение поля излучения на осцилляторы, квантование ос-

цилляторов, сопоставление с n-кратно возбужденным осциллятором коллектива из n фотонов в соответствующих $k\alpha$-состояниях.

С математической точки зрения разложение на осцилляторы выражается в представлении векторов поля через наборы дискретных амплитуд всех его компонент, после чего методом вторично квантования осуществляется переход к числам заполнения $N_{k\alpha}$ и операторам уничтожения $\hat{a}$ и рождения $\hat{a}^+$ фотонов, действующим на волновую функцию $\Psi(N_{k\alpha}, t)$. Рассмотрим основные этапы этих переходов на примере анализа поля в резонаторе.

Гамильтониан нормированного электромагнитного поля (2.47) аналогичен гамильтониану (11.6) совокупности осцилляторов с единичной массой. Поэтому квантование электромагнитного поля в резонаторе сводится к квантованию гармонических осцилляторов.

Согласно квантовой механике, каждой моде электромагнитного поля сопоставляется квантовый гармонический осциллятор.

Чтобы упростить обозначения, рассмотрим сначала поле с одной нормальной модой (2.40). В этом случае знак суммы в (2.47) и индекс l можно опустить и гамильтониан поля станет идентичным гамильтониану (11.6) гармонического осциллятора с единичной массой:

$$\hat{\mathcal{H}} = \frac{1}{2}(\hat{p}^2 + \omega^2 \hat{q}^2). \tag{11.9}$$

Чтобы перейти от классического анализа электромагнитного поля к квантовомеханическому, необходимо все входящие в последнее выражение переменные рассматривать как операторы, действующие на соответствующие волновые функции. При этом энергия одной нормальной электромагнитной моды (11.5) представляет собой собственное значение оператора Гамильтона (11.9), явно не зависящего от времени.

В методе вторичного квантования канонические переменные q и p заменяют соответствующими операторами, действующими на волновую функцию $\Psi(N_{k\alpha}, t)$ и имеющими вид

$$\hat{q} = \sqrt{\hbar}\frac{\hat{a}^+ + \hat{a}}{2\omega}, \quad \hat{p} = j\sqrt{\hbar\omega}\frac{\hat{a}^+ - \hat{a}}{2}, \tag{11.10}$$

где неэрмитовы ($\hat{a} \neq \hat{a}^+$), но сопряженные операторы

$$\hat{a} = \frac{\omega\hat{q} + j\hat{p}}{\sqrt{2\hbar\omega}}, \quad \hat{a}^+ = \frac{\omega\hat{q} - j\hat{p}}{\sqrt{2\hbar\omega}}. \tag{11.11}$$

После подстановки их значений (11.10) в (11.9) оператор

$$\hat{\mathcal{H}} = \hbar\omega(\hat{a}^+\hat{a} + 1/2) = \hbar\omega(\hat{N} + 1/2), \tag{11.12}$$

где

$$\hat{N} = \hat{a}^{+}\hat{a} = \frac{\hat{\mathcal{H}}}{\hbar\omega} - \frac{1}{2} \tag{11.13}$$

— эрмитов оператор.

Собственные значения энергии являются теми величинами, которые получаются в экспериментах. Поэтому для сравнения теории с экспериментом необходимо определить собственные значения $\mathscr{E}$ оператора $\hat{\mathcal{H}}$ из решения уравнения (11.4). Эта задача, согласно (11.12), полностью эквивалентна задаче о собственных значениях n оператора $\hat{N}$:

$$\hat{N}\psi_n = n\psi_n, \tag{11.14}$$

причем величина N является наблюдаемой, поэтому собственные функции ψ_n образуют полную ортонормированную систему $\int\psi_n^{*}\psi_m \, dV = \delta_{nm}$, где dV — элемент объема области изменения независимых переменных.

Собственные значения, расположенные равномерно по шкале энергий, определяют по соотношению (11.5). В случае электромагнитного поля возбуждение n-состояния гармонического осциллятора означает наличие n фотонов, каждый из которых обладает энергией $\hbar\omega$, т. е. состояние ψ_n — это состояние с n фотонами.

Оператор (11.13), удовлетворяющий уравнению (11.14), характеризует число фотонов в моде, поскольку среднее значение $\langle N\rangle = \int\psi_n^{*}\hat{N}\psi_n \, dV = \int\psi_n^{*} n\psi_n \, dV = \langle n\rangle$ равно числу фотонов. Поэтому оператор $\hat{N} = \hat{a}^{+}\hat{a}$ в электромагнитном поле называют *оператором числа фотонов*.

Состояние $\psi_{n=0}$, называемое *вакуумным*, соответствует нулевому энергетическому уровню, на котором отсутствуют фотоны, но имеется нулевая энергия $\hbar\omega/2$. Это состояние поля называется *состоянием с нулевыми флуктуациями поля*. Когда поле образовано суперпозицией разных l-состояний, каждый тип колебаний дает вклад в энергию, равный $\hbar\omega_l/2$. Нулевые флуктуации поля могут давать эффект, например, при использовании электрического заряда в качестве измерительного инструмента. При взаимодействии нулевых флуктуаций поля с электроном атома водорода они определяют лембовское смещение уровней $2P_{1/2}$ и $2S_{1/2}$. Нулевые флуктуации поля «индуцируют» спонтанное излучение электрона при переходе из состояния $2P_{1/2}$ в состояние $2S_{1/2}$. Они вызывают также то *спонтанное излучение* в мазерах, лазерах и параметрических усилителях, которое является источником *квантовых шумов*. Нуле-

вые флуктуации определяют и естественную ширину линий излучения атомов. Спонтанное излучение является чисто квантовым эффектом, поскольку при классическом подходе возбужденный свободный атом может существовать вечно.

При большом числе фотонов n дискретный характер энергии электромагнитного поля (11.5) становится незаметным и квантовые соотношения переходят в классические.

§ 11.3. Взаимодействие электромагнитного излучения с веществом

В квантовой электронике используются высокие плотности излучения, соответствующие высоким числам заполнения и фотонным потокам при больших квантовых числах. В соответствии с принципами квантовой механики это означает, что, если пренебречь шумовыми явлениями и спонтанным излучением, то электромагнитные поля в этих условиях описываются классически. Поэтому с достаточно хорошим приближением в данном параграфе будет использован подход, при котором поле описывается классически, а среда — квантовомеханически.

Если исходить из предположения, что в атоме имеется единственный электрон с зарядом e и массой m, взаимодействующий с бесконечно тяжелым ядром с эффективным потенциалом $U(\mathbf{r})$, и если пренебречь спином и релятивистскими эффектами, то для электрона во внешнем электромагнитном поле функция Гамильтона

$$\mathcal{H} = \frac{m}{2}\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)^2 + U - e\varphi = \sum p_i \frac{\mathrm{d}q_i}{\mathrm{d}t} - \mathcal{L}\left(q, \frac{\mathrm{d}q}{\mathrm{d}t}, t\right), \qquad (11.15)$$

где r — положение электрона относительно ядра; $p = m\dfrac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} - e\mathbf{A}$ — канонический импульс электрона; φ, $\mathbf{A}$ — скалярный и векторный потенциалы внешнего электромагнитного поля; индекс i характеризует направление координатных осей (например, в прямоугольной системе $\sum_i p_i \dfrac{\mathrm{d}q_i}{\mathrm{d}t} = p_x \dfrac{\mathrm{d}q_x}{\mathrm{d}t} + p_y \dfrac{\mathrm{d}q_y}{\mathrm{d}t} + p_z \dfrac{\mathrm{d}q_z}{\mathrm{d}t}$);

$$\mathcal{L} = \frac{1}{2} m \left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)^2 - U + e\varphi - e\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\mathbf{A} \qquad (11.16)$$

— функция Лагранжа.

Даже в дальней ультрафиолетовой области длины волн превышают сотни ангстрем, тогда как радиус боровской орбиты основного состояния

атома водорода равен 0,529 Å. В этих условиях пространственное изменение векторного потенциала **A** в пределах атома незначительно и его удобно разложить в ряд Тейлора:

$$\mathbf{A}(\mathbf{R}+\mathbf{r},t)=\mathbf{A}(\mathbf{R},t)+(\mathbf{r},\nabla_R)\mathbf{A}(\mathbf{R},t)+..., \qquad (11.17)$$

где **R** — положение ядра относительно начала координат; $\nabla_R = \mathbf{e}_x\frac{\partial}{\partial R_x}+\mathbf{e}_y\frac{\partial}{\partial R_y}+\mathbf{e}_z\frac{\partial}{\partial R_z}$; $\partial/\partial R_i$ — частная производная по направлению i-й координаты, взятая в точке расположения ядра.

Каждое из составляющих разложения (11. 17) приводит к появлению в гамильтониане членов, соответствующих электрическому дипольному, магнитному дипольному, электрическому квадрупольному и другим взаимодействиям с внешним электромагнитным полем. При электрическом дипольном приближении используют только первый член **A**(**R**, t) разложения (11.17), пренебрегая пространственным изменением векторного потенциала в пределах атома, что в большинстве случаев обеспечивает хорошую точность даже для оптических полей, простирающихся в далекую ультрафиолетовую область. К магнитному дипольному и другим приближениям приходится обращаться в том случае, когда правила отбора запрещают квантовые переходы, обусловленные электрическим дипольным взаимодействием.

В подобных расчетах обычно применяют кулоновскую калибровку векторного потенциала, при которой в свободной от зарядов области скалярный потенциал $\varphi = 0$ и напряженность электрического поля

$$\mathbf{E}=-\frac{\partial\mathbf{A}}{\partial t}-\operatorname{grad}\varphi=-\frac{\partial\mathbf{A}}{\partial t}. \qquad (11.18)$$

Дальнейшее упрощение математических операций можно получить, если к выражению (11.16) прибавить член $e\frac{\mathrm{d}}{\mathrm{d}t}\{\mathbf{r}\cdot\mathbf{A}(\mathbf{R},t)\}$, поскольку уравнения движения системы не изменяются при добавлении к функции Лагранжа полной производной по времени*. Таким образом, в дипольном приближении с учетом уравнений (11.15), (11.16) и (11.18) функция Гамильтона

$$\mathscr{H}=\sum_i p_i\frac{\mathrm{d}r_i}{\mathrm{d}t}-\frac{m}{2}\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)^2+U-e\varphi+e\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\mathbf{A}(\mathbf{R},t)-e\frac{\mathrm{d}}{\mathrm{d}t}\{\mathbf{r}\cdot\mathbf{A}(\mathbf{R},t)\}=$$

$$=\sum_i p_i\frac{\mathrm{d}r_i}{\mathrm{d}t}-\frac{m}{2}\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)^2+U-e\mathbf{r}\frac{\partial\mathbf{A}(\mathbf{R},t)}{\partial t}=\frac{\mathbf{p}^2}{2m}+U-\boldsymbol{\mu}\mathbf{E}(\mathbf{R},t), \qquad (11.19)$$

---

* **Гольдшейн Г.** Классическая механика. —М.: Гостехиздат, 1957.

где $\mathbf{p} = m\dot{\mathbf{r}}$ — канонический импульс ($p_i = \mathrm{d}\mathscr{L}/\mathrm{d}q_i$); $\boldsymbol{\mu} = -e\mathbf{r}$ — электрический момент и учтено, что на ядре

$$\frac{\mathrm{d}\mathbf{A}(\mathbf{R},t)}{\mathrm{d}t} = \frac{\partial \mathbf{A}(\mathbf{R},t)}{\partial t} = -\mathbf{E}(\mathbf{R},t).$$

Гамильтониан системы многих частиц в области, достаточно малой по сравнению с длиной электромагнитной волны, можно получить обобщением (11.19) путем простого перехода к сумме энергий всех электронов:

$$\hat{\mathscr{H}} = \frac{1}{2m}\sum \hat{p}_k^2 + U(\mathbf{r}_k) - \hat{\boldsymbol{\mu}}\mathbf{E} = \hat{\mathscr{H}}_0 + \hat{\mathscr{H}}', \tag{11.20}$$

где $U(\mathbf{r}_k)$ — потенциальная энергия, определяемая кулоновским взаимодействием между отдельными электронами с координатами $\mathbf{r}_k$ и их взаимодействием с ядрами; $\boldsymbol{\mu} = -e\sum \mathbf{r}_k$ — *электрический момент системы*;

$\hat{\mathscr{H}}_0 = \frac{1}{2m}\sum_k \hat{p}_k^2 + U(\mathbf{r}_k)$ — *гамильтониан невозмущенной системы* (в отсутствие внешнего электромагнитного поля);

$$\hat{\mathscr{H}}' = -\hat{\boldsymbol{\mu}}\cdot\mathbf{E} \tag{11.21}$$

— *гамильтониан взаимодействия*.

Перейдем к матричному представлению гамильтониана взаимодействия. Определим его матричные элементы с помощью собственных функций ψ_n невозмущенного гамильтониана, т. е. удовлетворяющих уравнению

$$\hat{\mathscr{H}}_0\psi_n = \mathscr{E}_n\psi_n. \tag{11.22}$$

При таком подходе предполагается, что

$$\hat{\mathscr{H}}'\psi(\mathbf{x}) = \hat{\mathscr{H}}'\sum_n c_n\psi_n(\mathbf{x}) = \sum_n c_n\hat{\mathscr{H}}'\psi_n(\mathbf{x}) = \sum_m b_m\psi_m(\mathbf{x}) = \varphi(\mathbf{x}),$$

т. е. $\sum_m b_m\psi_m(\mathbf{x}) = \sum_n c_n\hat{\mathscr{H}}'\psi_n(\mathbf{x})$. Умножив левую и правую части этого равенства на $\psi_m^*(\mathbf{x})$ и проинтегрировав их по всей области изменения независимых переменных $\mathbf{x}$, получим $b_m = \sum_n \mathscr{H}'_{mn}c_n$, где $\mathscr{H}'_{mn} = \int\psi_m^*(\mathbf{x})\hat{\mathscr{H}}'\psi_n(\mathbf{x})\mathrm{d}V$ — матричный элемент; $\mathrm{d}V$ — элемент объема области изменения независимых переменных.

Для электрического дипольного взаимодействия матричные элементы оператора (11.21) можно представить в виде

$$\mathscr{H}'_{mn} = -\int \psi_m^*(\mathbf{x})\hat{\boldsymbol{\mu}}\mathbf{E}\psi_n(\mathbf{x})\,dV = -\boldsymbol{\mu}_{mn}\mathbf{E}, \tag{11.23}$$

поскольку здесь не учитывается пространственное изменение электрического поля в пределах атомных размеров, $\mathbf{E}$ является только функцией времени и может быть вынесено из-под знака интеграла. Таким образом, оператор $\hat{\mathscr{H}}'$ зависит от матричных элементов

$$\boldsymbol{\mu}_{mn} = \int \psi_m^* \hat{\boldsymbol{\mu}} \psi_n \, dV \tag{11.24}$$

оператора электрического момента $\hat{\boldsymbol{\mu}}$, а напряженность электрического поля является общим сомножителем.

Матрица элементов $\boldsymbol{\mu}_{mn}$ позволяет определить, между какими состояниями возможны электрические дипольные переходы. Если матричный элемент между какими-либо двумя состояниями равен нулю, то переходы между ними запрещены (в электрическом дипольном приближении).

Например, если два собственных состояния ψ_m и ψ_n, принадлежащих различным собственным значениям энергии $\mathscr{E}_m$ и $\mathscr{E}_n$, имеют *одинаковую четность**, то матричный элемент

$$\boldsymbol{\mu}_{mn} = -e\int \psi_m^* \mathbf{r} \psi_n \, dV = 0,$$

поскольку подынтегральное выражение является нечетной функцией во всем пространстве. Отсюда следует одно из правил отбора: электрический дипольный переход может происходить только между состояниями противоположной четности. Согласно другому правилу отбора, магнитные дипольные переходы могут происходить только между состояниями одинаковой четности.

Нетрудно также видеть, что диагональные элементы оператора дипольного момента $\hat{\boldsymbol{\mu}}$, а следовательно, и гамильтониана взаимодействия (11.21) обращаются в нуль, если собственные состояния имеют определенную четность и невырождены:

$$\boldsymbol{\mu}_{nn} = -e\int \psi_n^* \mathbf{r} \psi_n \, dV = 0. \tag{11.25}$$

Поглощение и излучение электромагнитных волн, а также дисперсионные свойства вещества в ультрафиолетовой, видимой и инфракрасной областях спектра обычно обусловлены электрическими дипольными переходами между двумя энергетическими уровнями. Поэтому рассмотрим более подробно электрическое дипольное взаимодействие классического электромагнитного поля с *двухуровневой квантовой системой*. В этом случае предполагают, что среда состоит из набора атомов или молекул,

---

* $\psi(-\mathbf{r}) = \psi(\mathbf{r})$ или $\psi(-\mathbf{r}) = -\psi(\mathbf{r})$.

обладающих двумя невырожденными состояниями противоположной четности ψ_1 и ψ_2.

Основной гамильтониан такой невозмущенной системы, удовлетворяющий уравнению для собственных значений $\hat{\mathcal{H}}_0\psi_n = \mathcal{E}_n\psi_n$, имеет матричные элементы $(\mathcal{H}_0)_{mn} = \int\psi_m^*\hat{\mathcal{H}}_0\psi_n\,dV = \mathcal{E}_n\delta_{mn}$ и, следовательно, может быть записан в виде

$$\hat{\mathcal{H}}_0 = \begin{bmatrix} \mathcal{E}_1 & 0 \\ 0 & \mathcal{E}_2 \end{bmatrix}. \tag{11.26}$$

Оператор электрического момента для двухуровневой системы, обладающей невырожденными собственными состояниями противоположной четности, согласно (11.25),

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} 0 & \boldsymbol{\mu}_{12} \\ \dot{\boldsymbol{\mu}}_{12} & 0 \end{bmatrix}, \tag{11.27}$$

причем $\boldsymbol{\mu}_{21} = \dot{\boldsymbol{\mu}}_{12}$, поскольку оператор соответствует физически наблюдаемой величине и должен быть эрмитовым.

Соответствующий выражению (11.27) гамильтониан взаимодействия (11.21)

$$\hat{\mathcal{H}}' = -\hat{\boldsymbol{\mu}}\mathbf{E} = -\mu_x E_x - \mu_y E_y - \mu_z E_z = \begin{bmatrix} 0 & -\mu_\alpha E_\alpha \\ -\dot{\mu}_\alpha E_\alpha & 0 \end{bmatrix}, \tag{11.28}$$

где для упрощения записи в матрице опущены индексы 1, 2, использованные в (11. 27); $\alpha = x, y, z$ обозначает координатное направление (под повторяющимися индексами подразумевается суммирование), так что для оператора $\hat{\boldsymbol{\mu}}$ матричный элемент $\mu_\alpha = (\mu_\alpha)_{12} = \mathbf{e}_\alpha\int\psi_1^*\hat{\boldsymbol{\mu}}\psi_2\,dV$.

Различные ориентации молекул приводят к изменению квантово-механического среднего $\langle\mu_\alpha\rangle$ от молекулы к молекуле. Если в объеме V имеется $\mathcal{N}$ молекул, то полная макроскопическая поляризация

$$P_\alpha = \frac{1}{V}\sum_{i=1}^{\mathcal{N}}\langle\mu_\alpha\rangle^i = N\overline{\langle\mu_\alpha\rangle},$$

где $N = \mathcal{N}/V$ — число атомов или молекул в единице объема; черта сверху указывает усреднение по всем молекулам. В простейшем случае — газа из произвольно ориентированных анизотропных молекул — все направления α равноправны и

$$\overline{|\mu_x|^2} + \overline{|\mu_y|^2} + \overline{|\mu_z|^2} = 3\overline{|\mu_x|^2} = 3\overline{|\mu_\alpha|^2} = \overline{|\mu_{12}|^2}. \tag{11.29}$$

§ 11.4. Уравнения движения для дипольного момента

Нелинейные дифференциальные уравнения, описывающие динамику взаимодействия электромагнитного поля со средой, в квантовой механике выводят из уравнения Шредингера (11.1) с учетом гамильтониана дипольного взаимодействия (11.21) и оператора энергии внутренних взаимодействий, например, между атомом и решеткой или при столкновениях молекул. Влияние внутренних взаимодействий учитывают в первом приближении релаксационными членами с соответствующим временем релаксации. Так, постоянная времени T_1 характеризует время, в течение которого система переходит из возмущенного состояния в состояние энергетического равновесия с окружающей средой вследствие взаимодействия между ионом и решеткой,молекул при столкновениях в газе или в жидкости, а также под воздействием спонтанного излучения. Постоянную T_1 называют *временем продольной релаксации*, а в случае взаимодействия с решеткой — *временем спин-решеточной* или *диполь-решеточной релаксации*. Продольное время релаксации — это среднее время жизни частицы на энергетическом уровне.

Постоянная времени T_2 определяет время, через которое молекулы будут находиться в различных состояниях; она характеризует ширину линии перехода между этими состояниями. На это время влияют кроме факторов, дающих вклад в T_1, также одновременные встречные переходы между одной парой уровней при упругих столкновениях, когда происходит расфазировка. Постоянную времени $T_2 \ll T_1$ называют *временем поперечной*, *диполь-дипольной* или *спин-спиновой релаксации*.

Время поперечной релаксации T_2 есть среднее время, в течение которого ранее синфазно осциллирующие частицы приобретут случайные по отношению друг к другу фазы, т. е. время T_2 можно рассматривать как время фазовой памяти или время передачи энергии от частицы к частице по уровню.

В процессе релаксации происходит обмен энергией между подсистемой рассматриваемых частиц и всей системой в целом, заканчивающейся термодинамическим равновесием между ними.

При рассмотрении электромагнитного поля будем разделять полную поляризацию $P_{\text{полн}} = P_g + P$ на две компоненты. Одна из них (P) учитывает поляризацию в результате воздействия какого-либо определенного типа электромагнитных колебаний (например, резонансного перехода электрона с одного энергетического уровня на другой), а другая (P_g) определяется остальными типами колебаний (нерезонансным взаимодействием). Вклад этой компоненты будем учитывать диэлектрической проницаемостью ε (ее иногда называют фоновой диэлектрической проницаемо-

стью), измеряемой на частотах выше или ниже резонансной. В этом случае смещение $D = \varepsilon_0 E + P_{\text{полн}} = \varepsilon E + P$.

Различие между поляризационной составляющей P, возбуждающей электрическое поле, и полной поляризацией $P_{\text{полн}}$, так же как между локальным макроскопическим электрическим полем напряженностью $E_{\text{лок}} = \sqrt{L}E$ и макроскопическим полем E, связано в общем случае с влиянием поляризуемого вещества на локальное электрическое поле, которое действует на данный атом или молекулу. Различие между локальным и макроскопическим полем учитывает коэффициент Лоренца, который для изотропной среды $L = [(\varepsilon_r + 2)/3]^2$.

Вывод уравнений движения дипольного перехода приводится в курсе квантовой механики. В тех случаях, когда можно воспользоваться соотношением (11.29) и (как это обычно бывает) ширина линии перехода (T_2^{-1}) в двухуровневой системе мала по сравнению с резонансной частотой $\Omega = (\mathscr{E}_2 - \mathscr{E}_1)\hbar^{-1}$, уравнение движения усредненного электрического момента имеет вид

$$\frac{\partial^2 P}{\partial t^2} + \frac{2}{T_2}\frac{\partial P}{\partial t} + \Omega^2 P = \frac{2\Omega L|\boldsymbol{\mu}_{12}|^2}{3\hbar}(N_1 - N_2)E. \qquad (11.30)$$

Таким образом, поляризованный ансамбль двухуровневых квантовых систем ведет себя как осциллятор с собственной частотой Ω, возбуждаемый электрическим полем E с коэффициентом связи, пропорциональным разности населенностей ($N_1 - N_2$) энергетических уровней. Из выражения (11.30) также следует, что в отсутствии электромагнитного поля начальная поляризация затухает с постоянной времени, определяемой временем поперечной релаксации T_2 вследствие расфазировки индивидуальных диполей в результате их взаимодействия.

Разность населенностей энергетических уровней является мерой энергии, запасенной в квантованной среде. Она определяет мощность

$$\mathscr{P} = -\frac{\partial}{\partial t}\left[\hbar\Omega\frac{N_1 - N_2}{2}\right],$$

поглощаемую средой в единице объема, поскольку при переходе одного атома с первого на второй уровень разность уменьшается на 2, а запасенная в единице объема энергия увеличивается на $\hbar\Omega$.

Динамику изменения разности населенностей энергетических уровней описывает приближенное (при $\Omega T_2 \gg 1$) уравнение

$$\frac{\partial}{\partial t}(N_1 - N_2) + \frac{(N_1 - N_2) - (N_1 - N_2)^{\text{c}}}{T_1} \approx -\frac{2}{\hbar\Omega}\mathbf{E}\frac{\partial \mathbf{P}}{\partial t}, \qquad (11.31)$$

где $(N_1 - N_2)^c$ — стационарное значение разности населенностей энергетических уравнений. Уравнение (11.31) выражает баланс мощности, где изменение разности населенности определяет мощность $\mathbf{E}\frac{\partial \mathbf{P}}{\partial t}$, затрачиваемую полем на поляризацию среды. Второй член уравнения учитывает обмен энергии с окружающей средой. Начальная разность населенностей в отсутствие поля уменьшается, стремясь к равновесному (стационарному) значению с постоянной времени, равной времени продольной релаксации T_1.

Уравнения (11.30) и (11.31) для поляризации и разности населенностей описывают поведение среды в электромагнитном поле. Для полного математического описания взаимодействия поля с веществом в виде замкнутой системы уравнений необходимо выражения (11.30) и (11.31) дополнить уравнением поля, которое учитывает обратное воздействие динамических свойств среды на поле.

Уравнение электромагнитного поля может быть получено из уравнений Максвелла. Для изотропной среды без свободных зарядов ($\rho = 0$) источников ($E_{\text{стор}} = 0$) и намагниченности ($\mu = \mu_0$) совместное решение системы (6.18) приводит к уравнению

$$\operatorname{rot}\operatorname{rot}\mathbf{E} + \mu_0\sigma\frac{\partial \mathbf{E}}{\partial t} + \frac{\varepsilon_r}{c_0^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = -\mu_0\frac{\partial^2 \mathbf{P}}{\partial t^2}. \qquad (11.32)$$

При рассмотрении резонаторов уравнение поля можно получить в другой форме, если в волновое уравнение (11.32) подставить разложение по нормальным типам колебаний (2.50) и использовать выражения (2.55) и (2.56):

$$\ddot{p}_l + \frac{\mu_0\sigma c_0^2}{\sqrt{\varepsilon_r}}\dot{p}_l + \omega_l^2 p_l = \frac{1}{\sqrt{\varepsilon}}\int \ddot{p}_l E_l(R)\,\mathrm{d}V.$$

Если в резонаторе доминирует единственный тип колебаний $\mathbf{E}_l(\mathbf{R})$ и напряженность полного поля $\mathbf{E} = -p_l(t)\mathbf{E}_l(\mathbf{R})/\sqrt{\varepsilon}$, то предыдущее уравнение можно записать в виде

$$\ddot{\mathbf{E}} + \frac{1}{\tau_c}\dot{\mathbf{E}} + \omega_c^2\mathbf{E} = -\frac{1}{\varepsilon}\mathbf{E}_l(\mathbf{R})\int \ddot{\mathbf{P}}\mathbf{E}_l(\mathbf{R})\,\mathrm{d}V,$$

где $\tau_c = \sqrt{\varepsilon_r}/(\mu_0 c_0^2\sigma)$.

Дальнейшее упрощение уравнения возможно, когда источник поляризации целиком заполняет резонатор, поляризация сама задается полем и поэтому имеет такое же пространственное распределение, как и нормальный тип колебаний:

$$\ddot{\mathbf{E}}+\frac{1}{\tau_c}\dot{\mathbf{E}}+\omega_c^2\mathbf{E}=-\frac{1}{\varepsilon}\ddot{\mathbf{P}}. \qquad (11.33)$$

Уравнение поля в форме (11.32) или (11.33) вместе с уравнениями (11.30) и (11.31) образует замкнутую систему, достаточную для полного описания электрического дипольного взаимодействия классического электромагнитного поля с двухуровневой квантованной средой.

В основе классического аналога индуцированного излучения лежат свойства классического гармонического осциллятора, совершающего свободные колебания под воздействием внешнего резонансного излучения. Если колебания осциллятора и поля находятся в противофазе, то мощность, забираемая осциллятором из внешнего электромагнитного поля, становится отрицательной. Таким образом, осциллятор начинает передавать энергию внешнему полю под влиянием самого внешнего поля. Возникающее излучение можно трактовать как индуцированное, а его когерентный характер как следующий из классической гармонической природы осциллятора. Однако непротиворечивую классическую теорию индуцированного излучения можно создать, лишь рассматривая радиационные процессы излучения сгустков волн в ансамбле классических осцилляторов, когда расстояния между частицами в сгустках много меньше длины волны.

§ 11.5. Резонансные процессы

Взаимодействие электромагнитного поля с веществом можно рассматривать, согласно (11.30) и (11.33) как взаимодействие осцилляторов с частотами ω_c и Ω. При близости этих частот взаимодействие становится резонансным. К одному из резонансных процессов относится *поглощение* электромагнитной волны, связанное с переходами между любой парой энергетических уровней. Рассмотрим его в стационарном режиме.

В плоской волне изменение напряженности и поляризации в зависимости от времени и координаты имеет вид (1.22). Переход к комплексной форме превращает дифференциальное уравнение (11.30) в алгебраическое, решение которого для комплексных амплитуд вблизи резонансной частоты ($\omega \approx \Omega$) имеет вид

$$\dot{\mathbf{P}}_0=\frac{\frac{2\Omega}{\hbar}L\frac{|\boldsymbol{\mu}_{12}|^2}{3}(N_1-N_2)}{\Omega^2-\omega^2+j2\omega/T_2}\dot{\mathbf{E}}_0\approx\frac{L}{\hbar}\frac{|\boldsymbol{\mu}_{12}|^2}{3}\frac{N_1-N_2}{\Omega-\omega+j/T_2}\dot{\mathbf{E}}_0=\underline{\chi}\dot{\mathbf{E}}_0, \qquad (11.34)$$

если разность населенностей не зависит от времени. Отсюда восприимчивость

$$\underline{\chi} = \chi'' + j\chi'' = \frac{\pi}{\hbar} L \frac{|\mu_{12}|^2}{3} (N_1 - N_2) \underline{g}(\omega, \Omega), \tag{11.35}$$

где частотную характеристику

$$\underline{g}(\omega, \Omega) = \frac{1}{\pi} \frac{1}{\Omega - \omega + j/T_2} =$$

$$= \frac{1}{\pi} \frac{\Omega - \omega}{(\Omega - \omega)^2 + 1/T_2^2} - j \frac{1}{\pi} \frac{1/T_2}{(\Omega - \omega)^2 + 1/T_2^2} \tag{11.36}$$

называют *комплексной функцией Лоренца.*

Для плоской волны после подстановки выражения (11.34) в (11.32) можно найти комплексное волновое число

$$\underline{k}^2 = (\varepsilon_r + \underline{\chi}_r)\omega^2 c_0^{-2},$$

если пренебречь влиянием всех переходов, кроме рассматриваемого (*резонансного*). Для не очень сильного перехода, когда $\chi \ll \varepsilon$ и $\sqrt{1 + \chi/\varepsilon} \approx 1 + \chi/2\varepsilon$,

$$\underline{k} = \beta - j\alpha \approx \frac{1}{\sqrt{\varepsilon_r}} \frac{\omega}{c_0} \left(\varepsilon_r + \frac{\chi_r'}{2} \right) + j \frac{1}{\sqrt{\varepsilon_r}} \frac{\omega}{c_0} \chi_r''. \tag{11.37}$$

Уравнение (11.37) описывает *дисперсионные свойства* среды. Фазовая скорость волны равна сумме постоянной составляющей $c_0/\sqrt{\varepsilon_r}$ и быстроизменяющейся в окрестности перехода составляющей (рис. 11.1).

Частотная зависимость коэффициента фазы имеет вблизи частоты перехода форму, характерную для *аномальной дисперсии*, поскольку в этой области при возрастании частоты медленный рост β сменяется быстрым спадом.

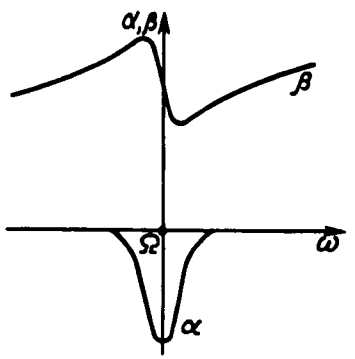


Рис. 11.1. Коэффициенты фазы и затухания в окрестности электрического дипольного перехода с частотой Ω

Показатель поглощения (9.53) при $\omega \approx \Omega$, согласно (11.35)—(11.37),

$$K \approx \frac{\pi L \Omega}{c_0 \varepsilon_0 \sqrt{\varepsilon_r} \hbar} \frac{|\mu_{12}|^2}{3} (N_1 - N_2) g(\omega, \Omega), \tag{11.38}$$

где

$$g(\omega, \Omega) = \frac{1}{\pi} \frac{\Delta\omega}{(\Omega - \omega)^2 + \Delta\omega^2} \tag{11.39}$$

— *функция Лоренца*; $\Delta\omega = T_2^{-1}$ — полуширина резонансной кривой на поло-

винном уровне. Из выражения (11.38) следует, что измерение поглощения дает возможность по экспериментальным данным вычислить матричный элемент $|\mu_{12}|$.

Поглощение пропорционально разности населенностей. Показатель поглощения (11.38) имеет максимальное значение, когда все атомы находятся в нижнем состоянии ($N_2 = 0$). При этом происходит поглощение энергии электромагнитного поля только атомами, находящимися на нижнем уровне, и переход их на верхний уровень. Если населенность $N_2 \neq 0$, то коэффициент поглощения уменьшается.

В условиях термодинамического равновесия населенности энергетических уровней атомов и молекул определяются законом Больцмана:

$$N_i^e = \frac{\sum_i N_i^e}{\sum_i e^{-\mathscr{E}_i/(k_Б T)}} e^{-\mathscr{E}_i/(k_Б T)}. \qquad (11.40)$$

В частности, для двухуровневой квантовой системы при $N = N_1 + N_2$

$$N_1^e = N \frac{e^{-\mathscr{E}_1/(k_Б T)}}{e^{-\mathscr{E}_1/(k_Б T)} + e^{-\mathscr{E}_2/(k_Б T)}};$$

$$N_2^e = N \frac{e^{-\mathscr{E}_2/(k_Б T)}}{e^{-\mathscr{E}_1/(k_Б T)} + e^{-\mathscr{E}_2/(k_Б T)}};$$

$$\frac{N_1^e}{N_2^e} = e^{\frac{\mathscr{E}_2 - \mathscr{E}_1}{k_Б T}} = e^{\frac{\hbar\Omega}{k_Б T}}. \qquad (11.41)$$

Следовательно, в равновесных условиях населенность верхнего энергетического уровня N_2^e всегда меньше населенности нижнего уровня N_1^e (при $\mathscr{E}_2 > \mathscr{E}_1$) и двухуровневая система ведет себя как поглощающая.

Из выражения (11.38) следует, что член, содержащий N_2, характеризует отрицательное поглощение или излучение, как бы противодействуя процессу поглощения атомами, находящимися на нижнем уровне. Противодействие, пропорциональное N_2 — это вынужденное излучение, когда происходит передача энергии от среды к полю с помощью атомов, сначала находившихся на верхнем уровне. Когда в невырожденной системе N_2 превосходит N_1, коэффициент K становится отрицательным и амплитуда электромагнитной волны нарастает в процессе ее распространения через среду. В этом и заключается *эффект квантового усиления*.

При наличии вырождения для возникновения отрицательного поглощения необходимо, чтобы число частиц, приходящихся на одно невырожденное состояние верхнего уровня превышало населенность каждого невырожденного состояния нижнего уровня $N_2/g_2 > N_1/g_1$ (g_1, g_2 —

кратности вырождения). Совокупность квантовых частиц с подобными населенностями уровней, т. е. среда с усилением распространяющегося в ней излучения, называется в квантовой электронике *активной средой.*

До сих пор предполагался низкий уровень электромагнитной волны, когда поглощаемая энергия быстро рассеивается в результате релаксационных процессов в окружающей среде; характеристики дипольной системы при этом не изменяются. Если же интенсивность излучения достаточно велика, поглощение энергии приводит к существенному изменению населенностей уровней. Их новое стационарное значение для бегущих волн вида (1.22) может быть получено из уравнений (11.31) и (11.34):

$$\frac{\hbar\Omega}{2}\frac{N_1 - N_2 - (N_1^e - N_2^e)}{T_1} = \mathrm{Re}\left(j\frac{\omega}{2}\dot{\mathbf{P}}_0\dot{\mathbf{E}}_0^*\right) \approx \frac{1}{2}\Omega\chi'' E_0^2,$$

откуда с учетом (11.35) стационарное среднее по времени значение разностей населенностей

$$N_1 - N_2 = \frac{N_1^e - N_2^e}{1 + \dfrac{I}{I_н}\dfrac{g(\omega,\Omega)}{T_2/\pi}}, \qquad (11.42)$$

где $I = \sqrt{\varepsilon_r}\varepsilon_0 c_0 E_0^2/2$ — интенсивность света или мощность, переносимая волной через единицу площади;

$$I_н = \frac{3c_0\varepsilon_0\sqrt{\varepsilon_r}}{(2T_1T_2/\hbar^2)L|\boldsymbol{\mu}_{12}|^2} \qquad (11.43)$$

— параметр насыщения, характеризующий интенсивность волны резонансной частоты, при которой разность населенностей уровней уменьшается вдвое по сравнению со случаем отсутствия излучения.

Из уравнения (11.42) следует, что при высоких интенсивностях излучения разность населенностей стремится к нулю, т. е. населенности верхнего и нижнего уровней становятся практически равными: $N_1 = N_2$. В этом заключается *эффект насыщения* дипольного перехода. Наибольшая эффективность насыщения перехода лучевым потоком достигается при резонансе ($\omega = \Omega$), когда в (11.42) $\pi g(\omega,\Omega)/T_2 = 1$. Вместе с тем из (11.42) следует, что при резонансном поглощении ни при какой интенсивности электромагнитного излучения невозможно добиться инверсии населенностей (по сравнению с равновесными значениями), т. е. получить больше атомов на верхнем уровне, чем на нижнем. Инверсия населенностей, необходимая для реализации эффекта квантового усиления, может быть достигнута при взаимодействии электромагнитного излучения с веществом при участии других энергетических уровней.

С учетом влияния электромагнитного поля на разность населенностей, согласно (11.38), (11.39) и (11.42), показатель поглощения

$$K(\omega)=\frac{\Omega L|\mu_{12}|^2}{3c_0\sqrt{\varepsilon_r}\varepsilon_0\hbar T_2}\frac{N_1^e-N_2^e}{(\Omega-\omega)^2+(1+I/I_н)T_2^{-2}}.$$

Таким образом, эффект насыщения проявляется также в уменьшении максимума показателя поглощения

$$K(\Omega)=\frac{[K(\Omega)]_{I=0}}{1+I/I_н} \qquad (11.44)$$

и увеличении ширины линии поглощения

$$\Delta\omega=(\Delta\omega)_{I=0}\sqrt{1+I/I_н};$$

т. е. проявляется в ее сглаживании и уширении (рис. 11.2).

По мере увеличения интенсивности электромагнитного излучения коэффициент поглощения уменьшается, в результате чего поглощаемая мощность, согласно (11.43) и (11.44), асимптотически приближается к предельному значению

$$\mathcal{P}_н(\omega)=[K(\Omega)]_{I=0}I_н=\frac{\hbar\Omega}{2}\frac{N_1^e-N_2^e}{T_1}.$$

Поскольку при полном насыщении дипольная система не может больше поглощать, любая дополнительная мощность, добавленная к мощности волны, передается через среду без ослабления. Такой переход называют *просветленным*.

С. И. Вавилов был первым, кто высказал и обосновал мысль о возможном уменьшении поглощения света при увеличении интенсивности облучения. Нелинейный характер процесса поглощения света большой интенсивности позволил ему ввести термин «нелинейная оптика» задолго до появления квантовой электроники.

Зависимость мощности поглощения или излучения квантовой системы от частоты называют *спектральной линией*. Спектральную линию излучения называют *линией люминесценции*. Ширина спектральной линии непосредственно связана с шириной энергетических уровней рабочего квантового перехода,

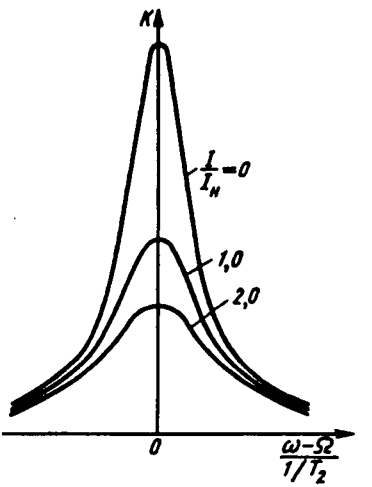


Рис. 11.2. Коэффициент поглощения при различных насыщениях

которая зависит от времени пребывания микрочастицы в данном состоянии. Вероятность спонтанных переходов изолированной неподвижной микрочастицы определяет наименьшую возможную ширину спектральной линии, которую называют *естественной шириной*.

Ширина спектральной линии вещества обычно значительно превышает естественную вследствие движения микрочастиц, их взаимодействия друг с другом и окружающей средой, в том числе со стенками сосуда и с электромагнитным излучением. Если при этом происходит только уменьшение времени жизни микрочастиц на энергетическом уровне, то такое уширение спектральной линии называют *однородным* и форма спектральной линии совокупности микрочастиц описывается функцией Лоренца. Если перечисленные факторы приводят к разному сдвигу центральных частот Ω_i квантовых переходов различных микрочастиц, то наблюдаемая экспериментально форма линии люминесценции всей совокупности возбужденных частиц представляет собой огибающую наложенных друг на друга однородно уширенных линий. Эта огибающая есть результат *неоднородного* уширения спектральной линии. Такое уширение возникает, например, из-за нерегулярного окружения активных ионов в матрице силикатного стекла и соответствующего изменения положения их энергетических уровней. Аналогичная ситуация возникает и в смеси разных газов (гелий—неон) вследствие эффекта Доплера в соответствии с хаотическим тепловым движением активных частиц. При неоднородном уширении форма спектральной линии описывается, как правило, функцией Гаусса. Ее ширина может значительно превышать однородное уширение элементарной линии люминесценции.

Уширение уровней энергии есть следствие сокращения их времени жизни и соотношения неопределенностей ($\Delta \mathcal{E} \Delta t \geqslant \hbar$): определение энергии может производиться за время, не превышающее время жизни в этом состоянии τ. Поэтому неточность нахождения энергии не может быть меньше $\hbar/\tau$. Постоянная τ является одновременно мерой времени, необходимого для того, чтобы возбужденная квантовая система отдала свою энергию.

§ 11.6. Квантовый усилитель

Среда может усиливать проходящее через нее электромагнитное излучение, если коэффициент поглощения (11.38) становится отрицательным и населенность верхнего уровня превышает населенность нижнего. Такое соотношение населенностей противоположно их соотношению (11.41) при тепловом равновесии и дает основание считать, что переход имеет *инверсию населенностей*, а среда возбуждена или обладает «отри-

цательной температурой». Инверсия населенностей обеспечивается источником возбуждения или *накачки*.

Инверсия населенностей энергетических уровней. Для получения инверсии населенностей уровней существует много различных способов. В первом квантовом приборе был использован метод сортировки пучка молекул в статическом неоднородном поле квадрупольного конденсатора, которое отклоняло от оси пучка молекулы аммиака, находящиеся на нижнем уровне. Поэтому на выходе пучка из конденсатора в его осевой части преобладали молекулы, находящиеся на верхнем уровне.

Возбуждение активной среды может также осуществляться пучком электронов (с энергией в десятки килоэлектрон-вольт), инжекцией неравновесных носителей заряда в плоскости *p-n*-перехода, в газовом разряде, химическими и тепловыми методами. В последнем случае используется различие времен релаксации нескольких энергетических уровней при быстром изменении температуры, когда между некоторыми из переходов образуется инверсия населенностей.

Методы вспомогательного электромагнитного излучения (накачки) получили наиболее широкое распространение для возбуждения парамагнитных и оптических квантовых усилителей и генераторов на примесных кристаллах и стеклах. В зависимости от числа используемых энергетических уровней различают трехуровневый, четырехуровневый и т. д. методы.

Трехуровневый метод основан на эффекте насыщения перехода из основного на самый верхний энергетический уровень для получения инверсии населенностей между другой парой трехуровневой системы. Если промежуточный (второй) уровень расположен ближе к верхнему (третьему) уровню, то инверсия населенностей возникает между третьим и вторым уровнями (рис. 11.3, *а*) при условии, что происходит насыщение перехода *1—3* и сохранение равновесной населенности второго уровня, показанной пунктирной линией. Если промежуточный уровень расположен ближе к нижнему уровню (рис. 11.3, *б*), то инверсия населенностей возникает между ним и основным (первым) уровнем. Существенно повышает

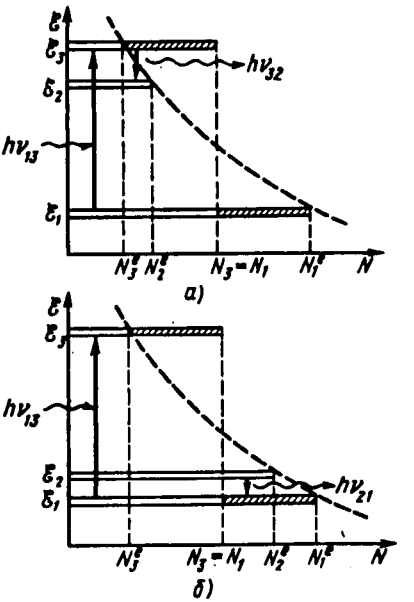


Рис. 11.3. Схема накачки и излучения в трехуровневой квантовой системе

эффективность накачки метастабильный характер промежуточного уровня. Таким свойством обладает, в частности, трехуровневая система ионов хрома Cr^{3+} в решетке Al_2O_3, свойственной рубину, в которой среднее время жизни для спонтанных переходов $\tau_{32} \approx 5 \cdot 10^{-8}$ с, $\tau_{21} \approx 3 \cdot 10^{-3}$ с и $\tau_{31} \approx 10^{-5}$ с.

При *четырехуровневом методе* накачки создается инверсия населенностей между двумя уровнями и в четырехуровневой квантовой системе с энергиями $\mathcal{E}_1 < \mathcal{E}_2 < \mathcal{E}_3 < \mathcal{E}_4$. Обычно инверсия возникает между промежуточными уровнями (вторым и третьим). Если второй уровень расположен достаточно далеко от основного уровня *1*, то его населенность существенно меньше населенности основного уровня. Поэтому отрицательная температура рабочего перехода (*3—2*) может быть достигнута при меньших населенностях уровня *4* и при меньших мощностях накачки. Значительный выигрыш в мощности накачки при переходе от трехуровневой к четырехуровневой системе может быть получен при $\hbar\Omega_{41} > 3k_{\text{Б}}T$. Выигрыш особенно значителен, когда разность энергий между уровнями *1—3* и *2—4* одинакова. Тогда излучение накачки с частотой $\omega = \Omega_{13} = \Omega_{24}$ кроме прямого обогащения верхнего уровня (*3*) рабочего перехода осуществляет одновременно прямое обеднение его нижнего уровня (*2*). Для возбуждения четырехуровневых систем требуется обычно мощность в десятки и сотни раз меньшая, чем для возбуждения трехуровневых систем.

Усиление бегущей волны. Изучение процессов усиления предполагает рассмотрение нестационарного решения уравнений (11.32), (11.30) и (11.31), соответствующего переходному процессу во времени и пространстве. Упрощает проблему приближение, при котором находится неполное нестационарное решение системы. Оно возможно благодаря тому, что во многих случаях приходится иметь дело с процессами и явлениями с двумя совершенно различными временны́ми характеристиками. Например, в оптическом диапазоне период электромагнитных колебаний составляет около 10^{-15} с, а время релаксации, характеризующее обмен энергии между полем и дипольной системой, обычно составляет около 10^{-10} с или более. Это позволяет рассматривать процессы обмена энергии между средой и полем, отделяя их от основных высокочастотных колебаний методом усреднения. Усреднение в пространстве облегчено также учетом различных пространственных масштабов рассматриваемых явлений. Одно из них соответствует длине волны электромагнитного излучения, а второе — заметному нарастанию или затуханию полей на расстояниях, гораздо больших длины волны. Подобное усреднение позволяет осуществить квазистационарное приближение (метод изменяющихся параметров), когда допускаются мед-

ленные изменения «стационарных значений» параметров в пространстве и времени.

В качестве примера рассмотрим усилитель бегущей волны, в котором волна (1.22) оптического излучения резонансной частоты распространяется вдоль оси z в среде с инверсией населенностей $N_{21}^{\text{н}} = (N_2 - N_1)^{\text{н}} > 0$. Исходя из изложенных соображений, будем предполагать, что параметры этих бегущих волн представляют собой стационарные колебания во времени с частотой $\omega = \Omega$, допускающие *медленные* изменения в пространстве комплексных амплитуд $\dot{\mathbf{P}}_0(z)$, $\dot{\mathbf{E}}_0(z)$ и разности $N_{21}(z) = N_2 - N_1$. Поэтому в уравнении (11.31) можно пренебречь малой компонентой $\partial N_{21}/\partial t$, а в уравнении (11.30) — компонентой $\partial^2 \dot{\mathbf{E}}_0/\partial t^2 \ll k\partial \dot{\mathbf{E}}_0/\partial t$. Тогда исходные уравнения (11.30)—(11.32) превращаются в систему уравнений для комплексных амплитуд

$$\dot{\mathbf{P}}_0(z) = j\frac{T_2}{\hbar} L \frac{|\boldsymbol{\mu}_{12}|^2}{3} N_{21}\dot{\mathbf{E}}_0,$$

$$\frac{N_{21} - N_{21}^{\text{н}}}{T_1} = \frac{j}{\hbar}\dot{\mathbf{P}}_0\dot{\mathbf{E}}_0,$$

$$j2\underline{k}\frac{\partial \dot{\mathbf{E}}_0}{\partial z} + j\underline{k}(c_0\mu_0\sigma/\sqrt{\varepsilon_r})\dot{\mathbf{E}}_0 = \Omega^2\mu_0\dot{\mathbf{P}}_0,$$

которая позволяет получить амплитуды и фазы медленно изменяющихся компонент поляризации и напряженности поля, а также разность населенностей как функции расстояния, пройденного волной в среде:

$$N_{21}(z) = \frac{N_{21}^{\text{н}}}{1 + T_1 T_2 L|\boldsymbol{\mu}_{12}|^2 (3\hbar)^{-1} E_0^2(z)}. \qquad (11.45)$$

Исключив из уравнения для поля поляризацию и умножив это уравнение на $\dot{\mathbf{E}}$, получим *кинетическое уравнение*

$$\frac{\partial}{\partial z}E_0^2(z) + \frac{c_0\mu_0\sigma}{\sqrt{\varepsilon_r}}E_0^2(z) = \frac{\Omega T_2 L|\boldsymbol{\mu}_{12}|^2}{3c_0\varepsilon_0\sqrt{\varepsilon_r}\hbar} N_{21}(z)E_0^2(z), \qquad (11.46)$$

поскольку оно характеризует скорость передачи энергии из среды в поле по мере распространения электромагнитной волны.

Нелинейное дифференциальное уравнение для средней мощности $I = c_0\varepsilon_0\sqrt{\varepsilon_r}\,E_0^2/2$, переносимой волной через единицу площади, можно получить из совместного решения (11.45) и (11.46):

$$\frac{\mathrm{d}I(z)}{\mathrm{d}z}=\left[K_0\frac{I(0)+I_н}{I(z)+I_н}-K_σ\right]I(z)\equiv K(I)I(z), \qquad (11.47)$$

где $K_0=\hbar\Omega N_{21}^{н}/2T_1[I(0)+I_н]$ — коэффициент усиления в плоскости $z=0$; $I(0)$ — входная мощность в плоскости $z=0$; $K_σ=c_0\mu_0\sigma/\sqrt{\varepsilon_r}$ — коэффициент нерезонансных потерь; $I_н$ — параметр насыщения (12.64);

$$K(I)=K_0\frac{I(0)+I_н}{I(z)+I_н}-K_σ \qquad (11.48)$$

— коэффициент усиления в плоскости $z>0$.

Уравнение (11.47) определяет закон изменения интенсивности бегущей волны в нелинейной активной среде. Здесь нелинейность вносится зависимостью инверсии населенностей от интенсивности волны. В зависимости от начальной интенсивности и соотношения между коэффициентом нерезонансных потерь и начальным коэффициентом усиления характер изменения амплитуды волны по мере ее распространения меняется.

При низком уровне электромагнитного сигнала, когда $I(0)\ll I_н$ и $K_0>K_σ$, волна заметно не меняет инверсию населенности, коэффициент усиления (11.48) остается практически постоянным ($K=K_0-K_σ>0$) и интенсивность волны начинает экспоненциально возрастать в соответствии с обращенным интегральным законом Бугера

$$I(z)=I(0)\mathrm{e}^{(K_0-K_σ)z}, \qquad (11.49)$$

пока она остается достаточно малой (нелинейность среды не проявляется).

При малом начальном усилении, когда $K_0<K_σI_н[I(0)+I_н]$ потери превышают эффект квантового усиления, общий коэффициент усиления всюду (при всех z) остается отрицательным и интенсивность волны непрерывно убывает ($\mathrm{d}I/\mathrm{d}z<0$).

При среднем усилении, когда $K_σ>K_0>K_σI_н[I(0)+I_н]$ коэффициент усиления (11.48) сначала отрицателен, а затем за счет поглощения увеличивается до некоторого положительного значения, при котором поглощение и потери уравновешиваются [$K(I_{ст})=0$]. В этом случае волна распространяется без затухания и поглощения с постоянной стационарной интенсивностью $I_{ст}=K_0I(0)/K_σ+(K_0-K_σ)I_н/K_σ$, меньшей начальной интенсивности $I(0)$.

При большом усилении, когда $K_0>K_σ$, коэффициент усиления (11.48) сначала имеет максимальное значение $K_0-K_σ>0$, обусловливающее рост амплитуды волны. Однако по мере распространения волны, увели-

чения ее интенсивности и снижения инверсии населенностей коэффициент усиления (11.48) непрерывно уменьшается, асимптотически приближаясь к нулю (потери полностью компенсируют квантовое усиление). В результате прекращается возрастание интенсивности с расстоянием и она достигает предельного значения

$$I(\infty) = \frac{K_0}{K_\sigma} I(0) + \frac{K_0 - K_\sigma}{K_\sigma} I_н > I(0), \qquad (11.50)$$

т. е. возникает *эффект насыщения усиления*.

Насыщение усиления является той нелинейностью, которая ограничивает выходную мощность усилителей и интенсивность излучения квантовых генераторов.

Рассмотрение частотной зависимости коэффициента усиления показывает, что ширина полосы пропускания сужается по мере роста усиления.

Усиление в резонаторе. Усиление электромагнитного поля резонатора с активной средой ($N_{21} > 0$) описывается уравнениями, соответствующими нестационарным стоячим волнам. Комплексные амплитуды поляризации и напряженности электрического поля резонансной частоты $\omega_c = \Omega$ могут быть представлены в виде

$$\dot{\mathbf{P}}(t) = \dot{\mathbf{P}}_0(t) e^{j\Omega t}; \quad \dot{\mathbf{E}}(t) = \dot{\mathbf{E}}_0(t) e^{j\Omega t},$$

где $\dot{\mathbf{P}}_0(t)$, $\dot{\mathbf{E}}_0(t)$ — медленно меняющиеся комплексные амплитуды поляризации и напряженности.

Уравнение для медленно меняющихся компонент разности населенностей можно получить, пренебрегая в уравнении (11.31) высокочастотными членами и приравнивая члены нулевой частоты:

$$\frac{\partial N_{21}}{\partial t} + \frac{N_{21} - N_{21}^н}{T_1} = \frac{j}{\hbar} \dot{\mathbf{P}}_0 \dot{\mathbf{E}}_0. \qquad (11.51)$$

Уравнение для медленно меняющейся компоненты поляризации можно найти, пренебрегая в (11.30) малым членом $\partial^2 \dot{P}_0 / \partial t^2 << 2\Omega \partial \dot{P}_0 / \partial t$ и учитывая малую ширину линии молекулярного резонанса $2/T_2 << \Omega$:

$$\frac{\partial \dot{\mathbf{P}}_0}{\partial t} + \frac{\dot{\mathbf{P}}_0}{T_2} = jL \frac{|\boldsymbol{\mu}_{12}|^2}{3\hbar} N_{21} \dot{\mathbf{E}}_0. \qquad (11.52)$$

Уравнение для медленно меняющейся компоненты напряженности электрического поля можно получить, пренебрегая в (11.33) малыми слагаемыми $\partial^2 \mathbf{E}_0 / \partial t^2 << 2\Omega \partial \mathbf{E}_0 / \partial t$, $d\mathbf{E}_0 / \partial t \ll \Omega \mathbf{E}_0$, $\partial \mathbf{P}_0 / \partial t \ll \Omega \mathbf{P}_0$ и учитывая малую ширину резонансной кривой добротного резонатора $1/\tau_c << \Omega$:

$$\frac{\partial \dot{\mathbf{E}}_0}{\partial t} + \frac{\dot{\mathbf{E}}_0}{2\tau_c} = -j\frac{\Omega}{2\varepsilon}\dot{\mathbf{P}}_0. \tag{11.53}$$

Дальнейшее упрощение уравнений возможно, если отказаться от фазовой информации о поле и учесть, что для твердого тела ширина молекулярной или атомной линии $2/T_2$ обычно значительно больше ширины резонансной кривой резонатора $1/\tau_c$. Тогда можно пренебречь производной $\partial \mathbf{P}_0/\partial t \ll P_0/T_2$ и по (11.51)—(11.53)

$$\frac{\partial N_{21}}{\partial t} + \frac{N_{21} - N_{21}^{\text{н}}}{T_1} = -\frac{T_2 L|\mu_{12}|^2}{3\hbar} N_{21} E_0^2 =$$

$$= -\frac{2\Omega T_2 L|\mu_{12}|^2}{3\varepsilon\hbar} N_{21} n_\nu = -2WN_{21}; \tag{11.54}$$

$$\frac{\partial E_0^2}{\partial t^2} + \frac{E_0^2}{\tau_c} = \frac{\Omega T_2 L|\mu_{12}|^2}{3\varepsilon\hbar} N_{21} E_0^2;$$

$$\frac{\partial n_\nu}{\partial t} + \frac{n_\nu}{\tau_c} = \frac{\Omega T_2 L|\mu_{12}|^2}{3\varepsilon\hbar} N_{21} n_\nu, \tag{11.55}$$

где

$$W = \frac{T_2 L}{2\hbar^2}\frac{|\mu_{12}|^2}{3} E_0^2 = \frac{1}{3}\Omega T_2 n_\nu; \tag{11.56}$$

$n_\nu = \varepsilon E_0^2/(2\hbar\Omega)$ — величина, пропорциональная плотности электромагнитной энергии. Ее лишь условно можно называть *плотностью фотонов*, поскольку это классическая величина, соответствующая пространственному распределению поля стоячей волны, тогда как законы квантовой механики (согласно принципу неопределенности) не допускают локализации отдельных фотонов внутри области пространства, меньшей длины волны. Поэтому не вполне корректно полагать, что фотоны, например, собираются в пучностях.

Уравнения (11.54) и (11.55) определяют скорость обмена энергией между полем и средой, и поэтому их называют *кинетическими* или *скоростными*.

Вероятность энергетических переходов. Кинетическое уравнение (11.54) для разности населенностей $N_{21} = N_2 - N_1$ можно переписать в виде

$$\frac{\partial N_2}{\partial t} + \frac{N_2 - N_2^{\text{н}}}{T_1} = -WN_2 + WN_1;$$

$$\frac{\partial N_1}{\partial t} + \frac{N_1 - N_1^{\text{н}}}{T_1} = WN_2 - WN_1, \tag{11.57}$$

если учесть условие $N = N_1 + N_2$.

Первый член в левой части уравнений характеризует полную скорость изменения населенности уровня, второй член — скорость изменения населенности уровня при безызлучательном взаимодействии с окружающей средой (без непосредственной связи с электромагнитным полем). Члены в правой части определяют скорость изменения населенности уровня, обусловленную взаимодействием с полем, поскольку вероятность излучения (11.56) пропорциональна плотности электромагнитной энергии.

Таким образом, величина W представляет собой скорость или *вероятность*, с которой один атом или молекула может переходить с одного уровня на другой под действием излучения в единицу времени. Переход на верхний уровень сопровождается поглощением фотона, а переход на нижний уровень — его излучением. Переходы на нижний уровень под воздействием электромагнитного поля определяют процесс индуцированного, стимулированного или *вынужденного излучения*.

Первой особенностью вынужденного излучения является то, что его вероятность (11.56) *пропорциональна* плотности энергии индуцирующего электромагнитного поля. Из уравнений (11.33) или (11.52) следует, что дипольная поляризация (дипольные переходы) находятся также в определенном фазовом соотношении с индуцирующим электромагнитным полем. Поэтому второй особенностью индуцированных переходов и вынужденного излучения является их *когерентность* (синфазность излучения всех одинаковых элементарных осцилляторов). Вероятность вынужденного излучения (W_{21}) согласно уравнениям (11.57), одинакова с вероятностью поглощения: $W_{12} = W = W_{21}$.

Строгий анализ взаимодействия излучения с веществом, учитывающий не только квантование среды, но и квантование поля, приводит после усреднения к уравнениям, аналогичным (11.54) и (11.56). Основное различие заключается в том, что скорость или вероятность процессов излучения W_{21} пропорциональна не n_ν, а $n_\nu + 1$. Второе слагаемое в этой сумме определяет составляющую вероятности излучения, которая не зависит от числа фотонов. Этой составляющей соответствует процесс излучения, происходящий в отсутствие внешнего поля и называемый поэтому *спонтанным излучением*. Спонтанные переходы возникают вследствие взаимодействия электронов в атоме с так называемым фотонным вакуумом или «нулевыми флуктуациями» электромагнитного поля. Поэтому спонтанный переход происходит самопроизвольно, случайно во времени, аналогично радиоактивному распаду. Между спонтанным излучением и внешним электромагнитным полем нет никаких связей, в том числе и фазовых. Поэтому спонтанное излучение некогерентно и в квантовых приборах играет роль шума.

Рассмотренный ранее полуклассический подход, при котором не учитывалось квантование поля, является хорошим приближением при большом числе фотонов ($n_\nu \gg 1$). Он справедлив для квантовых усилителей и генераторов до тех пор, пока нас не интересует порог возбуждения или их шумовые характеристики. Для точного расчета в уравнениях, полученных полуклассическим методом, необходимо для процессов излучения n_ν заменить на $n_\nu + 1$.

Соотношение между вероятностями спонтанного и вынужденного излучений нетрудно также получить из условий термодинамического равновесия. Для этого вероятность индуцированного излучения в единицу времени необходимо представить как произведение постоянного коэффициента B_{ij} на спектральную плотность излучения $u(\Omega_{ij}, T)$ вблизи частоты перехода Ω_{ij}, т. е. $W_{ij} = B_{ij} u(\Omega_{ij}, T)$. Аналогично вероятность индуцированного поглощения $W_{ji} = B_{ji} u(\Omega_{ij}, T)$. Вероятность спонтанного излучения A_{ij} не зависит от плотности внешнего излучения. Поэтому полная вероятность излучения в единицу времени $P_{ij} = A_{ij} + B_{ij} u(\Omega_{ij}, T)$, а вероятность поглощения $P_{ji} = W_{ji} = B_{ji} u(\Omega_{ij}, T)$.

В условиях термодинамического равновесия число переходов сверху вниз в единицу времени равно числу обратных переходов $P_{ij} N_i^e = P_{ji} N_j^e$ или с учетом (11.40)

$$\frac{P_{ji}}{P_{ij}} = \frac{B_{ji} u(\Omega_{ij}, T)}{A_{ij} + B_{ij} u(\Omega_{ij}, T)} = \frac{N_i^e}{N_j^e} = \mathrm{e}^{\frac{\mathcal{E}_j - \mathcal{E}_i}{k_\text{Б} T}}.$$

Если приравнять получающееся отсюда значение спектральной плотности излучения к плотности излучения абсолютно черного тела (формула Планка)

$$u(\Omega_{ij}, T) = \frac{A_{ij}}{B_{ji}\,\mathrm{e}^{\hbar\Omega_{ij}/(k_\text{Б} T)} - B_{ij}} = \frac{8\pi h \nu_{ij}}{c_0^3} \frac{1}{\mathrm{e}^{\hbar\Omega_{ij}/(k_\text{Б} T)} - 1},$$

то получим

$$B_{ij} = B_{ji}; \tag{11.58}$$

$$A_{ij} = \frac{8\pi\hbar\nu_{ij}^3}{c_0^3} B_{ij} = \frac{1}{\tau_{ij}}. \tag{11.59}$$

Таким образом, и при наличии спонтанных переходов вероятность индуцированных переходов вверх (поглощение) и вниз (излучение) в одной квантовой системе остается одинаковой для всех пар состояний.

Вероятность (11.59) пропорциональна кубу частоты перехода. Поэтому спонтанное излучение может быть доминирующим релаксационным процессом в оптических явлениях типа газового разряда. Однако при более низких частотах, например в диапазоне СВЧ, спонтанным излучением можно пренебречь (если не рассматривать усиление и шумы на пороге чувствительности).

Из приведенных рассуждений также следует, что пренебрежение спонтанным излучением входит, строго говоря, в противоречие с законами термодинамического равновесия.

Индуцированное излучение было постулировано, а его когерентность интуитивно угадана А. Эйнштейном в 1916 г. в процессе вывода формулы Планка, исходя из постулатов Бора. Обоснование существования индуцированного излучения дано П. А. Дираком в 1927 г. при разработке последовательной теории излучения и поглощения света. Однако, только в 1940 г. В. А. Фабрикантом была выявлена способность системы возбужденных атомов усиливать световое излучение.

Квантовый парамагнитный усилитель. Эффект квантового усиления возникает при взаимодействии электромагнитного излучения с активным веществом. Он описывается соответствующим матричным элементом гамильтониана взаимодействия. Введение в (11.15) и (11.16) разложения векторного потенциала (11.17) приводит к появлению в гамильтониане членов взаимодействия моментов различного порядка, вплоть до сколь угодно высокого. Необходимость учета моментов различного порядка возникает редко, так как с увеличением порядка взаимодействия его значение уменьшается. К средам с подобным гамильтонианом относятся, в частности, те, в которых матричный элемент оператора электрического момента (11.24) обращается в нуль (например, для переходов между состоянием противоположной четности) и необходимо учитывать магнитное дипольное взаимодействие. Для его рассмотрения в функцию Гамильтона (11.15) вводят второй член разложения (11.17) и учитывают компоненту взаимодействия $\mathscr{H}' = -\mathbf{m}\cdot\mathbf{B}(\mathbf{R},t)$, где $\mathbf{m} = -e\mathbf{r}\times\mathbf{p}/(2m)$ — магнитный момент; $\mathbf{p} = m\dfrac{\partial\mathbf{r}}{\partial t} + \dfrac{e}{2}\mathbf{r}\times[\nabla_R\times\mathbf{A}(\mathbf{R},t)]$ — канонический импульс.

Магнитное дипольное взаимодействие может быть рассмотрено тем же методом, что и электрическое дипольное взаимодействие. Оно приводит для правополяризованной волны с частотами, близкими к резонансной ($\omega \approx \Omega$) в стационарном режиме, к коэффициенту усиления

$$K_+ = \frac{\mu_0\sqrt{\varepsilon_r}}{2c_0}\pi\gamma^2\hbar\Omega(N_2 - N_1)g(\omega,\Omega) >> K_- \approx 0.$$

Магнитное дипольное взаимодействие лежит в основе действия квантовых парамагнитных усилителей СВЧ. Их иногда называют *мазерами* (maser — Microwave Amplification by Stimulated Emission of Radiation).

Квантовые парамагнитные усилители (КПУ) сочетают возможность перестройки частоты Ω в широком диапазоне (недоступную газовым мазерам и на молекулярном пучке) с чрезвычайно низким уровнем шумов (недоступным усилителям с электронными потоками). Это существенно, когда принимаемый сигнал настолько слаб, что никакими другими устройствами его не удается выделить на фоне собственных шумов приемной аппаратуры. Это относится, в частности, к радиоастрономии, космической связи и локации.

В мазерах используются электронные спиновые уровни парамагнитных ионов кристаллов во внешнем магнитном поле. Переходы между ними определяют процессы и усиления (при индуцированном излучении) и электронного парамагнитного резонанса (при поглощении).

Ион парамагнитного материала обладает результирующим магнитным моментом, и во многих случаях его можно рассматривать как жесткий магнитный диполь. Во внешнем постоянном поле магнитный момент парамагнитной частицы не может принимать любые значения как по модулю, так и по знаку (ориентации); он является квантовой величиной. Энергия отдельного парамагнитного иона в постоянном магнитном поле H_0 может иметь только дискретные квантовые значения, кратные магнетону Бора μ_B и зависящие от квантовых чисел L (полный орбитальный угловой момент иона) и S (полный спиновый момент электронов): $\mathscr{E} = \mu_B \mathbf{H}_0(\mathbf{L} + 2\mathbf{S})$.

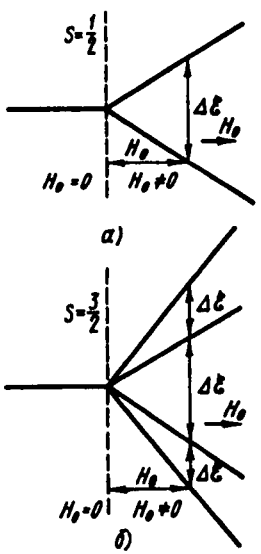


Рис. 11.4. Уровни энергии парамагнитной частицы во внешнем постоянном магнитном поле

Приведенное выражение показывает, что возможны $l + 2s$ ориентаций магнитного момента парамагнитной частицы. Для свободного электрона $s = 1/2$ и $l = 1$, при этом число возможных проекций магнитного момента на направление магнитного поля $2s + 1 = 2$, т. е. магнитный момент электрона во внешнем магнитном поле может иметь одно из двух значений: $+\hbar/2$ или $-\hbar/2$ и может быть направлен либо вдоль, либо против внешнего поля. Уровни энергии такого электрона изображены на рис. 11.4, *а*. Разность энергии между этими двумя энергетическими уровнями $\Delta\mathscr{E} = 2\mu_B H_0$.

Чаще всего используются вещества, атомы которых имеют спиновое число больше 1/2. В этом случае получается большее число возможных энергетических уровней (рис. 11.4, *б*).

Таким образом, при приложении магнитного поля к парамагнитной частице вследствие наличия у нее магнитного момента уровень энергии частицы расщепляется на несколько подуровней (эффект Земана). Интервалы между ними зависят от внешнего постоянного магнитного поля. Диапазону СВЧ соответствуют напряженности поля в десятки и сотни тысяч ампер на метр $[H = \hbar\omega/(2\mu_в)]$.

В парамагнитных усилителях применяются парамагнитные кристаллы, а не изолированные парамагнитные частицы. В кристаллической решетке диамагнитного вещества парамагнитные ионы подвергаются влиянию электрического поля решетки. Это приводит к существенному изменению картины возможных энергетических уровней, изображенных на рис. 11.4, *а*, *б*. Внутрикристаллическое поле резко ослабляет орбитальный магнитный момент и сводит магнитный момент иона в основном к спиновому моменту. Это объясняется тем, что в кристаллической решетке электроны в атомах движутся по своим орбитам в различных направлениях и результирующий орбитальный момент очень мал. Из-за связи остаточного орбитального момента с внутрикристаллическим полем решетки расположение возможных энергетических уровней зависит не только от напряженности постоянного внешнего поля, но и от угла θ между направлением этого поля и оптической осью кристалла. Кроме того, возникает начальное (при отсутствии внешнего магнитного поля) расщепление энергетических уровней, обусловленное внутренним электрическим полем кристалла (эффект Штарка). Диаграмма уровней энергии парамагнитного кристалла в зависимости от магнитного поля при различных ориентациях поля относительно оптической оси кристалла приведена на рис. 11.5.

Если на парамагнитный кристалл воздействовать переменным электромагнитным полем с частотой, соответствующей разности каких-либо двух энергетических уровней, то будут происходить переходы спинов с одного из двух подуровней на другой. Переходы сопровождаются изменением ориентации спиновых моментов (спинов).

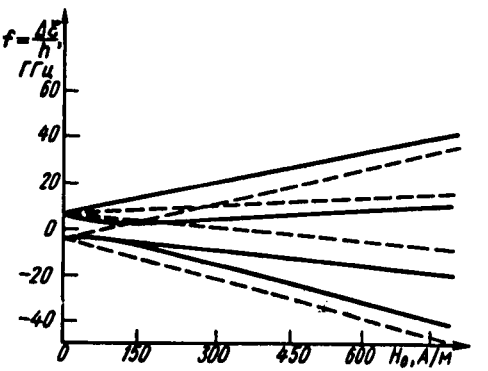


Рис. 11.5. Уровни энергии рубина при различных углах между направлением магнитного поля и оптической осью кристалла (при θ = 0 — пунктирная линия, при θ = 90° — сплошная)

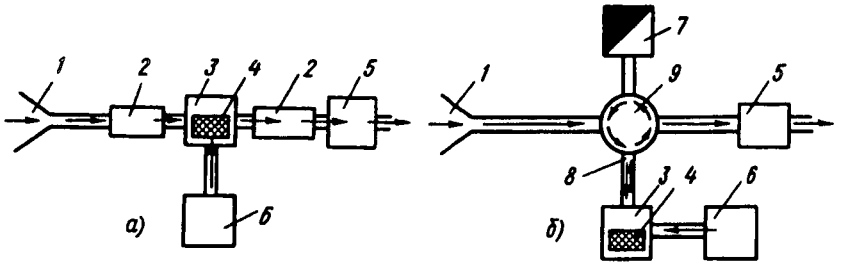


Рис. 11.6. Схемы квантовых усилителей с проходным (*а*) и отражательным (*б*) резонаторами

Применение в качестве усилителя отрезка волновода, заполненного веществом, непрактично, так как для получения нужных коэффициентов усиления он должен обладать чрезвычайно большой длиной. В существующих конструкциях квантовых усилителей используется то обстоятельство, что большие эффективные электрические длины можно получить в коротких (по геометрическим размерам) структурах. В *усилителе бегущей волны* волновод, заполненный активным веществом, укорачивается с помощью замедляющей структуры.

Другой путь состоит в помещении активного вещества в резонансный объем. Электромагнитная волна, попавшая в объемный резонатор с большой добротностью Q_0, прежде чем покинуть резонатор, претерпевает примерно Q_0 отражений от его внутренних стенок. Именно таким образом достигается в *резонаторном квантовом усилителе* высокая эффективная длина пути.

Электромагнитное поле в резонаторе, настроенном на частоту излучения, побуждает избыток возбужденных молекул, находящихся в резонаторе, к излучению. Усиление становится возможным в том случае, когда энергия, излучаемая возбужденными молекулами, достаточна для преодоления потерь в резонаторе.

В усилителе с проходным резонатором (рис. 11.6, *а*) поступающий от антенны *1* сигнал подводится к резонатору *3* с парамагнитным кристаллом *4* по одной линии, а отводится — по другой. Однонаправленное распространение сигнала (в направлении от антенны к приемнику *5*) обеспечивается включением в линию вентилей *2*.

Усилитель с проходным резонатором по сравнению с усилителем с отражательным резонатором (рис. 11.6, *б*) имеет полосу пропускания в два раза меньшую (при том же усилении). Объясняется это тем, что половина энергии, полученной кристаллом, возвращается во входную линию и теряется в вентиле или в выходном резисторе источника сигнала. Поэтому резонаторные КПУ строят, как правило, с отражательным

резонатором. В таком усилителе одна и та же линия используется для подачи в резонатор с кристаллом поступающего от антенны слабого сигнала, подлежащего усилению, и отведения усиленного сигнала. Разделение слабого и усиленного сигналов производится циркулятором. В схеме квантового усилителя циркулятор *9* (см. рис. 11.6, *б*) первым плечом подключен к антенне *1*, вторым — к резонатору *3*, третьим — к приемнику *5* (вернее, к последующим каскадам приемника) и четвертым — к согласованной нагрузке — поглотителю *7*. Сигнал от антенны через ферритовый циркулятор попадает по линии *8* в резонатор усилителя, куда поступает также энергия от генератора накачки *6*. Из резонатора усиленный сигнал по той же линии поступает в приемник. Шумы, возникающие во входном устройстве приемника, а также отраженная от его входа мощность (при плохом согласовании антенны и приемника) поступают только в поглотитель. Циркулятор исключает попадание и усиление шумов приемников в КПУ.

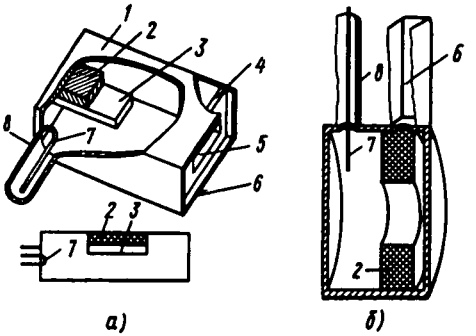


Рис. 11.7. Резонаторы КПУ

Резонатор КПУ должен иметь две резонансные частоты — сигнала и накачки. На рис. 11.7 показаны конструкции резонаторов КПУ: усилителя трехсантиметрового диапазона с генератором накачки односантиметрового диапазона (*а*) и усилителя двадцатисантиметрового диапазона с генератором накачки трехсантиметрового диапазона (*б*). Парамагнитный кристалл *2* в резонаторе первого усилителя (рис. 11.7, *а*) помещен вплотную к стенке волновода *1* в пучность магнитных силовых линий поля сигнала накачки. Резонатор на частоте сигнала образован четвертьволновой полосковой линией *3* и стенкой волновода *1*. На частоте накачки резонатором является отрезок волновода. Настройка резонатора производится с помощью поршня *4*. Сигнал подводится по коаксиальной линии *8*, заканчивающейся короткозамкнутым витком *7*. Накачка подводится по волноводу *6*, связь с генератором осуществляется щелью *5*.

Квантовый усилитель бегущей волны состоит из волновода и размещенного вдоль него активного вещества, по которому распространяются электромагнитные волны сигнала и накачки. Поле накачки возбуждает активное вещество, насыщая соответствующий переход, в результате чего образуется отрицательная температура между одной из пар энергетических уровней. Электромагнитная волна сигнала, взаимодействуя с парамагнетиком, вызывает индуцированное излучение и тем самым усиливается. Для получения заметного усиления длина обычного волновода

с парамагнитным веществом должна составлять несколько метров. Чтобы КПУ имел приемлемые для практики размеры, применяют структуры, обеспечивающие уменьшение групповой скорости сигнала. Чем медленнее волна, тем больше времени она взаимодействует с парамагнитным веществом и тем заметнее усиление.

Чтобы исключить рассогласование входной и выходной линий и получить однонаправленное усиление, активный материал располагают в структуре так, что он взаимодействует с полем сигнала только при одном направлении распространения волны. Кроме того, вводят специальный элемент, поглощающий обратную волну и позволяющий получить однонаправленные потери.

В структурах диэлектрического типа используется замедление в диэлектрике. Замедление зависит от диэлектрической проницаемости диэлектрика и его размещения в линии.

Периодические замедляющие структуры резонансного типа позволяют получать в некоторой полосе очень большие замедления волны (до 1000 раз). Одной из наиболее удобных является гребенчатая или штыревая замедляющая система, которая легко обеспечивает необходимую ориентацию постоянного и переменного магнитных полей, если постоянное магнитное поле приложено в направлении ряда штырей. Поле накачки, распространяясь в волноводе в виде основной волны типа H_{10}, образует сильное продольное магнитное поле у стенки волновода и сильное поперечное поле в его центре; при этом штыри не оказывают на эту волну большого влияния. По обе стороны штырей возникают области круговой поляризации с противоположным направлением вращения, причем плоскость вращения круговой поляризации перпендикулярна плоскости штырей. В случае гребенчатой структуры магнитное поле максимально у основания штырей. Используя такую благоприятную ориентацию высокочастотных полей, по одну сторону плоскости штырей помещают активное вещество — парамагнитный кристалл, предназначенный для усиления прямой волны, а по другую сторону — элемент с невзаимным затуханием — вентиль. В качестве вентильного элемента может быть использован тот же парамагнитный кристалл (например, рубин, но с большей концентрацией парамагнитных ионов) или специальный феррит. В первом случае поглощающий кристалл берут с такой концентрацией парамагнитных ионов, чтобы при выбранной мощности вспомогательного генератора не происходило насыщения по переходу накачки.

Во втором случае объем ферритового стержня, при котором достигается необходимое поглощение, получается значительно меньшим, чем объем стержня из парамагнетика. Расположение феррита в усилителе показано на рис. 11.8. Такой феррит очень слабо взаимодействует с полем накачки, что выгодно отличает его от парамагнитного кристалла аналогичного назначения. Для успешной работы усилителя бегущей волны

не нужны никакие внешние развязывающие устройства типа вентиля или циркулятора. Это обстоятельство делает усилители бегущей волны еще более малошумящими приборами, чем резонаторные квантовые усилители.

КПУ бегущей волны по сравнению с резонаторными обладают и другими преимуществами. Ширина полосы пропускания усилителя бегущей волны на порядок больше, чем в резонаторном усилителе и может достигать (или даже превышать) ширины линии парамагнитного резонанса активного вещества (для рубина — 75—100 МГц). Важным преимуществом усилителя бегущей волны является возможность его электронной перестройки путем изменения магнитного поля, частоты накачки или одновременного изменения этих параметров.

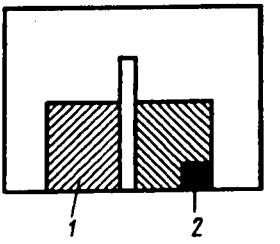


Рис. 11.8. Расположение кристалла рубина *1* и развязывающего феррита *2* в усилителе бегущей волны

В усилителях бегущей волны объем усиливающего материала во много раз больше (обычно в 10 раз), чем у резонаторных усилителей, поэтому и его выходная мощность также увеличивается (в 10 раз). В усилителе бегущей волны намного выше стабильность усиления.

Наилучшие коэффициент усиления, полоса пропускания и низкие входные шумы можно получить только при охлаждении КПУ до температуры жидкого гелия.

§ 11.7. Квантовый генератор

Квантовый усилитель состоит из двух основных частей: активной среды и системы возбуждения (генератора накачки). Классический метод превращения усилителя в генератор связан с введением в усилитель обратной связи, возвращающей часть энергии резонансной системы в эту же систему. В связи с этим можно считать, что квантовый генератор включает в себя три основных элемента: систему возбуждения, активную среду и систему обратной связи.

Основные элементы квантового генератора. Специфика обратной связи в квантовом генераторе обусловлена в первую очередь принципом квантового усиления в активной среде.

При рассмотрении процесса усиления можно было убедиться, что при достаточных инверсиях населенностей уровней ($K_0 > K_\sigma$) и пути (z) пробега волны активная среда уже не выполняет функции усилителя, поскольку амплитуда волны перестает зависеть от значения на входе. Более того, предельное значение интенсивности (11.50) достигается и в том случае, когда на входе вообще отсутствует излучение: $I(0) = 0$ (для запус-

ка лавинного стимулированного умножения фотонов достаточно одного фотона спонтанного излучения). Это означает, что при достаточно больших z активная среда начинает *генерировать* электромагнитное излучение. Переход активной среды из режима квантового усиления в режим стационарной генерации обусловлен эффектом насыщения усиления.

Главным условием возникновения генерации является, таким образом, определенная *протяженность* активной среды в направлении распространения волны z. Если это условие выполнено, то в квантовом генераторе не требуется специально сконструированного устройства для осуществления обратной связи. Здесь энергия от одной резонансной микросистемы (от одного активного центра или осциллятора) суммируется с энергией других резонансных микросистем (активных центров или осцилляторов) самим процессом квантового усиления. Оптимальное (синфазное, синхронное) суммирование обеспечивается автоматически фазовыми соотношениями, свойственными индуцированному излучению. Следовательно, в квантовом генераторе автоматическая реализация обратной связи связана с природой индуцированного излучения. Синхронность обусловлена также положительной обратной связью, осуществляемой при индуцированном испускании под действием излучения, накопленного в резонаторе в результате многократных отражений от стенок или зеркал резонатора. Процесс квантового усиления определяет связанную с этим высокую степень *когерентности* генерируемого излучения.

Практическое обеспечение основного условия генерации сводится к конструктивному обеспечению фактора пространственной избирательности: размеры активной среды должны быть достаточно большими в продольном направлении генерации и достаточно малыми в поперечных направлениях. Подобная избирательность может быть осуществлена тонким длинным активным волокном. Однако преимущественное распространение получил более компактный метод, в котором путь волны в активной среде увеличивается не столько за счет ее продольного размера, сколько за счет многократного прохождения отраженного излучения. В генераторах СВЧ для этой цели применяются *объемные резонаторы*, в оптическом диапазоне — параллельные или конфокальные зеркала, образующие *открытый резонатор*. Особенно широко распространены в настоящее время оптические квантовые генераторы (ОКГ), которые иногда называют оптическим мазерами или лазерами (laser — Light Amplification by Stimulatied Emission of Radiation). Появление лазеров ознаменовало революцию в оптике, ибо ранее использовавшиеся источники света не были когерентными и не были монохроматичными. Высокая степень когерентности излучения лазеров позволила манипулировать светом почти так же, как радиоволновым и микроволновым излучениями. Дальнейшее изложение будет посвящено именно этому виду квантовых генераторов.

Зеркала оптического резонатора обеспечивают многократное отражение, прохождение и максимальное усиление в активной среде только тех волн, которые распространяются в одном продольном направлении и синфазно складываются друг с другом. Именно таким образом резонатор обеспечивает фактор избирательности, благодаря которому сохраняются или выделяются только «разрешенные» фотонные состояния (стимулированные и спонтанные), формируется когерентный и *остронаправленный* световой луч. Именно в этом заключается главная функция резонатора. Обратная связь обеспечивается квантовым усилением и не является принципиальной функцией оптического резонатора.

При достаточно больших длинах активной среды даже в отсутствие резонатора формируется когерентное излучение, хотя и с относительно низкой монохроматичностью и направленностью. Оптические генераторы, не имеющие резонаторов, иногда называют лазерами с сверхлюминесценцией.

Расположение квантовой системы в объемном резонаторе, предложенное одновременно Н. Г. Басовым и А. М. Прохоровым в СССР и Д. Гордоном, Ч. Таунсом и Г. М. Цайгером в США, позволило создать первые квантовые генераторы. Именно этот момент считают началом возникновения квантовой электроники. Следующим важным шагом было предложение Н. Г. Басова и А. М. Прохорова (1955 г.) осуществить накачку трехуровневым радиационным методом. Предложение А. М. Прохорова использовать открытые резонаторы (1958 г.) определило возможность применения квантовой электроники в оптическом диапазоне.

Источником энергии квантового генератора является система возбуждения, обеспечивающая инверсию населенностей рабочих уровней активной среды. По типу фазового состояния активного вещества оптические квантовые генераторы подразделяют на газовые, твердотельные и жидкостные.

Порог стационарной генерации. При анализе порогового условия для дипольного перехода, в котором в результате возбуждения имеется начальная инверсия населенностей $N_{21}^{н}$ (в отсутствие излучения на частоте перехода), можно обратиться к кинетическому уравнению (11.55), переписав его в более удобной форме

$$\frac{\mathrm{d}n_{\nu}}{\mathrm{d}t}+\frac{n_{\nu}}{\tau_c}=\frac{N_{21}}{N_0}\frac{n_{\nu}}{\tau_c}, \tag{11.60}$$

где

$$N_0=\frac{3\hbar\varepsilon}{\tau_c\Omega T_2 L|\mu_{12}|^2} \tag{11.61}$$

— стационарное значение инверсии населенностей N_{21}, получаемое приравниванием нулю производной в уравнении (11.55).

Второе слагаемое левой части уравнения (11.60) характеризует скорость потерь энергии в резонаторе, а правая часть — скорость поступления энергии за счет вынужденного излучения. Для появления генерации необходимо, чтобы начала возрастать энергия (плотность фотонов) вынужденного излучения

$$\frac{\mathrm{d}n_\nu}{\mathrm{d}t}=\left(\frac{N_{21}}{N_0}-1\right)\frac{n_\nu}{\tau_c}>0,$$

т. е., чтобы поступление энергии стало превышать ее убыль и выполнялось условие $N_{21}>N_0$. На пороге генерации плотность электромагнитной энергии сначала исчезающе мала ($n_\nu\to 0$), и начальное значение инверсной населенности N_{21} определяется ее стационарным значением ($\mathrm{d}N_{21}/\mathrm{d}t=0$) в отсутствие излучения ($n_\nu=0$), т. е. будет, согласно (11.54), совпадать со стационарном значением $N_{21}^{\text{н}}$. Поэтому условием генерации является $N_{21}^{\text{н}}>N_0$. Пороговое значение инверсии населенностей

$$N_{21}^{\text{пор}}=N_{21}^{\text{н}}=N_0, \tag{11.62}$$

так как при большем значении имеет место нарастание электромагнитных колебаний, а при меньшем устойчивые колебания существовать не могут.

До сих пор предполагалось, что активная среда целиком заполняет резонатор и квантовая система не вырождена. Если источник поляризации заполняет активный объем V_a, меньший объема резонатора V, то поправку вводят с помощью коэффициента заполнения

$$F=\int_{V_a}E^2\,\mathrm{d}V\Big/\int_{V}E^2\,\mathrm{d}V.$$

С учетом вырождения и коэффициента заполнения *условие генерации* приобретает вид

$$\frac{N_2^{\text{н}}}{g_2}-\frac{N_1^{\text{н}}}{g_1}>\frac{3\hbar\varepsilon}{\tau_c T_2\Omega L|\mu_{12}|^2F},$$

где g_1, g_2 — кратности вырождения первого и второго энергетических уровней.

Время затухания резонатора τ_c первоначально было введено с учетом только электропроводности среды. Однако оно может учитывать все нерезонансные механизмы потерь электромагнитной энергии, поскольку при этом дифференциальные уравнения ангармоничного колебания (осциллятора) сохраняют свой вид. В частности, для оптического резонатора

$$\frac{1}{\tau_c}=\frac{1}{\tau_m}+\frac{1}{\tau_s}+\frac{1}{\tau_d}\approx\frac{2}{T_0}\frac{P_l}{P}, \qquad (11.63)$$

где τ_m — время затухания, определяемое потерями на излучение через зеркала; $\tau_s=1/(v\alpha)$ — время затухания, характеризуемое нерезонансными потерями на поглощение и рассеяние (α — коэффициент нерезонансных потерь); τ_d — время затухания, определяемое дифракционными и другими потерями, зависящими от формы зеркал, их размеров и расстояния между ними; $T_0=l/v$ — время прохождения светом расстояния l между зеркалами в резонаторе; P_l/P — относительные потери за одно прохождение.

Напомним, что кинетические уравнения (11.55) и (11.60) соответствуют случаю, когда нет спонтанного излучения и ширина линии резонатора $1/\tau_c$ много больше однородной ширины линии молекулярного резонанса $2/T_2$. В оптических резонаторах обычно это условие выполняется. Например, в рубине $2/T_2\approx$ 230 ГГц, а в резонаторе длиной 30 см с $P_l/P=0{,}25$ шириной линии $1/\tau_c\approx$ 80 МГц. При выводе условия генерации в целях упрощения предполагалось также, что в исходных уравнениях населенности и другие параметры, зависящие от координат, входят в качестве своих пространственно усредненных значений. В конечный результат это вносит незначительную погрешность.

В общем случае условие генерации сводится к требованию того, чтобы усиление в активном веществе за одно полное прохождение излучения через систему зеркал резонатора и активное вещество превышало все потери энергии в резонаторе за то же время.

Приближенное условие генерации можно получить из (11.49) для низкого уровня электромагнитного сигнала [$I(0)\approx 0$], учитывая только потери на излучение через зеркала оптического резонатора ($K_\sigma=0$). Излучение какой-либо точки активной среды испытывает два отражения от зеркал резонатора, прежде чем оно вернется в ту же точку и будет иметь то же направление. При каждом прохождении через активную среду интенсивность излучения увеличивается в $e^{K_0 l}$ раз (l — расстояние между зеркалами), а за полный цикл с двумя отражениями — в $G=R_1R_2\,e^{2K_0 l}=e^{2(K_0 l-\gamma)}$ раз, где R_1, R_2 — коэффициенты отражения зеркал; $\gamma=-\ln\sqrt{R_1R_2}$ — коэффициент потерь на излучение через зеркала; $K_0\approx\dfrac{\hbar\Omega N^{н}}{2T_1I_{н}}=\sigma_0\left(\dfrac{g_1}{g_2}N_2^{н}-N_1^{н}\right)$ — коэффициент усиления индуцированного излучения; $\sigma_0=\hbar\Omega/(2g_1T_1I_{н})$ — *поперечное сечение* излучения (поглощения).

Излучение затухает, если $G < 1$. В случае $G > 1$ или $K_0 l > \gamma$ интенсивность излучения на резонансной частоте быстро возрастает до тех пор, пока она не станет настолько большой, что произойдет уменьшение инверсии населенностей и коэффициента усиления (11.48). Следовательно, в этом случае условие генерации можно записать в виде

$$\frac{g_1}{g_2} N_2^{\text{н}} - N_1^{\text{н}} = \frac{\gamma}{\sigma_0 l}. \qquad (11.64)$$

В кристалле бледно-розового рубина $g_1 = g_2$ и $\sigma_0 = 2{,}5 \cdot 10^{-20}$ см.

Пороговая мощность возбуждения. Время продольной релаксации T_1 характеризует, согласно (11.31), скорость возвращения разности населенностей энергетических уровней $N_{12} = N_1 - N_2$ к равновесному значению $N_{12}^{\text{е}} = N_1^{\text{е}} - N_2^{\text{е}}$ в отсутствие внешнего электромагнитного излучения на резонансной частоте $\omega = \Omega$. Это время по (11.51) характеризует и скорость возвращения инверсии населенностей N_{21} к стационарному значению $N_{21}^{\text{н}} = N_2^{\text{н}} - N^{\text{н}}$ в процессе дипольного перехода после отключения внешнего электромагнитного сигнала на резонансной частоте, но при наличии постоянно действующей системы возбуждения. Следовательно, после отключения системы возбуждения инверсия населенностей $N_{21}^{\text{н}}$ возвращается к равновесному значению $N_{21}^{\text{е}}$ со скоростью $\dfrac{N_{21}^{\text{н}} - N_{21}^{\text{е}}}{T_1}$, характеризующей потери энергии в единицу времени за счет продольной релаксации. Эти потери должна компенсировать энергия источника возбуждения (накачки) для поддержания стационарного значения $N_{21}^{\text{н}}$. Таким образом, для обеспечения условия генерации (11.62) мощность накачки не должна быть меньше

$$\mathscr{P}_{\min} = \frac{\hbar\Omega}{2} \frac{N_0 - N_{21}^{\text{е}}}{T_1} = \frac{\hbar\Omega}{2T_1} \left(\frac{3\hbar\varepsilon}{\tau_c T_2 L |\mu_{12}|^2 F} - N_{21}^{\text{е}} \right). \qquad (11.65)$$

В оптическом диапазоне в условиях термодинамического равновесия практически все активные атомы N находятся на основном энергетическом уровне. Если нижний уровень *1* совпадает с основным, то $N_{21}^{\text{е}} \approx -N$. Если он располагается над основным на расстоянии нескольких $k_{\text{Б}}T$, то $N_{21}^{\text{е}} \approx 0$. Поэтому мощность накачки в четырехуровневой системе возбуждения значительно ниже, чем в трехуровневой. Для рубинового лазера обычно $N_0 \approx 10^{18}$ см$^{-3}$; $N \approx 2 \cdot 10^{19}$ см$^{-3}$; $T_1 \approx 4 \cdot 10^{-3}$ с; $\hbar\Omega \approx 2{,}7 \cdot 10^{-19}$ Дж и $\mathscr{P}_{\min} \approx 600$ Вт/см^3.

Пороговое значение мощности накачки (11.65) является приближенным, так как не учитывает потерь, связанных со спонтанными переходами и переходами на метастабильный уровень *2* с более высоких энергетических уровней трехуровневой и четырехуровневой систем возбуждения. Их можно рассмотреть с помощью понятия вероятности перехода и кинетических уравнений, аналогичных (11.57).

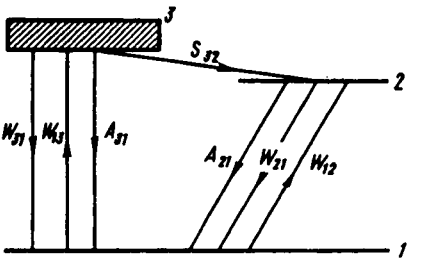


Рис. 11.9. Схема переходов в трехуровневом активном веществе:
W — вынужденные переходы; A — спонтанные излучательные переходы; S — спонтанный безызлучательный переход

Типичную трехуровневую систему образуют ионы хрома в кристалле розового рубина. Упрощенная диаграмма энергетических уровней иона Cr^{3+}, внедренного в решетку розового рубина, приведена на рис. 11.9, где основной уровень обозначен цифрой *1*. Возбуждение кристалла осуществляется в полосе волн 410...560 нм, поглощаемых внутри широкого уровня *3*. Бóльшая часть возбужденных ионов быстро ($\tau_{32} = 1/S_{32} = 5 \cdot 10^{-8}$ с) переходит безызлучательным путем на промежуточный узкий уровень *2*. Излучение, связанное со спонтанным возвращением с уровня *2* на основной уровень, проявляется в виде обычной флюоресценции. Она происходит при низком уровне излучения. При достаточно интенсивном возбуждающем излучении на уровне *2* можно получить больше атомов, чем их осталось на основном уровне. Спонтанное излучение, проходя через кристалл, индуцирует дополнительные фотоны. При этом индуцированное излучение может превзойти спонтанное. Вынужденное излучение возникает и в том случае, когда заполнение основного уровня *1* больше, чем заполнение уровня *2*, но поглощение больше, чем излучение и это вызывает уменьшение общего числа фотонов.

Имеется некоторая вероятность перехода ионов Cr^{3+} с уровня *3* не только на уровень *2*, но и на уровень *1*. Однако вероятность перехода на уровень *2* больше вероятности их перехода на уровень *1* ($S_{32} \gg A_{31}$). Время жизни $\tau_{31} = 1/A_{31} = 10^{-5}$ с. Вероятность перехода A_{21} с уровня *2* на уровень *1* меньше, чем вероятность перехода A_{31} с уровня *3* на уровень *1*. В рубине при комнатной температуре время жизни на метастабильном уровне *2* $\tau_{21} = 3$ мс.

Населенности уровней в трехуровневой системе рубина определяют кинетические уравнения

$$\frac{dN_3}{dt} = W_{13}N_1 - (W_{31} + A_{31} + S_{32})N_3 = W_{13}(N_1 - N_3) - \frac{N_3}{\tau_{31}} - \frac{N_3}{\tau_{32}};$$
$$\frac{dN_2}{dt} = W_{12}N_1 - (W_{21} + A_{21})N_2 + S_{32}N_3 = W_{21}(N_1 - N_2) - \frac{N_2}{\tau_{21}} + \frac{N_3}{\tau_{32}} \qquad (11.66а)$$

и уравнение баланса

$$N = N_1 + N_2 + N_3, \qquad (11.66б)$$

где W — вероятность вынужденного перехода, зависящая от плотности излучения; $A = 1/\tau$ — вероятность спонтанного перехода; S_{32} — вероятность безызлучательного перехода с уровня *3* на уровень *2*; N — полное число активных атомов в единице объема. С учетом (11.58) $W_{13} = W_{31}$ и $W_{12} = W_{21}$.

Упрощает нахождение стационарного ($d/dt = 0$) решения кинетических уравнений то обстоятельство, что в рассматриваемой системе вероятность безызлучательных переходов $S_{32} = \tau_{32}^{-1}$ обычно много больше вероятности излучательных переходов, в том числе и W_{21} (вблизи порога генерации $W_{21} \to 0$). В этом случае из (11.66)

$$\frac{N_2 - N_1}{N} \approx \frac{W_{13} - A_{21}}{W_{13} + A_{21} + 2W_{12}}. \qquad (11.67)$$

Из этого выражения следует, что *порогу инверсии* населенностей уровней в основном переходе соответствует вероятность $W_{13}^{\min} = A_{21} = \tau_{21}^{-1}$, а *порогу генерации*, согласно (11.64) и (11.67), — вероятность

$$W_{13}^{\text{пор}} \approx \frac{1 + \gamma/(\sigma_0 lN)}{1 - \gamma/(\sigma_0 lN)} A_{21}. \qquad (11.68)$$

Вероятность W_{13} пропорциональна мощности $\mathcal{P}_{13}$ излучения на резонансной частоте $\omega_{13} = \Omega_{13}$:

$$W_{13} = B_{13}u(\omega_{13}) = \varkappa_{13}n_{13} = \varkappa_{13}\mathcal{P}_{13}/(\hbar\omega_{13}S_{\text{н}}), \qquad (11.69)$$

где $\varkappa_{13}$ — поперечное сечение поглощения фотонов с частотой ω_{13} (в розовом рубине примерно 10^{-19} см^2); n_{12} — плотность этих фотонов; $S_{\text{н}}$ — поперечное сечение светового потока накачки, поглощаемого активной областью.

Пороговое значение мощности возбуждения $\mathcal{P}_{13}$ следует из (11.68) и (11.69)

$$\mathcal{P}_{\text{пор}} \approx \frac{\hbar\omega_{13}}{\tau_{21}\varkappa_{13}} \frac{1 + \gamma/(\sigma_0 lN)}{1 - \gamma/(\sigma_0 lN)} S_{\text{н}}. \qquad (11.70)$$

Приближенную оценку дает более простое выражение для порога инверсии: $\mathscr{P}_{\min} \approx \hbar\omega_{13}/(\tau_{21}\varkappa_{13})$. В рубине $\mathscr{P}_{\min}/S_{\text{н}} = 1200$ Вт/см^2.

Несмотря на приближенность выводов, соотношение (11.70) содержит основные факторы, позволяющие определить требуемую степень облучения активной среды и оценить их качественное влияние на эффективность накачки. В частности, из него следует, что необходимая мощность возбуждения пропорциональна вероятности спонтанного перехода и, следовательно, обратно пропорциональна времени жизни τ_{21} на метастабильном уровне. Очевидно также, что любое уменьшение коэффициента отражения зеркал резонатора должно быть скомпенсировано увеличением длины активной среды, в противном случае возрастает мощность, необходимая для выполнения условия генерации. Существенно также то, что генерация может быть сорвана уменьшением коэффициента отражения (увеличением γ).

В оптической системе возбуждения твердотельных лазеров обычно используются мощные импульсные лампы, окруженные со всех сторон вместе с активным веществом (стержнем) зеркальными отражателями. Поэтому электрическая мощность $P_{\text{н}}$, подводимая к лампе, намного больше мощности резонансного излучения:

$$\mathscr{P}_{21} = \eta_{\text{св}}\eta_{13}\eta_{\text{отр}}\eta_A\eta' P_{\text{н}} = \eta_{\text{н}} P_{\text{н}},$$

где $\eta_{\text{н}}$ — КПД системы накачки, учитывающий потери, связанные с превращением в импульсной лампе электрической энергии в световую ($\eta_{\text{св}}$), неиспользуемым излучением в других областях спектра (η_{13}), отражением от отражателей лампы ($\eta_{\text{отр}}$), а также со спонтанными переходами (η_A) и другими явлениями (η'). К ним относится, в частности, нерезонансное поглощение в активной среде, а также корректировка на реальную форму фронта светового потока накачки, отличную от формы фронта расчетной плоской волны.

В рубиновых лазерах обычно используется ксеноновая лампа. При этом $\eta_{\text{св}} = 0{,}5...0{,}6$; $\eta_{13} = 0{,}4...0{,}5$; $\eta_{\text{отр}} \approx 0{,}3$; $\eta_A \approx 0{,}8$; $\eta_{\text{н}} \simeq \eta_{\text{св}}\eta_{13}\eta_{\text{отр}}\eta_A \approx 0{,}05$. Таким образом, эффективность оптической накачки весьма мала, что определяет значительную пороговую мощность накачки $P_{\text{пор}} = \mathscr{P}_{\text{пор}}/\eta_{\text{н}}$ в рассмотренной системе возбуждения.

Выходная мощность в стационарном режиме. Стационарное значение плотности фотонов $n_0 = (N^{\text{н}} - N_0)\tau_{\text{с}}/(2T_1)$ следует из уравнения (11.54) при подстановке в него при $\mathrm{d}N_{21}/\mathrm{d}t = 0$ стационарного значения инверсной населенности (11.61). Соответствующая стационарная плотность энергии электромагнитного поля $w_0 = \hbar\Omega n_0 = \hbar\Omega\tau_{\text{с}}(N^{\text{н}} - N_0)/(2T_1)$.

Если суммарные потери лазера характеризовать постоянной времени (11.63), то в режиме релаксации $w(t) = w(0)\mathrm{e}^{-t/\tau_c}$, где $w(0)$ — начальная плотность энергии (при $t = 0$). При этом изменение энергии определяется всеми видами потерь. На это будет расходоваться мощность

$$\mathscr{P} = -\frac{\partial w}{\partial t} = w\left(\frac{1}{\tau_m} + \frac{1}{\tau_s} + \frac{1}{\tau_d}\right) = \mathscr{P}_m + \mathscr{P}_s + \mathscr{P}_d,$$

складывающаяся из мощностей потерь на излучение через зеркала ($\mathscr{P}_m = w/\tau_m$), нерезонансных потерь в активной среде ($\mathscr{P}_s = w/\tau_s$), дифракционных и других потерь света внутри резонатора ($\mathscr{P}_d = w/\tau_d$). В стационарном режиме все потери энергии в резонаторе покрывает генератор накачки мощностью $\mathscr{P}_{13} = \mathscr{P}_0 = w_0/\tau_c$.

Время релаксации τ_m можно вычислить, предполагая, что за время $t = 2l/v = 2T_0$, в течение которого электромагнитная волна проходит удвоенное расстояние между зеркалами, запасенная энергия $w = w(0)\mathrm{e}^{-t/\tau_m}$ уменьшится до значения $R_1R_2w(0)$. Поэтому

$$\tau_m = -\frac{2l}{v\ln(R_1R_2)} = \frac{T_0}{\gamma}. \qquad (11.71)$$

Однако то, что для резонатора в стационарном режиме будет мощностью $\mathscr{P}_m = w_0/\tau_m$, теряемой с излучением через зеркала для генератора, является выходной мощностью

$$\mathscr{P} = \frac{w_0}{\tau_m} = \gamma\hbar\Omega\tau_c\frac{N^{\text{н}} - N_0}{2T_0T_1} = \frac{\tau_c}{\tau_m}\mathscr{P}_{13}$$

в этом режиме.

Из вывода выражения (11.71) также следует, что для учета всех потерь в условии генерации (11.64) достаточно коэффициент потерь на излучение γ заменить коэффициентом потерь $\Gamma = T_0/\tau_c$ с учетом результирующего времени релаксации (11.63).

Расчет выходной мощности оптического квантового генератора может быть сведен к расчету такой его эквивалентной схемы, в которой потенциалам соответствуют населенности энергетических уровней, а проводимостям — вероятности энергетических переходов (сопротивлениям — времена соответствующих переходов). Например, уравнения (11.66) в стационарном режиме

$$W_{13}(N_1 - N_3) - (A_{31} + S_{32})N_3 = 0;$$
$$W_{12}(N_1 - N_2) - A_{21}N_2 + S_{32}N_3 = 0;$$

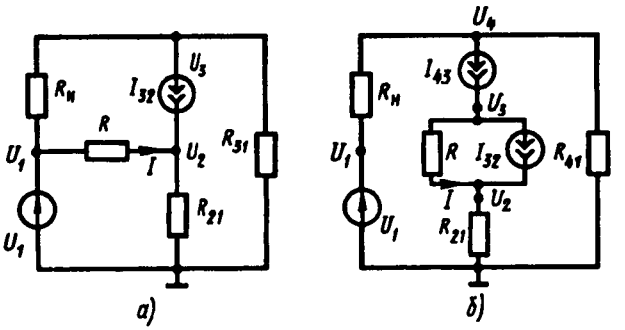


Рис. 11.10. Эквивалентные схемы трехуровневого (*а*) и четырехуровневого лазера (*б*)

$$N_1 + N_2 + N_3 = N$$

соответствуют уравнениям эквивалентной схемы

$$\frac{\varphi_1 - \varphi_3}{R_н} = \left(\frac{1}{R_{31}} + \frac{1}{R_{32}}\right)\varphi_3;$$

$$\frac{\varphi_1 - \varphi_2}{R} + \frac{\varphi_3}{R_{32}} = \frac{\varphi_2}{R_{21}}, \quad \varphi_1 + \varphi_2 + \varphi_3 = U,$$

где $R_{21} = 1/A_{21} = \tau_{21}$; $R_{32} = 1/S_{32} = \tau_{32}$; $R_{31} = 1/A_{31} = \tau_{31}$; $R = R_{12} = 1/W_{12} = 1/\sigma_{21}J$; $R_н = R_{13} = 1/W_{13} = 1/(\sigma_{13}I_н)$, J — интенсивность излучения; σ — поперечное сечение вынужденного излучения. Эквивалентная схема трехуровнего лазера (рис. 11.10, *а*) содержит источник ЭДС $U_1 = \varphi_1 = N_1$ и источник тока $I_{32} = \varphi_3/R_{32}$. Из нее следует, что при наличии инверсии выходной сигнал

$$I = (N_1 - N_2)W_{12} = \frac{U_2 - U_1}{R + \dfrac{2R_н/R_{31} + 2R_н/R_{32} + 3}{(R_н/R_{21} + 1)/R_{32} + (R_н/R_{31} + 1)/R_{21}}}.$$

Эквивалентная схема четырехуровневого лазера изображена на рис. 11.10, *б*, где $R_н = R_{14} = 1/W_{14} = 1/(\sigma_{14}J_н)$; $R_{21} = \tau_{21}$; $R_{32} = \tau_{32}$; $R_{41} = \tau_{41}$; $R_{43} = \tau_{43}$; $R = 1/W_{32} = 1/\sigma_{32}J$.

Пичковая генерация. В большинстве твердотельных лазеров на примесных кристаллах и стеклах используется импульсное возбуждение, причем длительность импульса накачки может быть сравнима с временами релаксации активной среды и резонатора (в рубиновых лазерах обычно так и бывает). В этом случае стационарный режим за

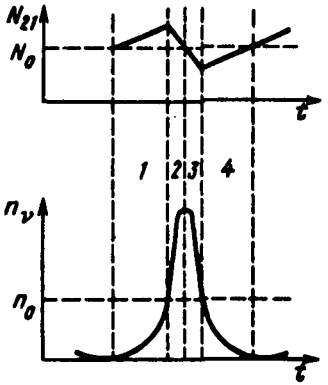


Рис. 11.11. Переходный процесс включения лазера и образования первого пичка

время импульса не может быть достигнут и генерацию когерентного излучения определяют нестационарные процессы, описываемые скоростными уравнениями (11.55) и (11.56), которые более удобно записать в виде

$$\frac{\mathrm{d}N_{21}}{\mathrm{d}t}+\frac{N_{21}-N_{21}^{\text{н}}}{T_1}=\frac{N_0-N_{21}^{\text{н}}}{T_1}\frac{N_{21}}{N_0}\frac{n_\nu}{n_0}, \tag{11.72}$$

$$\frac{\mathrm{d}n_\nu}{\mathrm{d}t}+\frac{n_\nu}{\tau_c}=\frac{N_{21}}{N_0}\frac{n_\nu}{\tau_c}. \tag{11.73}$$

Переходный процесс выходного излучения начинается с момента, когда инверсия населенностей N_{21} станет несколько превышать пороговое значение N_0. При умеренных мощностях накачки он протекает в четыре этапа (рис. 11.11).

1. Сначала, когда $n_0 \gg n_\nu \approx 0$, согласно (11.72), $\mathrm{d}N_{21}/\mathrm{d}t > 0$, и инверсия населенностей N_{21} возрастает. При этом по (11.73) $\mathrm{d}n_\nu/\mathrm{d}t > 0$ и энергия излучения (n_ν) начинает расти ($N_{21} > N_0$).

2. При умеренных мощностях накачки ($N_{21} \approx N_{21}^{\text{н}} \approx N_0$), после того как энергия излучения превысит стационарное значение ($n_\nu > n_0$), согласно (11.72), $\mathrm{d}N_{21}/\mathrm{d}t < 0$ и инверсия станет уменьшаться.

3. Когда инверсия населенностей станет меньше своего стационарного значения ($N_{21} < N_0$), одновременно с ее дальнейшим уменьшением начнет уменьшаться и энергия излучения (n_ν), поскольку в (11.73) $\mathrm{d}n_\nu/\mathrm{d}t < 0$. Этот процесс будет продолжаться до тех пор, пока энергия излучения не достигнет своего стационарного значения ($n_\nu = n_0$).

4. Дальнейшее уменьшение энергии излучения ($n_\nu < n_0$) сопровождается подъемом инверсии населенностей (поскольку при этом $\mathrm{d}N_{21}/\mathrm{d}t > 0$) вплоть до $N_{21} = N_0$, после чего система возвращается к условиям, близким к начальным, и цикл повторяется.

Таким образом, изменение энергии излучения и выходной мощности во времени будет представлять собой последовательность коротких импульсов (*пичков*), свойственных режиму, который получил название *пичковой генерации*. Пиковое значение мощности излучения может во много раз превышать ее стационарное значение. В рубиновых лазерах длитель-

ность пичка составляет около 1 мкс, что много меньше времени релаксации возбужденных ионов T_1 (около 4 мс).

Вблизи порогового режима ($N_{21} \approx N_0$) плотность энергии к концу каждого пичка спадает до уровня шумов и последующий пичок не коррелирован с предыдущим (*режим нерегулярных пичков*).

Когда мощность накачки существенно превышает пороговое значение и постоянная затухания резонатора достаточно велика, каждый последующий пичок коррелирован с предыдущим и они образуют регулярную последовательность импульсов с уменьшающейся амплитудой (*режим регулярных пичков*).

Модуляция добротности. На втором этапе рассмотренного нестационарного режима рост энергии излучения сопровождался уменьшением инверсии населенностей, поскольку при таких уровнях генерации опустошение метастабильного энергетического уровня за счет индуцированных переходов происходило быстрее восстановления его населенности от источника возбуждения. При этом началу излучения соответствовал сравнительно низкий уровень инверсии (11.61).

Начальное значение инверсии населенностей уровней может быть значительно больше N_0, если в переходном процессе начального роста инверсии под воздействием генератора накачки исключить опустошение метастабильного уровня индуцированным излучением. Тогда $W_{21} \approx 0$ и при достаточной мощности накачки ($W_{13} \to \infty$) или малой вероятности спонтанных переходов ($A_{21} \to 0$) инверсия населенностей (11.67) может вплотную подойти к своему предельному значению

$$N_{max} = (N_2 - N_1)_{max} \approx \left(\frac{W_{13} - A_{21}}{W_{13} + A_{21}}\right)_{max} N \to N.$$

Для этого достаточно исключить отражение от одного из зеркал ($R_1 R_2 = 0$), т. е. как бы выключить резонатор или сделать его добротность

$$Q = \omega\tau_c = \frac{\omega}{\tau_m^{-1} + \tau_s^{-1} + \tau_d^{-1}} = \frac{\omega}{\tau_d^{-1} + \frac{c_0\mu_0\sigma}{\sqrt{\varepsilon}} - \frac{v}{2l}\ln(R_1R_2)}$$

равной нулю.

После включения резонатора и восстановления первоначального значения его высокой добротности* происходит генерация излучения, называемого иногда гигантским импульсом, пиковая мощность которого может быть на несколько порядков выше, чем в стационарном режиме.

---

* Для рубиновых лазеров $\tau_c = 10^{-8}...10^{-9}$ с.

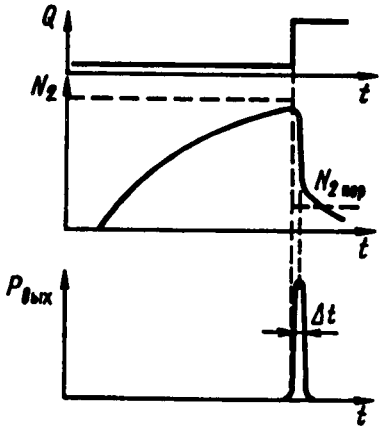


Рис. 11.12. Развитие лазерного импульса в режиме модулированной добротности

Значительный выигрыш в импульсной мощности достигается благодаря большей начальной энергии, запасаемой квантовой системой на метастабильном уровне, и более короткой длительности излучения этой энергии (рис. 11.12). Укорочение импульса происходит из-за большей скорости опустошения верхнего уровня ($\mathrm{d}N_{21}/\mathrm{d}t$) за счет большей разности начального и конечного состояний [сначала примерно в $(N-N_0)/(N_{21}^{н}-N_0)$ раз] и значительного увеличения вероятности индуцированных переходов W_{21} в условиях высоких плотностей энергии светового поля ($n_\nu \gg n_0$). Например, в рубиновых лазерах длительность импульса достигает 2—10 нс, а пиковая мощность — нескольких гигаватт.

Подобный способ генерации мощных импульсов называют *методом включения добротности* или *накопительным режимом*. При этом предполагают, что добротность резонатора изменяется за время, значительно меньшее времени жизни фотона τ_c.

Приближенный анализ рассматриваемого режима излучения с включением добротности можно привести в предположении, что уже в начальной стадии возникновения импульса плотность фотонов значительно превышает свое стационарное значение ($n_\nu \gg n_0$). Тогда в уравнении (11.72) можно пренебречь вторым слагаемым в левой части и с учетом (11.73) получить

$$\frac{\mathrm{d}n_\nu}{\mathrm{d}N_{21}} = -\frac{T_1}{\tau_c}\frac{N_{21}-N_0}{N_{21}^{н}-N_0}\frac{n_0}{N_{21}},$$

откуда

$$\frac{n_\nu}{n_0} = \frac{T_1}{\tau_c}\frac{N_0}{N_{21}^{н}-N_0}\left(\ln\frac{N_{21}}{N_{21}^{н}} + \frac{N_{21}^{н}}{N_0} - \frac{N_{21}}{N_0}\right), \qquad (11.74)$$

если $n_\nu = 0$ при $N_{21} = N_{21}^{н}$.

Поскольку пиковое значение излучаемой мощности ($\mathrm{d}n_\nu/\mathrm{d}t = 0$), согласно (11.73), соответствует моменту, когда $N_{21} \approx N_0$, то пиковое значение плотности фотонов (11.74)

$$n_{max} \approx n_0 \frac{T_1}{\tau_c} \frac{N_0}{N_{21}^{н} - N_0} \left(\ln \frac{N_0}{N_{21}^{н}} + \frac{N_{21}^{н}}{N_0} - 1 \right).$$

В рубиновом лазере при $N_{21}^{н}/N_0 = 2...4$ отношение $n_{max}/n_0 = 10^6$ и плотность пиковой мощности примерно равна 10 МВт/см^3.

Модуляция добротности может осуществляться несколькими методами. Для этого, например, можно использовать в качестве одного из зеркал *вращающуюся призму* с полным внутренним отражением. Резонатор оказывается включенным только в том случае, когда призма располагается параллельно неподвижному (выходному) зеркалу. Недостатком подобных механических методов является сравнительно медленное включение добротности. Даже при частотах вращения 20 000—30 000 об/мин время включения составляет 10^{-7} с. Преимуществами таких методов являются простота устройства, малые оптические потери, небольшие габариты, малое потребление энергии и устойчивость в работе.

Для управления добротностью применяют и электрооптические эффекты. В этом случае (рис. 11.13) между активной средой *1* и одним из зеркал располагают *электрооптический затвор*, состоящий из поляризационной призмы *2* и электрооптического кристалла *3*. Электрооптические затворы отличаются высоким быстродействием (до 10^{-9} с), отсутствием подвижных частей, возможностью электрического управления и высокой стойкостью к воздействию мощности оптического излучения. Осложняют их применение заметные оптические потери, высокие управляющие напряжения и необходимость точной юстировки.

Акустооптический затвор в резонаторе позволяет модулировать добротность путем отклонения светового потока от первоначального направления в результате дифракции лазерного луча на ультразвуковой волне, возбуждаемой в фотоупругой среде. В лазере ЛТИ-501 для этой цели используется высококачественный кварц. Просветляющие покрытия на боковых гранях звукопровода позволяют снизить остаточные отражения до значений, не превышающих 0,5%.

Простотой конструкции и высоким быстродействием обладают *фототропные затворы*, работа которых основана на эффекте просветления некоторых сред с нелинейными свойствами на резонансной частоте. К таким средам относятся ураниловые стекла и стекла КС-19, а также растворы органических красителей (полиметиновых; крипто- и фталоцианина).

Упрощенная диаграмма энергетических уровней фталоцианина ванадия изображена

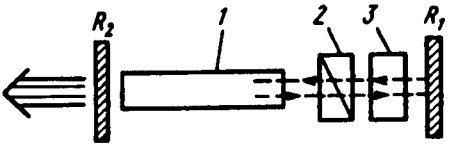


Рис. 11.13. Оптический квантовый генератор с электрооптическим модулятором

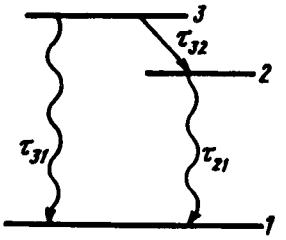


Рис. 11.14. Диаграмма энергетических уровней фталоцианина ванадия

на рис. 11.14. Частота перехода между уровнями *1* и *3* находится в красной области спектра и подбором подходящего растворителя, например, нитробензола или толуола совмещается с частотой когерентного излучения рубинового квантового генератора. Промежуточный уровень *2* является метастабильным ($\tau_{21} \approx 10^{-3}$ с), а переход с уровня *3* на уровень *2* — безызлучательным ($\tau_{32} \approx 10^{-7}$ с). Излучение при переходе с метастабильного уровня на основной происходит в инфракрасном диапазоне. Поскольку длительность импульса излучения генератора в данном случае мала по сравнению с временем перехода τ_{32}, то фототропную среду можно рассматривать как двухуровневую систему. Прозрачную кювету с раствором располагают между рубиновым стержнем и 100%-ным зеркалом ($R = 1$); она выполняет функции оптического затвора.

При малых уровнях спонтанного излучения рубина молекулы красителя сосредоточены на основном уровне *1*, коэффициент поглощения фототропной среды имеет максимальное значение. Поэтому потери, вносимые фототропной средой, не обеспечивают условия возникновения генерации. При приближении населенности метастабильного уровня *2* рубина к предельному значению мощность излучения возрастает, уменьшая разность населенностей уровней *1* и *3* фталоцианина, т. е. уменьшая его показатель поглощения. Когда показатель поглощения уменьшится до значения, при котором выполняется условие самовозбуждения, начинается генерация. При этом происходит быстрое насыщение перехода *1—3* красителя, после чего фототропная среда оказывается практически прозрачной. Благодаря тому, что отношение поперечных сечений поглощения фталоцианина и рубина велико ($\sigma_ф/\sigma_{Cr} \approx 10^4$), лавинообразный процесс просветления протекает практически мгновенно.

При этом рубиновый лазер генерирует импульс излучения длительностью 10^{-8} с КПД $\eta \approx 0{,}3\%$. После окончания индуцированного излучения добротность резонатора примет первоначальное (низкое) значение.

Таким образом, управление добротностью происходит только за счет свойств нелинейной среды. При этом не требуется никаких управляющих электромеханических и электрооптических устройств, а также вспомогательных электронных и высоковольтных схем. Поэтому подобные затворы называют также *пассивными*. Их преимуществами являются простота конструкции и высокое быстродействие. Однако фототропные среды об-

ладают сравнительно низкой стойкостью к воздействию мощного светового излучения.

КПД квантовых генераторов с модулированной добротностью ниже, чем в режиме свободной генерации, поскольку управляющий элемент вносит в резонатор дополнительные потери и увеличиваются потери со спонтанными переходами как за время накопления предельной инверсии, так и после окончания «гигантского» импульса (импульс накачки обычно имеет значительно бóльшую длительность).

Дальнейшее повышение пиковой мощности и укорочение импульса излучения достигается в лазерах с *переключением добротности*, где формирование индуцированного излучения сначала происходит без его вывода через 100%-ые зеркала. Электрооптический модулятор выводит луч из резонатора, минуя выходное зеркало в момент достижения наибольшей плотности фотонного потока, образуя на выходе гигантский импульс, который практически за один проход выводит наружу всю энергию в «глухом» резонаторе.

§ 11.8. Синхронизация мод

Режим одномодовой генерации возникает в лазере, когда генерация происходит на одной из мод оптического резонатора. При генерации одновременно на нескольких собственных частотах резонатора возникает *режим многомодовой генерации*. При этом возможно взаимодействие мод между собой. Рассмотрим в качестве примера модуляцию на $2n + 1$ аксиальных модах, разделенных частотным интервалом (2.35). Тогда напряженность результирующего электрического поля волны

$$\dot{E}(t) = \sum_{l=-n}^{n} E_l \, e^{j[(\omega_0 + l\Delta\omega)t + \varphi_l]},$$

где ω_0 — собственная частота центральной моды ($l = 0$); E_l, φ_l — начальные амплитуда и фаза l-й моды.

Интерференция мод возникает, если фазы всех мод каким-либо образом синхронизированы. Это явление называется *синхронизацией мод*. При равенстве амплитуд синхронизированных аксиальных мод ($E_l = E_0$)

$$\dot{E}(t) = \sum_{l=-n}^{n} E_0 \, e^{j(\omega_0 + l\Delta\omega)t} = E_0 \frac{\sin[(2n+1)\Delta\omega t/2]}{\sin[\Delta\omega t/2]} e^{j\omega_0 t}.$$

Временна́я зависимость выходной мощности лазера

$$P(t) \doteq \frac{\sin^2[(2n+1)\Delta\omega t/2]}{\sin^2[\Delta\omega t/2]}$$

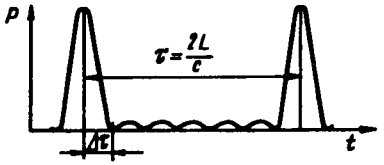


Рис. 11.15. Временна́я зависимость выходной мощности в режиме синхронизации мод

приведена на рис. 11.15 в случае генерации $2n + 1 = 7$ синхронизированных мод.

Синхронизация и интерференция мод приводит к тому, что энергия излучается в виде последовательности коротких световых импульсов, разделенных временем $\tau = 2\pi/(\Delta\omega) = 1/(\Delta\nu) = 2L/c$, необходимым фотону для полного прохода резонатора. Ширина импульса приблизительно равна времени наступления ближайшего минимума:

$$\Delta\tau = \frac{\tau}{2n+1} \approx \frac{1}{\Delta\nu_c}. \qquad (11.75)$$

Следовательно, для получения очень коротких световых импульсов необходимо возбуждение модового спектра большой ширины, т. е. нужны активные материалы с широкой линией люминесценции $\Delta\nu_c \approx (2n+1)\Delta\nu$. В газовых лазерах ширина спектральной линии усиления составляет около 10^9 Гц. Синхронизация мод позволяет получить импульсы длительностью около 1 пс. В твердотельных лазерах уширение линии люминесценции за счет сильного поля кристаллической решетки дает возможность генерировать импульсы длительностью 1 пс и менее. При этом пиковая мощность в импульсе в $2n + 1$ раз превышает сумму мощностей отдельных мод.

Каждый импульс может нести энергию около 1 Дж и иметь мощность около 10^{12} Вт. Если учесть, что за время длительности импульса 1 пс свет проходит расстояние 0,3 мм, то оказывается, что в области пространства 0,3 мм образуется сгусток мощностью 10^3 ГВт, движущийся со скоростью света. Примечательно, что выражение (11.75) для длительности волнового сгустка хорошо удовлетворяет соотношению неопределенностей $\delta\mathscr{E}\delta t = \hbar\Delta\nu_c\Delta\tau \approx \hbar$. Это позволяет рассматривать сгустки как своеобразные квазичастицы, подчиняющиеся многим законам квантовой механики.

Синхронизацию мод можно осуществить путем модуляции усиления и потерь в лазере с частотой, равной или кратной межмодовому интервалу (2.35).

Синхронизацию мод в случае внешней принудительной модуляции называют *активной*. Ее можно выполнить путем управления колебаниями глухого зеркала оптического резонатора или пропусканием электрооптического или акустооптического модулятора, введенного в резонатор, с частотой следования световых импульсов $\Delta\nu = c/(2L)$.

Пассивной синхронизацией называют самосинхронизацию в нелинейной среде, находящейся в резонаторе лазера.

Самосинхронизация мод в *активном элементе* основана на использовании взаимного влияния мод лазера в процессе установления инверсных населенностей атомных уровней. В зависимости от пространственной модовой структуры и типа уширения линии атомного перехода моды могут конкурировать либо существовать совместно. При этом происходит самомодуляция скоростей переходов и уровня инверсии. Излучение достигает максимального уровня, когда фазовые соотношения между оптимальными лазерными модами обеспечивают максимальную скорость вынужденных излучательных переходов.

Синхронизация в *поглотителе* (нелинейном фильтре) происходит также в результате самофазировки мод. Его начальное поглощение невелико и не препятствует возникновению генерации. Лазерное излучение «просветляет» фильтр. При генерации на нескольких модах зависимость поглощения от поля содержит члены, изменяющиеся с частотой межмодовых биений, т. е. возникает модуляция пропускания фильтра на частоте межмодовых биений, что и приводит к синхронизации мод. В этом случае просветляющийся фильтр работает как амплитудный модулятор («экспандер»), автоматически настраивающийся на частоту межмодовых биений.

Экспандер должен обладать быстрым нелинейным откликом: просветление, а также возврат в состояние с сильным поглощением должны происходить за время, меньшее двойного времени прохода света через резонатор. Таким очень небольшим временем возбужденных электронных состояний обладают молекулы красителей, применяемые для синхронизации мод.

Оптическое просветление в стационарном режиме было рассмотрение в § 11.5. Однако при распространении коротких мощных световых импульсов возникает новое своеобразное явление.

Экспериментальные и теоретические исследования ультракороткого лазерного импульса ($\tau < T_2$) в пассивной двухуровневой среде на резонансной частоте показали, что волна быстро переходит в волновое поле, которое содержит несколько импульсов, четко локализованных в пространстве и распространяющихся в пространстве с постоянными амплитудой и скоростью на два-три порядка меньше фазовой скорости света в данной среде. После столкновения такие импульсы сохраняют свою первоначальную форму, амплитуду и скорость, т. е. обладают свойствами солитонов. Световые солитоны распространяются сквозь оптическую среду, как сквозь полностью прозрачную систему. Это явление получило название *самоиндуцированной прозрачности*. Оно обусловлено инвертированием населенностей энергетических уровней на фронте импульса и

их переходом в основное состояние с индуцированной эмиссией на спаде импульса, когда запасенная в возбужденной квантовой системе энергия возвращается обратно. В этих условиях нелинейные волны с дисперсией в виде солитонов ведут себя подобно материальным частицам, демонстрируя и здесь дуализм волна-частица.

Математическое моделирование этой задачи в приближении плоской волны в среде без потерь исходит из волнового уравнения (11.32) в виде

$$\frac{\partial^2 E}{\partial t^2} - v^2 \frac{\partial^2 E}{\partial x^2} = -\mu_0 \frac{\partial P^2}{\partial t^2}$$

и уравнения Шредингера для двухуровневого атома в виде

$$j\hbar \frac{\partial}{\partial t}\begin{bmatrix}\dot{\Psi}_2 \\ \dot{\Psi}_1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}\hbar\omega & 2E \\ 2E & -\hbar\omega\end{bmatrix}\begin{bmatrix}\dot{\Psi}_2 \\ \dot{\Psi}_1\end{bmatrix},$$

где $\dot{\Psi}_2$ соответствует возбужденному состоянию атома, а $\dot{\Psi}_1$ — основному. Предполагается, что индуцированная поляризация $P \doteq \langle \dot{\Psi}_2 \dot{\Psi}_1^* + \dot{\Psi}_1 \dot{\Psi}_2^* \rangle = -\sin\varphi$ определяется усреднением электрических моментов множества микроскопических осцилляторов. Дальнейший расчет приводит к уравнению синус-Гордона

$$\frac{\partial^2 \varphi}{\partial x^2} - \frac{\partial^2 \varphi}{\partial t^2} = \sin\varphi,$$

решение которого имеет вид

$$\varphi_{\pm} = 4\,\mathrm{arctg}\left[\exp\left(\pm \frac{x - ut}{\sqrt{1-u^2}}\right)\right].$$

Функция φ_+ определяет свойства солитона, функция φ_- — свойства антисолитона; $\sin\varphi \doteq P$; огибающая импульса $E_0 \doteq \partial\varphi / \partial t$.

При образовании солитонов в одномодовых световодах стабилизация формы импульса была связана с компенсацией уширения импульса из-за дисперсии нелинейным изменением показателя преломления. При синхронизации мод с помощью насыщающегося поглотителя стабилизация формы импульса обусловлена компенсацией уширения импульса $\Delta\omega/(\Delta\omega)_{I\to 0}$ насыщением поглощения.

Синхронизация мод насыщающимся поглотителем позволяет получить импульсы много короче, чем при активной синхронизации мод. При активной синхронизации длительность импульса соответствует времени полного прохода резонатора излучением. В насыщающемся поглотителе изменения коэффициента поглощения сравнимы по длительности с шириной импульса и могут быть меньше времени релаксации обычной

поглощающей среды. При насыщении поглощения «центр» мощного импульса ослабляется меньше, чем малоинтенсивные «крылья». За счет этого и происходит сужение проходящего импульса.

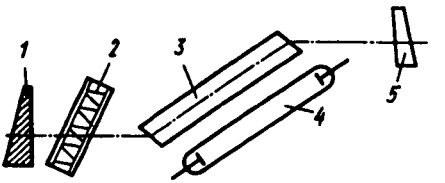


Рис. 11.16. Схема лазера с пассивной синхронизацией мод

В лазере (рис. 11.16) насыщающийся фильтр *2* расположен около глухого зеркала *1*. Торцы активного элемента *3* расположены под углом Брюстера, что исключает дополнительные отражения внутри резонатора (*4* — система накачки). Грани оптических элементов, в том числе, и зеркал *1* и *5* слегка отклонены от нормали к оптической оси резонатора для исключения образования дополнительных резонаторов со своим периодом межмодовых биений внутри основного. В лазерных генераторах такого типа длительность импульса составляет 1...10 пс, а энергия цуга импульсов — 1..2 Дж. Одними из распространенных импульсных лазеров высоких энергий являются те, которые используют в качестве активной среды стекла, активированные ионом Nd^{3+}, и четырехуровневую систему возбуждения. Они позволили получить импульсную энергию до десятков килоджоулей. Высокая оптическая однородность стекол является одним из его достоинств. Однако их теплопроводность низка. Это приводит к возникновению в стекле неоднородного температурного поля, термоупругих напряжений и оптических искажений, ограничивает их лучевую стойкость.

Вклад в разрушение материала вносит и эффект самофокусировки, обусловленный нелинейностью показателя преломления активной среды $n \approx n_0 + n_2 E^2$. Нелинейный показатель преломления n_2 в лучших лазерных стеклах составляет менее 10^{-22} $м^2В^{-2}$; для рубина и иттриево-алюминиевого граната (ИАГ) он в 2—4 раза выше.

Синхронизация продольных мод означала прорыв в область пикосекундных лазерных импульсов. Их стали называть сверхкороткими. Пиковые мощности таких импульсов доходили до 10 ГВт.

Рубеж 10^{-13} с был преодолен в начале 80-х годов, после чего началось быстрое освоение фемтосекундного диапазона длительностей сигналов (1 фс = 10^{-15} с). К 1987 г. были достигнуты предельно возможные длительности световых импульсов, равные одному периоду колебаний в инфракрасном диапазоне (40 фс) и трем периодам в видимом диапазоне (6 фс). Это область ультракоротких импульсов.

§ 11.9. Формирование ультракоротких импульсов

Ультракороткие импульсы получают сжатием (компрессией) сверхкоротких импульсов. Источником световых сверхкоротких импульсов обычно служат пикосекундные лазеры с синхронизацией мод; в некоторых случаях одновременно с синхронизацией мод в них используют модуляцию добротности первичного лазера. Компрессия импульсов происходит в результате быстрого самовоздействия в пассивной среде, приводящего к самомодуляции амплитуды и фазы компонент сигнала.

Впервые ультракороткие импульсы были получены в лазерах с новой разновидностью пассивной синхронизации мод — с синхронизацией мод сталкивающихся импульсов. Интерференция встречных импульсов в нелинейном поглотителе приводит приблизительно к четырехкратному сокращению длительности генерируемого импульса по сравнению с режимом, когда импульсы проходят поглотитель поодиночке. Таким способом в лазере на красителе были сформированы импульсы длительностью 56 фс.

Длительность оптических импульсов, генерируемых лазерами, может быть уменьшена в несколько раз при использовании свипирования (модуляции) несущей частоты (так называемого чирпа) при дисперсионном самовоздействии. Предельно короткие световые сигналы формируют сжатием фазомодулированного импульса в диспергирующем устройстве, называемом компрессором.

Сжатие фазово-модулированного сигнала оптическим компрессором — это временной аналог пространственной фокусировки светового пучка с помощью линзы. «Фокусировка» во времени осуществляется специальным подбором начальной фазовой модуляции импульса и знака дисперсии. В общем случае устройство сжатия («лупа времени») состоит из модулятора (фазового) и компрессора, осуществляющего фазировку фурье-компонент сигнала.

Фазовая самомодуляция может происходить, как было показано в §§ 3.15, 7.6 и 7.7, в трех режимах: бесдисперсионном (нелинейном), дисперсионном (линейном) и смешанном (солитонном).

Компрессию сигнала на временной плоскости можно трактовать как результат «нагона» в диспергирующей линии задержки низкочастотных спектральных компонент, располагавшихся на фронте импульса, высокочастотными компонентами, первоначально сгруппированными на его хвосте. Такой процесс происходит в диспергирующей среде и, в частности, в световодах с аномальной дисперсией.

Компрессию импульсов осуществляют, используя дисперсию и в процессах дифракции и преломления. В простейшем решеточном компрессоре световой луч проходит последовательно через пару дифракционных решеток, расположенных параллельно друг другу. Существенно

бóльшие значения дисперсии можно реализовать с помощью отражающих решеток. Разработаны также призменные компрессоры, в которых световой луч проходит через несколько прозрачных призм. Их преимущества — малые энергетические потери, отсутствие пространственного смещения частот и возможность компенсации дисперсии высоких порядков.

Для формирования и генерации фемтосекундных лазеров возможно также применение солитонных эффектов. Для получения импульсов предельно малой длительности было осуществлено самосжатие в волоконном световоде не только одного солитона, но и N-солитонного импульса. Подобные процессы лежат в основе работы *солитонных лазеров*. В одном из них световой луч проходит через два связанных резонатора. Основной резонатор представляет собой синхронно-накачиваемый лазер, генерирующий в автономном режиме импульсы длительностью 8 пс. В солитонном режиме часть излучения вводится во вспомогательный резонатор, содержащий одномодовый волоконный световод. Испытавший в световоде самосжатие импульс инжектируется в основной резонатор синхронно с накачкой, что приводит к генерации более короткого импульса и т. д. вплоть до выхода на стационарный режим. В стационарном режиме импульс представляет собой связанное состояние двух солитонов. Минимальная длительность импульсов в лазере была доведена до 100 фс.

Эффективные источники ультракоротких импульсов открывают возможность реализации предельных скоростей обработки и передачи информации. Переход к фемтосекундным импульсам — это и очередной скачок по шкале интенсивностей. Так, длительности импульса 100 фс и сравнительно небольшой его энергии 0,1 Дж соответствует уровень мощности 10^{12} Вт, который недавно удавалось получить только в мультикилоджоульных установках для управляемого термоядерного синтеза. В поле сфокусированных фемтосекундных импульсов получены напряженности светового поля 10^{10} В/см — на порядок выше, чем во внутриатомных полях.

§ 11.10. Инжекционные лазеры

Генерация излучения в полупроводниках происходит на переходах между зоной проводимости и валентной зоной, между зонами и примесными уровнями или между примесными уровнями.

Действие инжекционного лазера основано на стимулированной излучательной рекомбинации электронов и дырок, заполняющих состояния вблизи дна зоны проводимости и вблизи потолка валентной зоны. В отличие от других твердотельных лазеров, где генерация определяется переходами между уровнями примесных ионов, в полупроводниках активным веществом является сама кристаллическая матрица, а примеси слу-

жат источниками электронов и дырок, рекомбинация которых приводит к возникновению фотонов. Полупроводники от других твердотельных активных материалов отличает также электропроводность, а от газовых систем — более широкие линии излучения и возможность создания высокой концентрации активных частиц. В квантовой электронике применить полупроводниковые кристаллы предложил Н. Г. Басов в 1959 г.

Инверсия населенностей уровней в полупроводниках может быть получена путем инжекции носителей заряда, облучения пучком быстрых электронов, оптической накачки и ударной ионизации (лавинным пробоем). Благодаря своей высокой эффективности и простоте реализации наиболее широкое распространение получил первый тип возбуждения.

Условие возникновения вынужденного излучения в *p-n*-переходе. Одной из особенностей полупроводников является возможность взаимодействия электромагнитного излучения не с двумя узкими энергетическими уровнями, а с целым спектром энергетических состояний. Наиболее отчетливо это можно продемонстрировать на процессе усиления электромагнитного излучения частоты ν при межзонных переходах, когда энергия фотона ($h\nu$) достигает ширины запретной зоны ($\mathscr{E}_g = \mathscr{E}_c - \mathscr{E}_v$).

Если число фотонов N, а вероятность нахождения электронов на энергетических уровнях, между которыми происходит переход ($\mathscr{E}_2 - \mathscr{E}_1 = h\nu \approx \mathscr{E}_c - \mathscr{E}_v$), — f_n и f_p, то в единицу времени число переходов с излучением

$$\frac{\partial N_{\text{изл}}}{\partial t} = AW_{\text{пр}} f_n (1 - f_p) N,$$

где A — коэффициент, пропорциональный концентрации неравновесных носителей; $W_{\text{пр}}$ — вероятность излучательного перехода.

Число переходов, связанных с фотонами и сопровождающихся поглощением,

$$\frac{\partial N_{\text{погл}}}{\partial t} = AW_{pn} f_p (1 - f_n) N,$$

где W_{pn} — вероятность перехода с поглощением энергии, причем для индуцированных переходов $W_{pn} = W_{\text{пр}}$.

Критерием инверсии населенностей уровней может служить преобладание индуцированного излучения над поглощением, когда

$$f_n = \left(1 + e^{\frac{\mathscr{E}_2 - F_n}{k_Б T}}\right)^{-1} > \left(1 + e^{\frac{\mathscr{E}_1 - F_p}{k_Б T}}\right)^{-1} = f_p$$

или

$$F_n - F_p > \mathscr{E}_c - \mathscr{E}_v = \mathscr{E}_g. \qquad (11.76)$$

Отсюда следует ряд принципиальных выводов. Во-первых, инверсия населенностей возможна лишь в диодах, в которых хотя бы одна из областей образована вырожденным полупроводником (по квазиуровням Ферми). Во-вторых, инверсия населенностей может быть достигнута лишь при достаточных для этого уровнях инжекции, когда напряжение на p-n-переходе $U = (F_n - F_p)e^{-1} \geqslant \mathscr{E}_g e^{-1}$. В-третьих, частота излучаемых колебаний должна удовлетворять неравенству $eU > h\nu > \mathscr{E}_g$ (рис. 11.17).

Для получения инверсии в полупроводниковом диоде требуется сильное легирование. При достаточном внешнем напряжении смещения квазиуровень Ферми в n-области располагается выше дна зоны проводимости и в эту область происходит инжекция электронов, которые излучают фотоны $h\nu$ в процессе рекомбинации с дырками валентной зоны. Может происходить переход и дырок из p-области в n-область с последующей рекомбинацией в n-области. Таким образом, инверсия населенностей и излучение могут иметь место лишь в области, непосредственно примыкающей к плоскости p-n-перехода и ограничиваемой глубиной диффузии за время жизни неосновных носителей.

Ширина активной области обычно не превышает нескольких микрометров. Однако излучение распространяется в более широкой области. Одной из причин этого является не идеально плоская поверхность p-n-перехода. Поскольку поглощения квантов вынужденного излучения в неинвертируемых областях гомогенных переходов практически нельзя избежать, то его стараются свести к минимуму выбором соответствующих областей.

Несмотря на привлекательность рассмотренной модели межзонных оптических переходов, она не соответствует большинству экспериментальных данных: в инжекционных лазерах энергия излучаемых фотонов

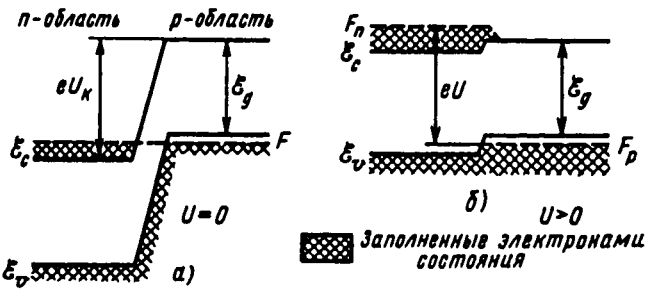


Рис. 11.17. Энергетическая диаграмма вырожденного p-n-перехода:
а — без смещения; *б* — при смещении

обычно меньше ширины запрещенной зоны. Это связано с использованием сильно легированных полупроводников, в которых излучательные процессы происходят с участием примесных состояний.

При рассмотрении сильно легированных полупроводников можно использовать в качестве первого приближения результаты рассмотренной модели межзонных переходов, внеся поправки, отражающие влияние примесей. Например, когда в активной области доминируют оптические переходы между зоной проводимости и акцепторным уровнем, условие инверсии представляют в виде $F_n - F_p \geqslant h\nu \geqslant \mathscr{E}_c - \mathscr{E}_A$, где $\mathscr{E}_A$ — энергия акцепторного уровня.

Однако ближе к опытным данным более строгий подход, учитывающий реальное заполнение квантовых состояний и их участие в индуцированных переходах. Имеется в виду, что у краев зоны проводимости и валентной зоны плотность квантовых состояний асимптотически спадает внутрь запрещенной зоны, образуя так называемые «хвосты». Кроме того, мелкие примесные уровни размываются в примесную зону, которая сближается и в пределе сливается с ближайшей зоной.

Экспериментальное и теоретическое рассмотрения показывают, что спектр излучения полупроводников обладает следующими качественными особенностями: значительной спектральной шириной и зависимостью от температуры, смещением пика полосы вынужденного излучения в сторону большей частоты с ростом тока накачки (по мере заполнения состояний), несовпадением спектров спонтанного и вынужденного излучений.

Устройство инжекционного лазера. Наибольшее распространение в полупроводниковых лазерах получил оптический резонатор из двух плоских зеркал, образуемых параллельными поверхностями кристалла. Коэффициент отражения от них (0,2...0,4) достаточен для создания необходимой положительной обратной связи. Оптическая ось резонатора (нормальная к поверхности зеркал) совпадает с направлением оптического потока и обычно расположена в плоскости p-n-перехода (*поперечная накачка*). Кристаллодержатель и электроды обеспечивают подвод тока и отвод теплоты от кристалла (рис. 11.18).

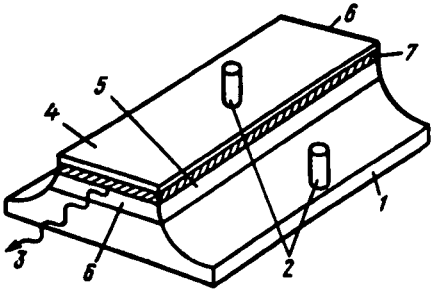


Рис. 11.18. Схема инжекционного лазера:
1 — молибденовая пластина; *2* — проводники; *3* — выходящее излучение; *4* — p-область; *5* — n-область; *6* — полированные торцевые поверхности; *7* — активная область кристалла

Эффект генерации когерентного излучения может быть обнаружен почти во всех прямозонных чистых монокристалли-

ческих полупроводниках. В непрямозонных полупроводниках вероятность непрямых излучательных межзонных переходов мала, не способна конкурировать с внутризонным поглощением и рекомбинацией через примеси, которая в большинстве этих материалов также не дает достаточного для генерации оптического усиления. Одним из наиболее распространенных материалов для инжекционных лазеров является арсенид галлия. Особенностью этого интерметаллического соединения является очень большая вероятность излучательной рекомбинации, а также доступность получения кристаллов необходимой чистоты. Электронно-дырочные переходы создаются обычно путем диффузии акцепторных примесей (цинк, кадмий и др.) в материал, легированный донорными примесями (теллур, селен, олово и др.). Благодаря высокому коэффициенту электромагнитного излучения условия генерации удается выполнить для миниатюрных образцов (от десятков микрометров до миллиметра).

Как показали исследования, активная область диффузионного диода образуется со стороны *p*-области. При очень большой концентрации свободных носителей заряда* излучательная рекомбинация подавляется ожерекомбинацией (ударная рекомбинация, сопровождаемая передачей кинетической энергии электронно-дырочной пары третьему свободному носителю заряда). Поэтому активный слой с одной стороны ограничен *p-n*-переходом, а с другой — границей высокой концентрации акцепторов.

Оптическая схема инжекционного лазера включает в себя кроме плоского активного слоя две прилегающие полубесконечные пассивные области с разными комплексными диэлектрическими проницаемостями. Распространение волны в активном слое связано с распространением в пассивных областях волн с затухающими амплитудами. Существование таких волн может быть обусловлено поглощением излучения вне активного слоя и соответствует высокому уровню дифракционных потерь. Если же вещественная компонента диэлектрической проницаемости среднего слоя больше, чем у прилегающих слоев, то в результате полного внутреннего отражения на границах слоев возникает волна поверхностного типа; поток излучения ограничен средним слоем, образующим диэлектрический волновод. Опыт показывает, что обычно имеет место именно такой *волноводный эффект* и связанная с ним малость оптических потерь. Эффект обусловлен градиентом концентрации свободных носителей заряда, температуры и упругих деформаций в активном слое.

Лазеры, состоящие из одного полупроводника, называют *лазерами на гомоструктурах* или *гомолазерами*. Типичный гомолазер имеет форму параллелепипеда с прямоугольным или трапециевидным основанием. Для его получения используется материал *n*-типа, в котором предвари-

---

* В арсениде галлия этот процесс начинается при концентрации дырок более 10^{19} см$^{-3}$.

тельно диффузией примеси p-типа образуется p-переход, параллельный грани [100]. Лучшим способом получения плоскопараллельных зеркал является скол кристалла вдоль кристаллических плоскостей [110] перпендикулярно плоскости p-n-перехода. Общей чертой инжекционных гомолазеров являются очень высокие плотности порогового тока при комнатной температуре, что предопределило невозможность создания непрерывного режима работы при этой температуре.

Толщина d лежащего в p-области активного слоя, в котором происходит усиление излучения, определяется длинной диффузией L_n инжектированных электронов. Концентрация инжектированных электронов в направлении x, перпендикулярном плоскости p-n-перехода, уменьшается с расстоянием пропорционально $\exp(-x/L_n)$. Поэтому накачка и, следовательно, усиление не однородны. Кроме того, часть электронов диффундируют за пределы области усиления излучения и не участвует в нем. В сильно легированном высококачественном арсениде галлия при комнатной температуре $L_n \approx 1...3$ мкм.

Показатель преломления в активном слое выше, чем в соседних слоях. Различие показателей преломления в этих слоях обусловлено различием оптических свойств n- и p-областей, связанным с температурными градиентами, инжекцией носителей тока и усилением излучения. На границе раздела областей с различными показателями преломления происходит полное отражение лучей, электромагнитное излучение распространяется в направлении, параллельном границам раздела между слоями, образующих диэлектрический волновод внутри активного слоя. В результате волноводного эффекта в полупроводниковых инжекционных лазерах происходит также ограничение светового поля. Коэффициент оптического ограничения Г определяет долю электромагнитного излучения, приходящуюся на возбужденную область. В гомолазере разность между показателями преломления центральной области, излучающей свет, и прилегающими пассивными областями мала ($\sim 10^{-3}$), и поэтому оптическое ограничение мод в гомолазере слабое.

Коэффициент усиления в арсениде галлия при комнатной температуре в области значительного усиления $G = \beta(\eta J/d - J_0)$, где J — плотность тока; η — квантовая эффективность; β, J_0 — параметры кристалла. Внутренние потери, характеризуемые показателем поглощения K_g, обычно обусловлены поглощением на свободных носителях и рассеянием на оптических неоднородностях. Составляющая потерь за счет излучения через зеркала с коэффициентом отражения R равна $l^{-1}\ln R$, где l — длина резонатора. На пороге усиление должно равняться сумме всех потерь, т. е. $\Gamma G = K_g - l^{-1}\ln R$. Плотность тока на пороге генерации

$$J_{\text{пор}} = J_0 d/\eta + (d/\beta\eta\Gamma)(K_g - l^{-1}\ln R). \qquad (11.77)$$

В гомолазерах скачки показателя преломления на краях области усиления малы, что приводит к утечке световой энергии в окружающие области, в которых потери велики. Это проникновение излучения в пассивные области приводит к дополнительному вкладу в увеличение потерь K_g. Поэтому в выражении (11.77) слабое оптическое ограничение приводит к малым значениям Г и большим K_g, в то время как толщина активной области d, определяемая диффузионной длиной, равна нескольким микрометрам. При этом плотности тока $J_{\text{пор}}$ при комнатной температуре превышают 26 кА/см^2.

В инжекционных лазерах оптический резонатор образован из диэлектрического волновода конечной длины l. Если ширина волновода w бесконечна, то продольные моды резонатора образуют непрерывный спектр собственных частот (2.34). При конечной ширине возникает дискретный спектр углов $\theta_m = \text{arctg}[ml/(wq)]$, где m — поперечный индекс моды (в направлении широкой стенки волновода): q — продольный индекс моды.

В этом случае собственные частоты резонатора находят по формуле $\nu_{mq} = \dfrac{v\,q}{2l}\sqrt{1+\left(\dfrac{ml}{qw}\right)^2}$. При $lm/(wq) \ll 1$ $\nu_{mq} \approx \nu_{0q} + v\,lm^2/(2qw^2)$, где $\nu_{0q} = v\,q/(2l)$.

Малая толщина d диэлектрика препятствует возбуждению поперечных типов колебаний в направлении узкой стенки волновода (d) [в симметричном волноводе при $(2\pi/\lambda_0)d\sqrt{\varepsilon_{r2} - \varepsilon_{r1}} < \pi$].

Дифракционная расходимость выходящего излучения диффузионных лазерных диодов из арсенида галлия в импульсном режиме при 77 К в вертикальной плоскости составляет 10—30°. В горизонтальной плоскости (в плоскости p-n-перехода) для стандартных лазеров расходимость меньше (около 3°) вследствие большей ширины излучающей области (доли миллиметра).

Степень поляризации внешнего излучения зависит главным образом от волноводных характеристик (в обычных резонаторах). Для большинства типов лазеров вблизи порога генерации характерна тенденция к преимущественному возбуждению волн типа Н.

Схема преобразования энергии накачки в полупроводниковом лазере представлена на рис. 11.19. В стационарном режиме отношение полезной мощности $\mathcal{P} = \hbar\omega N_e$, связанной с N_e когерентно излучаемыми во внешнюю среду фотонами (в единицу времени), к мощности накачки

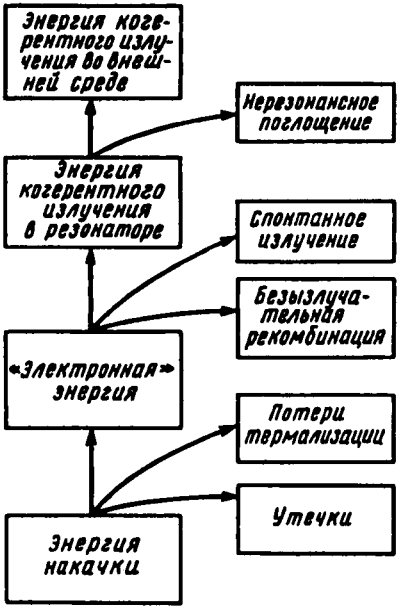


Рис. 11.19. Схема преобразования энергии накачки в полупроводниковом квантовом генераторе

$P_н = IU = I(U_{pn} + RI)$ определяет *коэффициент полезного действия лазера*

$$\eta = \frac{\hbar\omega N_e}{IU} = \frac{\hbar\omega}{eU_{pn}} \frac{1}{1 + IR/U_{pn}} \frac{N_e}{I/e} \approx$$

$$\approx \frac{U_{pn}}{U_{pn} + IR} \frac{N_e}{I/e} =$$

$$= \frac{U_{pn}}{U_{pn} + IR} \frac{r_{st}}{r_{st} + r_{sp}} \frac{N_e}{N_i} =$$

$$= \frac{U_{pn}}{U_{pn} + IR} \eta_e = \frac{U_{pn}}{U_{pn} + IR} \eta_p \eta_i,$$

где $eU_{pn} \approx \hbar\omega$; U_{pn} — напряжение на p-n-переходе; R — объемное сопротивление p- и n-областей диода; N_i — число фотонов, возникающих в p-n-переходе; r_{st}, r_{sp} — вероятности вынужденного излучения и спонтанной рекомбинации (включая безызлучательную); $\eta_i \approx r_{st}/(r_{st} + r_{sp})$ — *внутренний квантовый выход. Внешний квантовый выход*

$$\eta_e = \frac{N_e}{I/e} = \frac{r_{st}}{r_{st} + r_{sp}} \frac{N_e}{N_i} = \eta_i \eta_p$$

определяется характером рекомбинации и потерями на поглощение и рассеяние в активной области. Отношение выходной мощности излучения к мощности излучения, генерируемой в кристалле, называют *функцией выхода* $\eta_p = N_e/N_i$. Она зависит от поглощения и рассеяния излучения в кристалле, размеров резонатора и коэффициента отражения зеркал; ее значение может приближаться к единице.

С увеличением длины кристалла внешний квантовый выход уменьшается (рис. 11.20), так как возрастает доля нерезонансных потерь. В материалах с прямой зонной структурой (гомоструктурой) значительное уменьшение внеш-

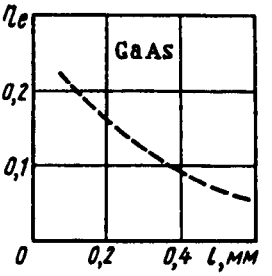


Рис. 11.20. Зависимость внутреннего квантового выхода от длины резонатора при $T = 300$ K

него квантового выхода происходит из-за собственного поглощения, достигающего значительных величин (10^3 $см^{-1}$ в арсениде галлия, легированном цинком).

Внутренний квантовый выход учитывает потери при переходе «электронной» энергии (энергия электронно-дырочных пар) в энергию когерентного излучения, он характеризует предельно достижимую эффективность преобразования электрической энергии в световую. Выбором материала хорошего качества безызлучательную рекомбинацию можно свести к пренебрежимо малому значению. Однако потери, связанные со спонтанным излучением, не могут быть меньше уровня, соответствующего пороговому значению вынужденного излучения. Внутренний квантовый выход в большой степени зависит от совершенства структуры и чистоты материала в активной области. Неконтролируемые примеси и дефекты образуют глубокие центры, рекомбинация через которые создает длинноволновое излучение или является безызлучательной. В расход электронной энергии на тепловые потери и спонтанное излучение вносят также вклад те участки полупроводникового кристалла, где когерентное излучение отсутствует или ниже порогового уровня. Уменьшение отношения толщины излучающего слоя к толщине инвертированного слоя снижает полезное использование энергии накачки, что равносильно уменьшению внутреннего квантового выхода. Его уменьшение с ростом температуры (рис. 11.21) является одной из причин значительного роста порогового тока с повышением температуры. С ростом температуры также увеличиваются вероятность безызлучательных переходов и нерезонансные потери.

Одним из методов уменьшения потерь и ограничения толщины активной области до оптимальной является использование гетероструктур.

Гетеропереход представляет собой переход между двумя различными полупроводниками, образованными в одном кристалле. Основное различие этих полупроводников в *гетеролазерах* заключается в разных значениях ширины запрещенной зоны и показателя преломления. При этом снижение порогового тока (11.77) происходит за счет уменьшения d и увеличения Г. Это достигается помещением активного слоя GaAs между более широкозонными слоями $Al_xGa_{x-1}As$, в результате чего одновременно происходит ограничение локализации инжектированных носителей тока и оптического поля.

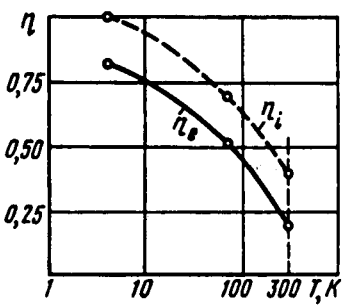


Рис. 11.21. Зависимость квантового выхода от температуры

Первыми инжекционными лазерами, которые стали работать в непрерывном режиме при комнатной температу-

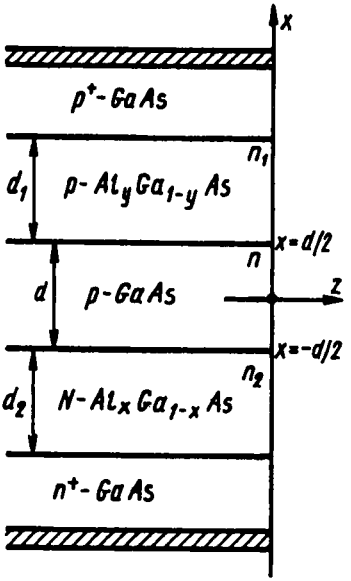


Рис. 11.22. Структура лазера с двусторонней гетероструктурой

ре, были лазеры с двусторонней гетероструктурой (ДГС). Обычное расположение слоев в ДГС-лазере показано на рис. 11.22. Молярная доля Al обозначается через x или y (в ДГС-лазерах $x = y$). Верхний слой p^+-GaAs обеспечивает создание омических контактов. В правильно выполненном лазере толщины слоев d_1 и d_2 достаточно велики, чтобы исключить влияние границ с контактным слоем p^+-GaAs или n^+-подложкой на распределение оптического поля.

На границе раздела между p-GaAs с узкой запрещенной зоной и более широкозонным $Al_xGa_{1-x}As$ происходит резкий скачок показателя преломления (не менее 10^{-2}). Центральный активный слой p-GaAs имеет более высокий показатель преломления, образуя более сильный симметричный диэлектрический волновод, чем в гомолазере (скачки показателя преломления на p-n-переходе и на границе GaAs-$Al_xGa_{1-x}As$ различаются по меньшей мере в 10 раз). Толщина активного слоя d может быть уменьшена более, чем на порядок (от нескольких микрометров в гомолазерах до 0,1 мкм в ДГС-лазерах). Потери излучения снижаются не только благодаря увеличению коэффициента оптического ограничения Г, но и из-за того, что часть излучения, проникающая за пределы активного слоя, находится в более широкозонном полупроводнике и не поглощается (как это происходит в гомолазере). Использование ДГС позволило понизить пороговые плотности тока при комнатной температуре до значений порядка 10^3 А/см^2.

ДГС-лазер на основе GaAs-$Al_xGa_{1-x}As$ при комнатной температуре позволяет получить излучение с длиной волны 900 нм. При этом показатели преломления $n = 3{,}59$ и $n_1 = n_2 = 3{,}385$. При решении дисперсионного уравнения для волн типа H_m в симметричном плоском волноводе найдено, что для $d < 0{,}6$ мкм может существовать только мода низшего порядка H_0 (волна основного типа), для которой не существует условия отсечки. При $0{,}6 < d < 1{,}5$ мкм могут возникать волны типа H_2, при $d > 1{,}5$ мкм — волны типа H_4 и т. д. Из дисперсионного уравнения для нечетных типов волн следует, что волна типа H_1 может возникнуть при $d > 0{,}3$ мкм, волна типа H_3 — при $d > 1{,}1$ мкм.

Как показывают расчеты и эксперимент в ДГС-лазерах поляризация излучения соответствует волнам типа Н. Коэффициент оптического ограничения растет с увеличением толщины активного слоя. Для высших типов волн он мал в режиме отсечки и при увеличении приближается к своему значению для основного типа колебаний.

Поле излучения в дальней зоне и расходимость луча инжекционного лазера могут быть получены преобразованием Фурье выражения для распределения поля волновода на поверхности зеркала оптического резонатора. Поле в дальней зоне лазера определяет способ ввода излучения в оптическое волокно, его эффективность, апертуру линзы для формирования и считывания голограмм и характеризует его распределение внутри лазерного волновода.

Пороговое значение тока (11.77) зависит не только от коэффициента оптического ограничения моды, но и от коэффициента отражения R_m на торцевых гранях этой моды, поскольку он входит в пороговое условие в виде величины $l^{-1}\ln R_m$. По мере роста толщины активного слоя коэффициент оптического ограничения увеличивается (внутри слоя локализуется большая часть моды), однако коэффициент R_0 сначала растет, а потом уменьшается при $d > 0{,}3$ мкм. В ДГС-лазерах для получения непрерывного режима работы при комнатной температуре толщину активного слоя выбирают в пределах 0,1...0,3 мкм, а толщину прилегающих слоев — 1,0...3,0 мкм. Отбор преимущественного типа волн и поляризации излучения обусловлен именно различием коэффициентов отражения на торцевых гранях для Н- и Е-волн. Поэтому при выборе толщины активного слоя с учетом возбуждения волн высших типов учитывают не только изменение коэффициента оптического ограничения, но и изменение коэффициента отражения. Более высокие значения коэффициентов отражения для волн типа Н по сравнению с их значением для волн типа Е определяют выбор этой поляризации для ДГС-лазеров.

Ширина запрещенной зоны арсенида галлия высокой чистоты для прямых переходов составляет 1,519 эВ при абсолютном нуле и 1,424 эВ при комнатной температуре. Запрещенная зона сужается при увеличении концентрации примесей. Это сужение составляет примерно 0,013 эВ при концентрации дырок 10^{18} см$^{-3}$. Энергетический зазор в $Al_xGa_{1-x}As$ $\mathscr{E}_g = 1{,}424 + 1{,}247x$ (эВ).

В ДГС-лазерах, используемых в системах волоконно-оптической связи $x = 0{,}08$, а длина волны излучения в активном слое, при которой в оптическом волокне потери имеют минимум, равна 820 нм.

Энергия уровня Ферми является свободной энергией электронов и дырок или химическим потенциалом. При тепловом равновесии уровень Ферми должен быть постоянным на протяжении всего гетероперехода (следует из второго закона термодинамики). Поэтому изгиб зон на грани-

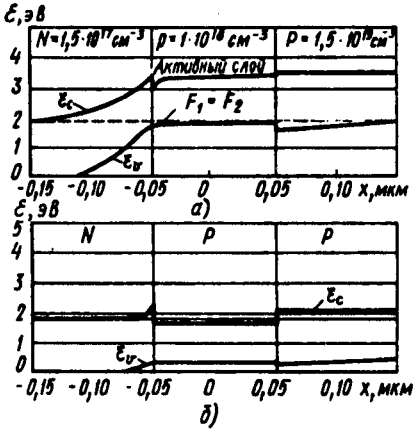


Рис. 11.23. Зонная диаграмма двусторонней гетероструктуры

це раздела определяется разностью уровней Ферми полупроводников, образующих переход.

Энергетическая зонная диаграмма GaAs-$Al_{0,3}Ga_{0,7}As$ двусторонней гетероструктуры *N-p-P*-типа приведена на рис. 11.23, *а* при нулевом смещении (контактная разность потенциалов 1,643 В), а на рис. 11.23, *б* — при прямом смещении (контактная разность потенциалов 1,43 В). Разрыв зоны проводимости $\Delta\mathscr{E}_c$ создает потенциальный барьер для электронов на *p-P*-гетеропереходе, тем самым ограничивая инжектированные электроны активным слоем арсенида галлия. Разрыв валентной зоны $\Delta\mathscr{E}_v$ вместе с оставшейся после приложения прямого смещения контактной разностью потенциалов создает потенциальный барьер для дырок на *p-N*-гетеропереходе, препятствуя инжекции дырок в *N*-слой. Эти барьеры приводят к ограничению как неосновных, так и основных носителей заряда в двойной гетероструктуре.

В арсениде галлия усиление возникает при концентрациях инжектированных носителей заряда около 10^{18} $см^{-3}$, а генерация в полосковой структуре — при концентрации $2 \cdot 10^{18}$ $см^{-3}$.

Ограничение электронов и дырок не является полным, поскольку имеется тепловой поток небольшой части электронов с энергиями, превышающими энергию потенциального барьера на гетеропереходе. Это та часть электронов, инжектированных в активный *p*-слой ДГС-структуры *N-p-P*-типа, которая проходит через *p-P*-гетеропереход и становится неосновным носителем заряда в *P*-слое. В дальнейшем эти носители рекомбинируют с дырками по мере диффузии по направлению к положительному контакту.

Ранее рассматривались свойства лазеров с широким контактом. Однако большая часть лазеров, используемых вне лаборатории, имеет полосковую геометрию, где область растекания тока ограничена и в направлении, параллельном плоскости *p-N*-перехода. Конструкция полосковой структуры, в которой металлическая полоска вытравливается в слое SiO_2, показана на рис. 11.24. Ширина полосок обычно составляет 5—30 мкм. Ограничение ширины в боковом направлении преследует несколько целей. Во-первых, уменьшается площадь поперечного сечения токовой области, а, следовательно, и рабочий ток. Во-вторых, при ши-

рине полоски меньше 15 мкм осуществляется генерация в одном световом канале с основным поперечным (в направлении плоскости p-N-перехода) типом волны. В-третьих, вследствие удаленности активной области от поверхности улучшаются деградационные характеристики.

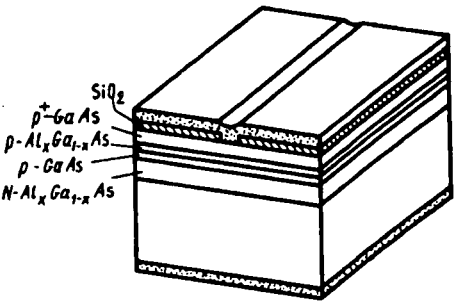


Рис. 11.24. Полосковый гетеролазер

Вместе с тем из-за меньшей площади сечения получаемые выходные мощности полосковых лазеров меньше, чем у лазеров с широким контактом. При ширине полоски 15 мкм и длине резонатора 250 мкм световая мощность в непрерывном режиме при рабочем токе 200 мА от одного зеркала обычно составляет 20—30 мВт. Полный угол расходимости светового пучка на уровне половинной интенсивности обычно равен 45° в направлении, перпендикулярном плоскости p-n-перехода, и 9° в направлении вдоль этой плоскости.

Полосковые ДГС-лазеры являются удобным источником излучения для систем волоконно-оптической связи из-за малого размера, малых токов (200 мА) и напряжений (1,7 В), коэффициента полезного действия свыше 5%, большого срока службы и малой стоимости. Привлекательна возможность прямой модуляции излучения изменением тока накачки.

Полосковые ДГС-лазеры с импульсной модуляцией позволяют осуществить световодную связь со скоростями от 50 до 400 Мбит/с

При температурах жидкого гелия и азота удается получить КПД инжекционных лазеров до 50 %, а при комнатной температуре — несколько процентов. Несмотря на имеющиеся потери, КПД этих лазеров выше КПД лазеров на других активных средах, что обусловлено преобразованием в полупроводниковом кристалле квантового генератора энергии электрического тока непосредственно в энергию когерентного излучения [энергетический выход $\hbar\omega/(eU) \approx 1$].

Рабочие режимы полупроводниковых лазеров могут соответствовать высоким плотностям мощности накачки (10^8 Вт/см^3) и оптического потока (10^7 Вт/см^2), не свойственным другим полупроводниковым приборам. В этих условиях возникают деградационные явления, проявляющиеся в монотонном уменьшении выходной мощности без внешних признаков разрушения кристалла. Обычно срок службы оценивается временем, в течение которого выходная мощность уменьшается до половины своего первоначального значения. Лишь у отдельных лабораторных образцов он достигает 10^4 ч. Старение замедляется и срок службы резко увеличивает-

ся при снижении тока инжекции. Уменьшение его в два раза увеличивает срок службы некоторых диодов более чем в пять раз.

Эффективность накачки. Отношение электронной энергии к затраченной на ее создание энергии накачки называют *эффективностью накачки* $\eta_\gamma = GV/(I/e)$, где G — скорость инжекции электронно-дырочных пар; V — объем активной области.

Скорость инжекции может быть оценена из предположения, что ток инжекции определяет диффузионный ток, изменяющийся по экспоненциальному закону по мере удаления (x) от p-n-перехода. В p-области плотность электронного тока $J_n(x) = J_n(0)\mathrm{e}^{-x/L}$, а плотность полного тока $J(x) = J_n(x) + J_p(x) = \mathrm{const} = J$.

При сильном легировании n-области и слабом легировании p-области, а также при $m_p \gg m_n$ (например, в арсениде галлия) с достаточной точностью коэффициент инжекции $\gamma = J_n(0)/J \approx 1$. Поэтому для тонкого активного слоя, где $x < d \ll L$,

$$I_n(x) \approx J\,\mathrm{e}^{-x/L} \approx J(1 - x/L).$$

По уравнению непрерывности скорость инжекции

$$G = \left(\frac{\partial \Delta n}{\partial t}\right)_{\text{инж}} = -\frac{1}{e}\operatorname{div}\mathbf{J}_n = -\frac{1}{e}\frac{\mathrm{d}J}{\mathrm{d}x} \approx \frac{J}{eL} \tag{11.78}$$

и эффективность инжекции $\eta_p = GV/(J/e) \approx d/L$.

Ватт-амперная характеристика. В кинетическом уравнении (11.73) для одномодового приближения правая часть выражает увеличение плотности фотонов (в единице объема за единицу времени) за счет вынужденного излучения. Это увеличение пропорционально плотности фотонов n_ν, а коэффициент пропорциональности $r_{st} = N_{21}/(\tau_c N_0)$ может быть назван также частотой или скоростью вынужденного излучения. С учетом спонтанного излучения уравнение (11.73) приобретает вид

$$\frac{\partial n_\nu}{\partial t} + \frac{n_\nu}{\tau_c} = r_{sp} + r_{st} n_\nu, \tag{11.79}$$

где r_{sp} — скорость спонтанного излучения.

Аналогичным образом можно записать кинетическое уравнение для неравновесных электронов:

$$\frac{\partial \Delta n}{\partial t} + \frac{\Delta n}{\tau_e} = -(r_{sp} + r_{st} n_\nu) + G, \tag{11.80}$$

где τ_e — время жизни неравновесных электронов в отсутствие излучения и инжекции.

Наблюдения показывают, что зависимость скорости вынужденного излучения и коэффициента оптического усиления (в отсутствие насыщения) от концентрации электронов описывает степенная функция

$$r_{st} = a\Delta n^m, \tag{11.81}$$

где a — коэффициент пропорциональности; m = 1...3 в зависимости от условий и типа лазера.

Если выражение (11.81) подставить в уравнения (11.79), (11.80) и пренебречь спонтанными переходами ($r_{sp} = 0$), то из совместного решения этих уравнений получим для стационарных значений $\Delta n_0 = (2a\tau_c)^{-1/m}$ и $n_{v0} = 2\tau_c(G - \Delta n_0/\tau_e)$.

Величина $\Delta n_0/\tau_e = G_T$ представляет собой часть скорости инжекции, которая компенсирует безызлучательные потери электронов и может быть названа *пороговой скоростью инжекции* ($n_{v0} = 0$ при $G = G_T$). Из ее выражения следует, что в рассмотренном приближении соответствующее значение пороговой концентрации электронов совпадает со стационарным значением концентрации электронов. Таким образом, концентрация электронов в режиме за порогом генерации стабилизируется на пороговом значении и вся сверхпороговая накачка идет на увеличение плотности фотонов:

$$n_{v0} = 2\tau_c(G - G_T). \tag{11.82}$$

Интенсивность электромагнитного излучения во внешнюю среду пропорциональна плотности фотона в резонаторе. Поэтому формулы (11.78) и (11.82) могут быть использованы для получения выходной характеристики инжекционного лазера, т. е. зависимости мощности его излучения

$$P = \beta'\tau_c(G - G_T) = \beta(J - J_T) = \eta_{WA}(I - I_T) \tag{11.83}$$

от тока накачки I (I_T — пороговый ток инжекции; η_{WA} — ватт-амперная дифференциальная эффективность; β, β' — коэффициенты пропорциональности).

Линейность выходной характеристики (11.83) у инжекционных лазеров вблизи порога обычно наблюдается на опыте даже в нестационарных условиях, например в режиме пичковой генерации и при импульсной инжекции (для средней мощности). Отклонения от простой линейной формы при больших токах (рис. 11.25) возникают в результате нагрева активной области, измене-

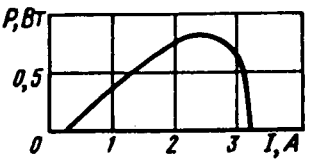


Рис. 11.25. Ватт-амперная характеристика инжекционного лазера в стационарном режиме при температуре 4,2 К

ния ее объема, вариации модового состава излучения и деградации (необратимого ухудшения параметров).

При малых токах в предпороговом режиме излучение обусловлено в основном спонтанными переходами, что приводит к падению усиления, росту нерезонансного поглощения, ослаблению электронного и оптического ограничений. Наличие спонтанной составляющей в общем световом потоке приводит к тому, что стационарные пороговые условия оказываются выполненными при усилении, несколько меньшем, чем потери, а не при точном равенстве этих величин.

Мощность излучения инжекционных лазеров в непрерывном режиме при глубоком охлаждении может доходить до 12 Вт, а при комнатной температуре — до 0,2 Вт. В диффузионных переходах пороговая плотность тока колеблется от 30 А/см^2 (при 4,2 К) до 100 кА/см^2 (при 300 К). Наиболее низкие значения пороговой плотности тока при комнатной температуре получены в двусторонних гетероструктурах (0,85 кА/см^2). Использование многослойных гетероструктур позволяет осуществить еще более гибкую оптимизацию оптических и электрических характеристик. В пятислойной гетероструктуре пороговая плотность тока при комнатной температуре доведена до 0,39 кА/см^2.

Мощности излучения в импульсном режиме. Рост тока накачки вызывает, с одной стороны, увеличение концентрации инжектированных носителей заряда, а с другой — нагрев излучающего кристалла. Эти два процесса оказывают противоположное влияние на мощность излучения, а их соотношение определяет как значение этой мощности, так и границы токов накачки, при которых возможна генерация излучения.

Изложенное следует из выражений для мощности излучения (11.83) и порогового тока накачки

$$I_T = A + BT^n, \tag{11.84}$$

где A, B — коэффициенты, зависящие от активной среды; T — температура диода; n — параметр, зависящий от температуры.

В арсениде галлия при температурах выше 60 К ток (11.84) пропорционален третьей степени температуры ($n = 3$). При более низких температурах пороговое значение тока мало и постоянно ($n = 0$). При температурах, близких к 0 К, $n \approx 1$.

К одному из способов снижения температуры кристалла относится установление импульсного режима накачки, когда ток инжекции длится лишь небольшую часть t_1 периода повторения импульсов t_2. Превышение температуры ΔT p-n-перехода температуры окружающей среды можно рассчитать с помощью упрощенной эквивалентной схемы тепловой цепи (рис. 11.26), где R_T — тепловое сопротивление; C_T — тепловая емкость; P_T — рассеиваемая в образце мощность. Согласно этой схеме,

$$C_T \frac{\partial \Delta T}{\partial t} + \frac{\Delta T}{R_T} = P_T.$$

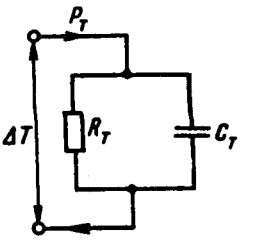


Рис. 11.26. Эквивалентная схема тепловой цепи

Решение данного уравнения с учетом начальных условий позволяет получить выражение для прироста температуры p-n-перехода после воздействия n одинаковых прямоугольных импульсов:

$$\Delta T = R_T P_T (1 - e^{-t_1/\tau})[1 + e^{-t_2/\tau} + e^{-2t_2/\tau} + \ldots + e^{-(n-1)t_2/\tau}],$$

где $\tau = R_T C_T$ — тепловая постоянная времени.

Для бесконечной последовательности импульсов

$$\Delta T = R_T P_T \frac{1 - e^{-t_1/\tau}}{1 - e^{-t_2/\tau}} = \frac{R_T P_T}{Q}, \qquad (11.85)$$

где $Q = \dfrac{1 - e^{-t_2/\tau}}{1 - e^{-t_1/\tau}} \approx \dfrac{t_2}{t_1}$ практически равно скважности (t_2/t_1), если $t_1 < t_2 \ll \tau$.

Учитывая, что температура окружающей среды T_0 связана с пороговым током соотношением

$$I_0 = A + BT_0^n \qquad (11.86)$$

и перегрев

$$\Delta T = T - T_0, \qquad (11.87)$$

из уравнений (11.84)—(11.87) получим

$$\frac{I_T - A}{I_0 - A} = \left(\frac{T}{T_0}\right)^n = \left(1 + \frac{R_T P_T}{Q T_0}\right)^n. \qquad (11.88)$$

Мощность, рассеиваемая в образце,

$$P_T = rI^2 + (1 - \eta_c) I U_{pn} = P_н - P, \qquad (11.89)$$

где r — электрическое сопротивление контактов; η_c — коэффициент преобразования электрической энергии в световую; U_{pn} — напряжение на p-n-переходе; $P_н$ — мощность накачки; P — мощность излучения.

Если в порядке первого приближения положить $A = 0$, то из уравнений (11.88) и (11.89) следует

$$\frac{I_T}{I_0} = \left[1 + \frac{H}{Q}\left(\frac{I}{I_0}\right)^2 + U_{pn}\frac{H}{Q}\frac{1 - \eta_c}{rI_0}\frac{I}{I_0}\right]^n, \qquad (11.90а)$$

где $H = rI_0^2 R_T / T_0$.

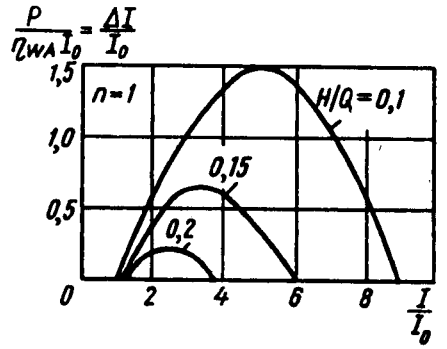


Рис. 11.27. Зависимость мощности излучения от тока инжекции при различных скважностях

Из уравнения (11.90а) следует два вывода. Во-первых, пороговый ток накачки зависит не столько от абсолютных значений длительности и частоты импульсов накачки, сколько от их отношения, т. е. скважности. Во-вторых, пороговый ток является функцией тока накачки. Поэтому при определенном нагреве образца достижение порога генерации оказывается невозможным (при $I_T > I$).

Полученные выражения позволяют также выбрать оптимальный режим работы лазерного диода. Зависимость P от эффективного тока накачки $\Delta I = I - I_T$ наглядно прослеживается в случае идеализированного диода ($\eta_c = 1$) и гелиевых температур ($n = 1$), когда

$$\frac{\Delta I}{I_0} = \frac{I}{I_0} - 1 - \frac{H}{Q}\left(\frac{I}{I_0}\right)^2. \qquad (11.90б)$$

Из уравнения (11.90б) и соответствующего ему графика (рис. 11.27) видно, что ΔI равно нулю в двух точках, определяющих граничные значения тех токов накачки, которые могут вызвать излучение. И верхняя граница тока накачки, и максимальное превышение зависят от Q/H. Они тем больше, чем больше скважность.

При снижении скважности ниже некоторого критического значения генерация прекращается: $(\Delta I)_{max} = 0$. При $n = 1$ для идеализированного образца $(H/Q)_{кр} = 0{,}25$ и при $n = 3$ $(H/Q)_{кр} = 0{,}07$.

Связь между мощностью излучения и током накачки можно найти путем подстановки в уравнение (11.83) выражения (11.90а) с заменой произведения $\eta_c I U_{pn}$ на P. Тогда

$$P = \eta_{WA} I_0 \left\{ \frac{I}{I_0} - \left[1 + \frac{H}{Q}\left(\frac{I}{I_0}\right)^2 + q\frac{H}{Q}\frac{I}{I_0} - \frac{H}{Q}\frac{P}{rI_0^2} \right]^n \right\},$$

где $q = U_{pn}/(rI_0)$ — *качество лазерного диода.*

При охлаждении до гелиевых температур ($n = 1$) и $\eta_c = 1$ зависимость $P/(\eta_{WA} I_0)$ от тока I совпадает с аналогичной зависимостью, определяемой уравнением (11.90б) и изображенной на рис. 11.27. Из расчета и

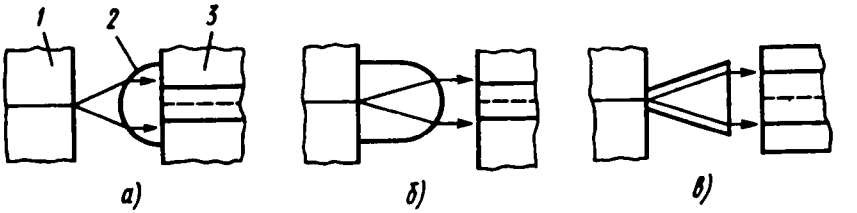


Рис. 11.28. Сочленения лазера с оптическим волноводом в виде микролинзы на торцах волновода (*а*), лазера (*б*) и фокона (*в*):

1 — лазер; *2* — микролинза; *3* — волновод

графика рис. 11.27 следует существование максимума излучаемой мощности и предельного тока накачки, выше которого нагрев делает генерацию невозможной (происходит срыв генерации). В импульсном режиме мощность гомолазеров на арсениде галлия при $T = 77$ К достигает 100 Вт.

Инжекционные лазеры в интегральной оптике. В интегральной оптике используют две категории источников когерентного излучения: миниатюрные лазеры на монокристаллах и пленочные лазеры.

При использовании лазеров на монокристаллах возникает задача эффективного *сочленения* инжекционного лазера с оптоэлектронной системой и, в частности, с оптическим волноводом. Для этой цели могут быть применены микролинзы и фоконы (рис. 11.28), хотя возможно и непосредственное сочленение, если расхождение лазерного пучка меньше угловой апертуры волокна (при оптимальном воздушном зазоре была получена эффективность 8%).

Принцип направленного ответвителя использован в структуре рис. 11.29, *а* для связи активного *1* и пассивного *2* волноводов. Предполагается равенство

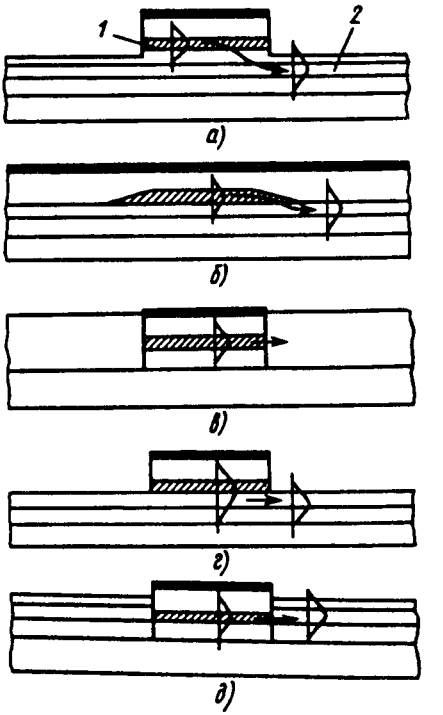


Рис. 11.29. Соединения активного и пассивного волноводов в лазерах на монокристаллах

фазовых скоростей мод в обоих волноводах и туннелирование светового излучения через разделяющий слой с меньшим, чем в этих волноводах, показателем преломления. Эффективность связи через сужающийся волновод (рис. 11.29, *б*) может быть доведена почти до 100% выбором достаточной длины суженного участка (много больше поперечных размеров). Свет подается на пассивный волновод в процессе многократного внутреннего отражения. Связь встык (рис. 11.29, *в*) характеризуется малой эффективностью. В лазерах с большим оптическим резонатором (рис. 11.29, *г*) связь осуществляется за счет «исчезающих полей». Здесь предполагается небольшое различие показателей преломления активной области и волноводного слоя. Эффективность связи достигает 98% при использовании встроенного торцевого волновода (рис. 11.29, *д*) за счет согласования эффективных показателей преломления и профилей полей в обоих волноводах.

В пленочных лазерах на основе GaAlAs, совместимых с планарной технологией гетероструктур интегральной оптики, вместо зеркальных граней применяют распределенную структуру, обусловленную периодическими изменениями лазерных или волноводных свойств.

Периодическая структура (решетка) может быть образована пространственным изменением одного из параметров, от которого зависят условия распространения света: толщины кристалла, показателя преломления или коэффициента усиления. При дифракции световой волны на периодических неоднородностях возникает ряд дифракционных максимумов и проходящая волна получает добавочную энергию от противоположно движущихся волн. Это вместе с усилением создает механизм обратной связи, распределенной вдоль всей периодической структуры и обеспечивающей условия для генерации когерентного излучения. Следовательно, в таком резонаторе решетка играет роль зеркал классических лазеров, т. е. является одновременно отражателем и фильтром: подавляет в спектре излучения волны, которые не должны усиливаться, и увеличивает путь генерируемых волн. Обе функции не локализованы на концах резонатора, как в обычных лазерах, а распределены по всей его длине, поэтому такие лазеры называют *лазерами с распределенной структурой*.

Расстояние d между неоднородностями (период решетки) связано с длиной волны соотношением Брэгга $n\lambda = 2d\sin\theta$, где n — порядок отражения; θ — угол, соответствующий дифракционному максимуму. Высокая спектральная селекция излучения обеспечивается чувствительностью дифракции Брэгга к длине волны. Значительно упрощает согласование лазеров с распределенной структурой с оптическими интегральными схемами возможность вывода излучения перпендикулярно p-n-переходу ($\theta = 90°$). При этом расходимость излучения на порядок меньше, чем при выводе излучения из торцов активного слоя.

Лазеры с распределенной решеткой на границе активной области позволяют получить более низкий порог и лучшую избирательность мод, чем лазеры с распределенной структурой в объеме активной среды. На рис. 11.30, *а*, *б* приведены структуры пленочных лазеров с распределенной структурой в виде гофрированной поверхности активного слоя (*а*) и широкозонного эмиттера (*б*). На рис. 11.30, *в* показана структура дифракционной решетки на границе широкозонного эмиттера и активного слоя двойной гетероструктуры. Расходимость излучения составляет 30′, а ширина спектральной линии — 0,03 нм.

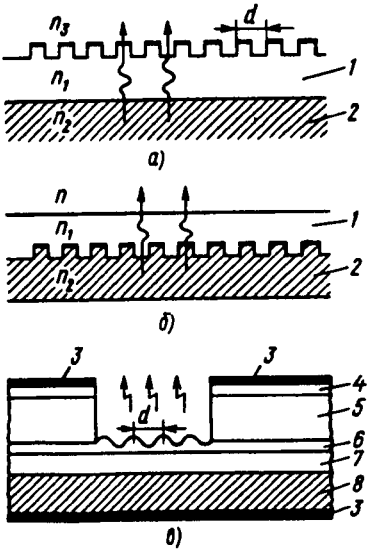


Рис. 11.30. Структура пленочных лазеров с гофрированной поверхностью активного слоя:

1 — активный слой; *2* — подложка; *3* — омический контакт; *4* — слой p^+-CaAs; *5* — слой p-Al_xGaAs; *6* — слой N-GaAs; *7* — слой N-$Al_xGa_{1-x}As$; *8* — подложка n^+-GaAs

Примером монолитной интегрально-оптической схемы, объединяющей на одной подложке лазер, электрооптический модулятор и детектор, может служить структура, представленная на рис. 11.31. Путь светового потока обозначен стрелками, электроды заштрихованы. Область модуляции света получена селективным вытравливанием верхних структур. Ее связь с фотоприемником и лазерной гетероструктурой (с прямым смещением) осуществляется через суживающийся край GaAs волновода. Свет в слое $Ga_{0,7}Al_{0,3}As$ будет распространяться, когда его показатель преломления выше, чем в изолирующем слое $Ga_{0,5}Al_{0,5}As$. В волноводном слое $Ga_{0,7}Al_{0,3}As$ могут быть реализованы все логические функции, модуляция, переключение, связь между схемами, фильтрация, анализ, модовая селекция и т. д. (на схеме не показаны). Фотоприемник на схеме образован поглощающей областью (GaAs) p-n-перехода, где под действием поля обратного смещения возникает фото-

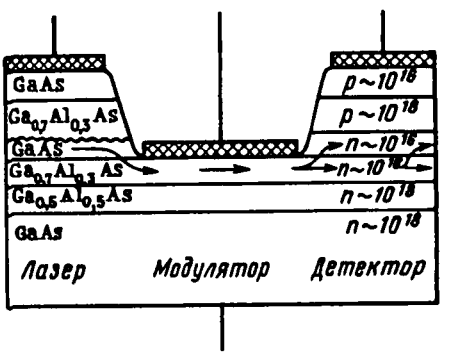


Рис. 11.31. Монолитная интегрально-оптическая схема с пленочным лазером

ток электронно-дырочных пар, образованных при поглощении света.

Интегральная оптика открывает перспективу создания в одной монолитной структуре оптоэлектронных микропроцессоров с оптическим вводом-выводом информации и многоканальных оптических систем связи, на входе которых излучение нескольких одномодовых гетеролазеров, работающих на различных длинах волн, сводится в один оптический волновод, а на выходе установлена система селективных фотоприемников на полупроводниках с градиентом ширины запрещенной зоны.

Гетеролазеры в оптической связи. Широкое применение полупроводниковых лазеров в качестве источников оптических сигналов в волоконно-оптических системах связи обеспечено в значительной мере благодаря разработке эффективных гетероструктур с полосковой геометрией и ультратонкими активными слоями, а также благодаря решению проблемы долговечности лазерных диодов. Малоинерционность лазерного диода при изменении тока накачки позволяет совместить в лазерном диоде функции генератора несущей частоты и модулятора с частотой до 10 ГГц. Поэтому простейший передатчик оптических сигналов может содержать один лазерный диод.

Для повышения длины оптической линии или расстояния между ретрансляторами в системе связи необходимо обеспечить эффективный ввод излучения инжекционного лазера в оптический волоконный световод. Многомодовый волоконный световод для высокоскоростной системы связи имеет обычно угловую апертуру 5—10°. Даже при оптимальном воздушном зазоре и расходимости лазерного пучка меньше угловой апертуры волокна эффективность непосредственного сочленения не превышает 8 %.

В гетеролазерах дифракционная расходимость испускаемого светового пучка больше угловой апертуры оптического волновода. При непосредственной стыковке со световодом эффективность ввода излучения в лучшем случае достигает 30%. Для повышения эффективности ввода применяются микролинзы и фоконы (см. рис. 11.28). Сферическая линза может быть образована путем оплавления многомодовых волокон, обеспечивая эффективность сочленения до 62%. На эффективность ввода излучения влияет и режим накачки, особенно для полосковых гетеролазеров без бокового оптического ограничения.

Лазерные диоды, работающие при предпороговых токах и создающие спонтанное излучение, называют *суперлюминесцентными*. По излучаемой мощности они занимают промежуточное положение между светодиодами и инжекционными лазерами.

Лазерные диоды могут быть использованы также в качестве *оптических усилителей*. Выполнение зеркальных граней под углом Брюстера позволяет осуществить в полосковом лазере усилитель бегущей волны. Лазерные диоды с резонатором используются в регенеративных усилите-

лях. В этих усилителях обнаружен *отрицательный фотоэффект*, обусловленный тем, что под действием внешнего резонансного излучения и вызванного им вынужденного излучения в активной среде происходит уменьшение концентрации активных частиц (избыточных электронно-дырочных пар в полупроводнике). При малых токах накачки (ниже порога инверсии) диод функционирует как обычный фотодиод, в котором фото-ЭДС обусловлена контактными явлениями на *p-n*-переходе. Вблизи порога инверсии, когда в объеме активного слоя уменьшение концентрации неравновесных носителей начинает превалировать над разделением зарядов на контакте, фото-ЭДС изменяет знак и при дальнейшем возрастании накачки увеличивается по абсолютному значению вплоть до порога генерации. Выше порога генерации режим захвата переходит в режим свободной генерации и отрицательная фото-ЭДС убывает.

Фотоответ оптического усилителя можно оценить, если учесть влияние плотности фотонов в активной среде на длину диффузии. Так, по кинетическому уравнению для концентрации электронов

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{J}{ed} - \frac{n}{\tau} - Mnn_{\nu}$$

в стационарном режиме

$$n_{\text{ст}} = \frac{J}{ed}\frac{\tau}{1 + M\tau n_{\nu}},$$

где τ — время жизни носителей заряда при $n_{\nu} = 0$; d — толщина активной зоны; M — коэффициент пропорциональности для вынужденной рекомбинации. Полученное выражение соответствует введению эффективного времени жизни $\tau_{\text{эф}} = \dfrac{\tau}{1 + M\tau n_{\nu}}$ и эффективной длины диффузии $L = \sqrt{\tau_{\text{эф}} D} = L_0 \Big/ \sqrt{1 + M\tau n_{\nu}}$.

ВАХ инжектирующего контакта можно представить в виде

$$J = \frac{eDn_{\text{э}}}{L} \mathrm{e}^{eV/kT} \operatorname{th}\frac{d}{L},$$

где $n_{\text{э}}$ — концентрация инжектируемых носителей в эмиттере, имеющих достаточную энергию для преодоления потенциального барьера с высотой, равной контактной разности потенциалов в инжектирующем контакте. Подстановка эффективной длины диффузии в это выражение при $d \ll L$ позволяет найти напряжение на инжектирующем контакте:

$$V = \frac{kT}{e}\ln\left(\frac{J}{J_0}\frac{1}{1 + M\tau n_{\nu}}\right) = V_0 - \frac{kT}{e}\ln(1 + M\tau n_{\nu}) = V_0 + \Delta V,$$

где V_0 — напряжение в отсутствии излучения; $\Delta V = -\frac{kT}{e}\ln(1+M\tau n_\nu)$ — изменение напряжения при появлении фотонов с плотностью n_ν. Если $M\tau n_\nu \ll 1$, то $\Delta V \approx -\frac{kT}{e}M\tau n_\nu = -\frac{kT}{e}M\tau\frac{n'}{c}\frac{\Phi}{\hbar\omega} = -\frac{kT}{e}\frac{M\tau n'P}{c\hbar\omega S}$, где $P = \Phi S$ — мощность пучка сечением S на выходе усилителя. Для полоскового AlGaAs-лазера при комнатной температуре $\tau = 10^{-9}$ с; $M = 3\cdot10^{-7}$ см^3/с; $S = 10^{-7}$ см^2; $n' = 5$. При усилении 10 дБ $\Delta V/P = -5$ В/Вт.

Следовательно, лазерный усилитель может совмещать функции оптического усилителя и источника информации, т. е. может генерировать *электрический сигнал*, адекватный проходящему оптическому сигналу (лазерный диод выполняет здесь функции фотоприемника без поглощения). Использование лазерных диодов позволяет создать линии двусторонней передачи информации с идентичными оконечными устройствами. На макете такой линии получена скорость передачи 100 Мбит/с.

Для оптических систем связи с частотным уплотнением представляет интерес возможность квантового усилителя независимо усиливать несколько входных сигналов с различными частотами (в пределах контура усиления), если их суммарная мощность не приводит к режиму насыщения. Вместе с тем режим насыщения усилителя по входному сигналу, полностью подавляя фон, позволяет в некоторых случаях обойтись без спектральной фильтрации на входе.

§ 11.11. Динамические нелинейные явления в инжекционных лазерах

Нелинейные процессы в оптическом волноводе инжекционного лазера определяются, в первую очередь, влиянием свободных носителей заряда на показатель преломления, а также влиянием интенсивности излучения на концентрацию свободных носителей. Зависимость показателя преломления $n' = n'_0 + n_2\Phi$ от интенсивности Φ получила название *нелинейной рефракции*.

Нелинейная рефракция в полупроводниках. Влияние свободных носителей заряда вблизи края собственного поглощения приводит к тому, что нелинейная рефракция в полупроводниках проявляется как «гигантская» нелинейность. Если в диэлектрических средах (с самофокусировкой) коэффициент нелинейности $n_2 > 0$ составлял 10^{-16}—10^{-13} см^2/Вт, то в арсениде галлия он равнялся — $4\cdot10^{-4}$ см^2/Вт и в антимониде индия — $3\cdot10^{-3}$ см^2/Вт, что на десять порядков больше (по абсолютному значению), чем в классических нелинейных материалах. В результате проти-

воположного эффекта, производимого электромагнитным полем на невозбужденную и инвертированную среду, знак коэффициента нелинейности будет в этих двух случаях противоположным.

При внешней подсветке полупроводника нелинейность обусловлена фотоносителями, в лазерах — инжектированными носителями. В первом приближении можно принять линейную зависимость приращения показателя преломления $\delta n' = n_2 \Phi = \sigma n$ от концентрации носителей заряда n (σ — коэффициент пропорциональности). Тогда $n_2 \approx \eta\sigma\tau K/(\hbar\omega)$, где η — относительная доля фотоэлектрического поглощения; K — показатель поглощения; τ — время жизни избыточных носителей заряда.

Динамику изменения концентрации электронов под влиянием внешнего излучения можно приближенно представить, если пренебречь диффузионным членом:

$$\frac{\mathrm{d}n}{\mathrm{d}t} = -\frac{n}{\tau} + G - mn_\nu, \tag{11.91}$$

где τ — время жизни электронов проводимости; G — объемная скорость сторонней накачки электронов. Коэффициент взаимодействия электронов с фотонами в линейном приближении

$$m = \xi(n - n_{\text{пор}}), \tag{11.92}$$

где ξ — коэффициент пропорциональности; $n_{\text{пор}}$ — концентрация носителей заряда на пороге инверсии в отсутствии поглощения и усиления. Стационарная концентрация ($\mathrm{d}n/\mathrm{d}t = 0$)

$$n_0 = \frac{G + \xi n_{\text{пор}} n_\nu}{\tau^{-1} + \xi n_\nu}; \tag{11.93}$$

$$\frac{\mathrm{d}n_0}{\mathrm{d}n_\nu} = -\xi \frac{n_0 - n_{\text{пор}}}{\tau^{-1} + \xi n_\nu}.$$

Поэтому

$$n_2 = \frac{\mathrm{d}n'}{\mathrm{d}n}\frac{\mathrm{d}n}{\mathrm{d}n_\nu}\frac{\mathrm{d}n_\nu}{\mathrm{d}\Phi} = -\xi \frac{\mathrm{d}n'}{\mathrm{d}n}\frac{n'}{c\hbar\omega}\frac{n_0 - n_{\text{пор}}}{\tau^{-1} + \xi n_\nu} = \frac{\xi}{\hbar\omega}\left|\frac{\mathrm{d}n'}{\mathrm{d}n}\right|\frac{n'}{c}\frac{n_0 - n_{\text{пор}}}{\tau^{-1} + \xi n_\nu}, \tag{11.94}$$

так как $\mathrm{d}n'/\mathrm{d}n < 0$; $\Phi = c\hbar\omega n_\nu / n'$.

Из выражения (11.93) следует, что если $G > n_{\text{пор}}/\tau$, то $n_0 > n_{\text{пор}}$, $n_2 > 0$ и нелинейность соответствует самофокусировке. При $G < n_{\text{пор}}/\tau$, $n_2 < 0$, что соответствует «дефокусировке» (в том числе и при $G = 0$). Таким образом, знак нелинейности зависит от уровня накачки и изменяется при переходе через порог инверсии.

Из выражения для скорости изменения плотности фотонов $\frac{\mathrm{d}n_\nu}{\mathrm{d}t} = -\frac{n_\nu}{\tau_\phi} + Kn_\nu$ следует что, в условиях стационарной интенсивности излучения $(\mathrm{d}n_\nu/\mathrm{d}t = 0)$ $m = \tau_\phi^{-1}$. Тогда согласно (11.92), $n = n_{\text{пор}} + 1/\xi\tau_\phi = n_c$, где n_c — концентрация носителей заряда, обеспечивающая поддержание необходимого (в том числе, порогового) уровня усиления в лазере. Уровень усиления не зависит от интенсивности излучения. Это означает, что в лазерном режиме разным стационарным значениям Ф соответствуют одинаковые значения показателя преломления.

В отсутствие накачки (G = 0) и в случае поглощения излучения $n_\nu \ll 1/(\tau\xi)$, падающего извне, согласно (11.94),

$$n_2 \approx -\frac{\tau K}{\hbar\omega}\left|\frac{\mathrm{d}n'}{\mathrm{d}n}\right|,$$

где $K = n'\xi n_{\text{пор}}/c$ — показатель поглощения в состоянии равновесия.

Самофокусировка в инжекционном лазере. Самофокусировочной неустойчивости в инжекционном лазере способствует зависимость показателя преломления и диэлектрической проницаемости от плотности фотонного потока, существенно влияющего на концентрацию носителей заряда. Связь между плотностью фотонов и концентрацией неравновесных электронов можно получить из уравнения (11.80), пренебрегая в первом приближении спонтанными переходами ($r_{sp} \approx 0$) и полагая в (11.81) m = 1. Тогда в стационарном режиме

$$\Delta n = \frac{\Delta n_0}{1 + a\tau_e n_0} \approx \Delta n_0(1 - a\tau_e n_0),$$

где Δn_0 — стационарная концентрация неравновесных электронов, когда плотность фотонов $n_0 = 0$. Тогда, пренебрегая потерями ($n'' = 0$, $\varepsilon'' = 0$), в линейном приближении

$$n' = n'_0 + \frac{\mathrm{d}n'}{\mathrm{d}\Delta n}\Delta n = n_1 + n_2\Phi,$$

где $n_1 = n'_0 + \Delta n_0 \frac{\mathrm{d}n'}{\mathrm{d}\Delta n}$; $n_2 = -\frac{\mathrm{d}n'}{\mathrm{d}\Delta n}\frac{an'_0\tau_e\Delta n_0}{c\hbar\omega}$; $\Phi = c\hbar\omega n_0/n'_0$ — интенсивность фотонного потока. Если исходить из того, что нелинейность диэлектрической проницаемости также определяется интенсивностью или мощностью светового потока, то $\varepsilon' = \varepsilon_1 + \varepsilon_2 E^2$, где коэффициент нелинейности $\varepsilon_2 = cn_2(n'_0)^3\sqrt{\varepsilon_1}$.

Распределение напряженности электрического поля найдем из совместного решения волновых уравнений (6.18) в виде

$$\nabla^2 E = \mu_0 \frac{\partial^2 D}{\partial t^2} = \mu_0 \varepsilon_1 \frac{\partial^2 E}{\partial t^2} + \mu_0 \varepsilon_2 \frac{\partial^2 E^3}{\partial t}. \qquad (11.95)$$

Активная область инжекционного лазера представляет собой тонкий слой, толщина которого (вдоль оси x) контролируется волноводным эффектом. Поэтому сжатие светового потока из-за самофокусировки происходит преимущественно в горизонтальной плоскости (в направлении оси y). Поэтому решение волнового уравнения можно искать в виде $E(y, z, t) = E_m(y)\sin(\omega t - \beta z)$. Тогда согласно уравнению (11.95)

$$\frac{\mathrm{d}^2 E_m}{\mathrm{d}y^2} - (\varepsilon_1 \mu_0 \omega^2 - \beta^2) E_m + \varepsilon_2 \mu_0 \omega^2 E_m^3 \cos^2(\omega t - \beta z) = 0.$$

После усреднения по времени этого выражения получим

$$\frac{\mathrm{d}^2 E_m}{\mathrm{d}y^2} - \varkappa^2 E_m + \frac{\varepsilon_2 \mu_0 \omega^2}{2} E_m^3 = 0,$$

где $\varkappa^2 = \varepsilon_1 \mu_0 \omega^2 - \beta^2$. Решение уравнения $E_m = E_0/\mathrm{ch}\,\varkappa y$ описывает профиль канала самофокусировки, ширина которого $w = 1{,}76/\varkappa^2 = 3{,}52/\omega\sqrt{\varepsilon_2 \mu_0} E_0$, т. е $wE_0 = 3{,}52/\omega\sqrt{\varepsilon_2 \mu_0} E_0 \approx \mathrm{const}$. Следовательно, после того, как электрическое поле в центре лазерного канала шириной w превысит значение E_0, канал начнет сжиматься обратно пропорционально E_0 ($w = \mathrm{const}/E_0$).

При комнатной температуре в арсениде галлия в волноводном канале толщиной 1 мкм и шириной 10 мкм с $n_2 = 10^{-8}$ см^2/Вт эффект самофокусировки возникает при накачке всего на 7% выше порога (при мощности излучения около 4,5 мВт) при перепаде показателя преломления на боковой границе канала около 10^{-3}.

Вещественные и мнимые компоненты диэлектрической проницаемости связаны дисперсионными соотношениями Крамерса—Кронинга. Поэтому изменения в спектре $\varepsilon'' = cn'K/\omega$ влекут соответствующие изменения в спектре $\varepsilon' = n'^2 - n''^2$, а изменения в спектре показателя преломления

$$n'(\omega) - 1 = \frac{2}{\pi} \int_0^\infty \frac{\Delta n''(\omega')}{\omega'^2 - \omega^2} \mathrm{d}\omega'.$$

Изменения концентрации носителей заряда могут привести к *насыщению поглощения* вблизи края собственного поглощения. В арсениде

галлия вблизи длины волны 880 нм при температуре 300 К основной вклад в изменение концентрации вносят межзонные переходы. Приближенная теоретическая оценка показывает, что каждый переход из валентной зоны в зону проводимости дает вклад в показатель преломления

$$\Delta n' \approx \frac{2\pi|\mathbf{p}|^2}{n_0\hbar(\omega-\omega_0)},$$

где $\mathbf{p}$ — эффективный электрический момент перехода; ω_0 — центральная частота перехода. Следовательно, в области $\omega < \omega_0$ показатель преломления уменьшается при увеличении концентрации фотоносителей. С учетом взаимодействия излучения с различными энергетическими уровнями

$$\sigma = \frac{dn'}{dn} \approx -\frac{2\pi|\overline{\mathbf{p}}|^2}{n_0\hbar(\overline{\omega}_0-\omega)},$$

где $\overline{\mathbf{p}}$, $\overline{\omega}_0$ — эффективные средние значения $\mathbf{p}$ и ω_0. Для арсенида галлия в рассмотренных условиях наиболее вероятное значение $\sigma = -6{,}9\cdot10^{-21}$ См^3. Значения этого коэффициента оказываются бо́льшими в узкозонных полупроводниках, т. е. проявления нелинейности такого рода более сильно выражены в длинноволновых лазерах.

К узкозонным полупроводникам относится, в частности, антимонид индия InSb, у которого ширина запрещенной зоны составляет около 0,2 эВ при температуре 77 К. Он непрозрачен в видимой части спектра, но прозрачен для некоторой области инфракрасной части спектра. Энергии 0,2 эВ соответствует длина волны приблизительно 5 мкм, генерируемая лазером на оксиде углерода. Ширина запрещенной зоны уменьшается при повышении температуры. Поэтому заметное изменение («переключение») порядка интерференции и функции Эйри в пленке антимонида индия удалось получить и при комнатной температуре в луче лазера с длиной волны 10 мкм. Благодаря малой ширине запрещенной зоны нелинейность антимонида индия в 100—1000 раз превосходит нелинейность арсенида галлия, так как нелинейность показателя преломления уменьшается пропорционально квадрату ширины запрещенной зоны.

В арсениде галлия энергия переключения может быть доведена до 1 фДж, причем даже в многомодовых инжекционных лазерах мощностью менее 10 мВт. При комнатной температуре в бистабильном устройстве на сверхрешетке (61 пара слоев GaAs толщиной 336 нм и $Al_{0,27}Ga_{0,73}As$ толщиной 40,1 нм с полной толщиной GaAs 2 мкм) переключение происходит при плотности мощности излучения 1 мВт/мкм^2 и времени переключения от 20 до 40 нс.

Бистабильность в лазерах. Одним из проявлений нелинейной рефракции является смещение частоты продольных мод плоского активного

резонатора с увеличением тока накачки, причем знак смещения противоположен знаку смещения при температурном эффекте. При этом в полупроводниковых резонаторах наряду с оптической возникает и электрооптическая бистабильность. Резонансы, перестраиваемые под действием приложенного напряжения вследствие электрооптического эффекта, приводят в нелинейном активном резонаторе к появлению волнообразной ВАХ с падающими участками при внешнем облучении. В резонаторе с p-n-переходом зависимость фототока от интенсивности подсветки имеет петлю гистерезиса, свидетельствующую об оптической бистабильности.

В неполосковых гетеролазерах возможно формирование в активной области пространственных каналов (нитей), вызванное самофокусировкой лазерного потока в поперечном направлении, вдоль волноводного слоя. Возникновение пятнистой структуры излучения сопровождается появлением ступеньки, т. е. нелинейности на ватт-амперной характеристике.

Если инжекционный лазер снабдить внешним отражательным селектором, то реализуется бистабильный режим, обусловленный свойствами составного («трехзеркального») резонатора и, в частности, зависимостью оптической длины нелинейного участка резонатора от интенсивности излучения. При спектральной перестройке внешнего резонатора наблюдается гистерезисный характер зависимости фототока от интенсивности подсветки.

Механизм возникновения гистерезиса связан с изменением добротности составного резонатора при тонкой перестройке внешнего элемента (меняется взаимное положение резонансов селектора и диода). При некоторых фиксированных значениях тока накачки настройка позволяет пройти порог генерации и получить лазерное излучение. При увеличении добротности составного резонатора, увеличении интенсивности излучения и снижения концентрации свободных носителей происходит некоторое увеличение показателя преломления и оптической длины резонатора, т. е. продольный резонанс смещается в сторону более длинных волн. Если настройка внешнего элемента осуществляется от более длинных в короткие волны, то она сопровождается встречным смещением нелинейного резонанса; в противоположном направлении настройка вызывает «убегание» продольного резонанса. Поэтому зависимость интенсивности излучения от настройки спектрально асимметрична. При достижении добротностью составного резонатора порогового значения переход в режим генерации возникает скачкообразно, а асимметрия перестроечной кривой приводит к тому, что включение и выключение лазерного режима происходят при разных длинах волн (кривая приобретает форму петли гистерезиса).

В области бистабильности лазер имеет два уровня мощности, соответствующие спонтанному и лазерному излучениям, т. е. он представляет систему с памятью, переключение которой возможно введением в систе-

му оптического сигнала от другого лазера. В данном случае для переключения не требуются электронные цепи, необходимые в бистабильных оптических схемах с обратной электронной связью. В отличие от нелинейных эталонов Фабри—Перо бистабильность в лазере проявляется при малом уровне мощности (около 1 мВт). Аналогичную природу имеет эффект бистабильности и в резонансном усилителе на основе полупроводникового лазера.

Нелинейная рефракция в полупроводниковых лазерах позволяет создать бистабильные оптические элементы с оптическим и электрическим управлением с высокой чувствительностью, способные работать в непрерывном режиме при комнатной температуре. Меньшие затраты мощности делают их конкурентоспособными с бистабильными лазерными схемами на основе просветляющихся сред (с нелинейным поглощением).

Спектральные эффекты в связанных резонаторах. Если идентичные лазерные диоды с длиной резонаторов L_1 и L_2 расположить параллельно друг другу и осуществить оптическую связь с помощью внешнего сферического зеркала, то в общем спектре возникают биения, определяемые межмодовой дистанцией продольных мод лазеров. Интенсивность суперлюминесценции при этом оказывается промодулированной по длине волны с периодом $\Delta\lambda = \lambda^2/2n^*|L_1 - L_2|$, где n^* — групповой показатель.

Такая спектральная модуляция обусловлена сложением почти периодических функций спектральной зависимости добротности продольных мод двух резонаторов («нониусный» эффект). Если моды суперлюминесценции совпадают по длине волны в обоих резонаторах, то наложение спектральной модуляции на огибающую продольных мод позволяет получить дополнительное сужение суперлюминесценции. Добротности несовпадающих мод уменьшаются, в результате чего эти моды подавляются. Таким образом, введение оптической связи между лазерными диодами является одним из методов спектральной селекции в инжекционных лазерах. Эта связь может быть использована для получения режима одной продольной моды и электрооптической перестройки инжекционных лазеров, в частности, в стыковочных узлах с волоконными световодами и в интегрально-оптических схемах.

Сдвиг резонансов в одной из секций можно изменять изменением накачки, поскольку это меняет концентрацию носителей заряда и показатели преломления. Сравнительно легко позволяет осуществлять автоподстройку электрооптическая модуляция. В этом случае в модулируемой части составного резонатора используется электрооптический материал (например, ниобат лития или ниобат бария-натрия). Перестройка ускоряется «нониусным» эффектом.

Другой вариант связанных резонаторов получил название *C^3-лазеров* (cleaved-coupled cavity lasers — лазеры со сколотыми связанными резона-

торами). Их образует пара лазерных диодов, последовательно расположенных вдоль оси резонатора и изготовленных раскалыванием по спайности одной диодной заготовки. При этом диоды остаются близко расположенными (зазор менее 5 мкм), реализуя сильную оптическую связь между ними. Раздельная накачка дает возможность управлять режимом генерации с помощью двух независимых токов инжекции в обеих секциях лазерного устройства.

В волоконно-оптических линиях связи C^3-лазеры позволяют обеспечить динамический одномодовый режим, т. е. режим одной продольной моды при сверхбыстрой модуляции тока накачки. Электрическая перестройка частоты дает возможность также реализовать спектральное уплотнение каналов связи. Для этого достаточно незначительно изменить постоянную составляющую тока в одной («модуляторной») из секций лазера, поскольку скорость перестройки существенно увеличена «нониусным» эффектом. Изменение тока модуляторной секции C^3-лазера приводит в некоторых случаях к спектральной бистабильности, применяемой для осуществления логических операций. Переключение лазерной длины волны за время 1 нс получено с помощью ультракоротких электрических импульсов (0,2 нс) с энергией 10^{-12} Дж, наложенных на ток постоянного смещения. При этом возникала петля гистерезиса и в ВАХ.

Сильная спектральная селективность составных (многозеркальных) резонаторов позволяет обеспечить динамический одночастотный режим, который оказывается недоступным обычным лазерным диодам. В них СВЧ-модуляция тока накачки приводит к срыву одномодового режима.

Гистерезисные явления. Гистерезисные и разрывные ватт-амперные характеристики инжекционных лазеров могут быть объяснены нелинейностью оптических потерь, зависимостью показателя поглощения от интенсивности излучения. Одним из проявлений нелинейности оптических потерь является просветление поглощающих областей. К таким областям относятся периферийные области, где имеется электромагнитное поле, а накачка отсутствует или слаба, а также приповерхностная область вблизи торцевых зеркал, где концентрация свободных электронов снижена из-за их поверхностной рекомбинации. В лазерах с неоднородной накачкой поглощающая область намеренно вводится в виде участка с пониженной накачкой или без нее.

Самофокусировка в пространственных каналах генерации приводит к локальному уменьшению концентрации электронов и увеличению показателя преломления с последующей самофокусированной неустойчивостью и явлению «деформации мод». Это приводит к изменению эффективности накачки, оптических потерь и усиления. Уменьшение эффективности накачки и зарождение новых каналов уменьшает наклон ватт-амперной характеристики. Снижение оптических потерь приводит к резкому нарастанию мощности излучения вплоть до скачкообразного при

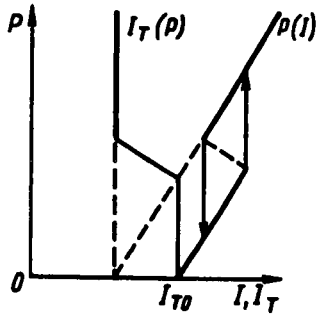


Рис. 11.32. Возникновение гистерезиса выходной мощности инжекционного лазера

положительной обратной связи — с ростом интенсивности излучения ускоряется просветление пассивных участков. В этом случае на ватт-амперной характеристике могут появиться гистерезисные петли и разрывы. В качестве примера на рис. 11.32 показано, как кусочная зависимость порогового тока I_T от мощности P приводит к S-образной расчетной ватт-амперной характеристике. Переход с одной ветви характеристики на другую на концах линейных участков образует петлю гистерезиса, создается возможность бистабильного режима работы лазера, что может быть использовано в устройствах памяти.

Генерация ультракоротких импульсов света. В настоящее время наиболее широко инжекционные лазеры применяются в волоконно-оптических линиях связи и системах высокоскоростной оптической обработки информации. Создание компактных лазерных систем, излучающих ультракороткие импульсы, предвещает появление нового поколения сверхбыстродействующих оптоэлектронных приборов. Эти системы имеют важное значение, поскольку быстродействие электронных приборов уже приближается к своему физическому пределу. Для систем оптической передачи и обработки информации наибольший интерес представляют источники пикосекундных импульсов на основе инжекционного лазера, поскольку он является единственном лазером, совместимым с элементной базой интегральной оптики и волоконно-оптических линий связи. Получение импульсов длительностью в единицу пикосекунд сулит в перспективе возможность передачи информации в системах оптической связи с огромной скоростью (свыше 100 Гбит/с), а, в конечном счете, и создание оптических компьютеров.

Генерация импульсов длительностью до 30 пс в инжекционном лазере легко достигается модуляцией тока накачки. Для получения более коротких импульсов необходима синхронизация мод. Препятствует ее реализации в резонаторе, образованном сколом граней кристалла, очень большое межмодовое расстояние (100...200 ГГц в диодах толщиной 200...400 мкм). Осуществлена активная синхронизация мод в инжекционных лазерах с использованием внешнего резонатора, обеспечивающего уменьшение межмодового расстояния до 1...10 ГГц.

Обеспечение полной синхронизации мод в спектре генерации позволяет получить минимальную длительность у импульсов с ограниченной шириной спектра. В них отсутствуют нерегулярные модуляции огибающей импульса или его фазы, присущие другим импульсам. Для обеспечения полной синхронизации мод могут быть введены селективные элемен-

ты, например, в виде отражательной дифракционной решетки, что проще и удобнее, чем селекция с помощью эталона Фабри—Перо. Пассивная синхронизация мод может быть получена с помощью двухкомпонентных лазеров с управляемой током областью поглощения. Введением насыщающегося участка в активную область инжекционного лазера с помощью протонной бомбардировки были получены импульсы длительностью 0,7 пс и частотой повторения 0,6...2,7 ГГц.

Дальнейшее укорочение длительности импульсов возможно при наличии чирпа внутри пикосекундных импульсов, генерируемых инжекционными лазерами на AlGaAs в режиме синхронизации мод. Свипирование несущей частоты генерируемых импульсов обусловлено здесь сильной зависимостью показателя преломления активной среды от концентрации свободных носителей заряда. Поэтому при прохождении ультракороткого импульса через активную среду он не только усиливается, но и воздействует на свой спектр.

В гетеролазерах из AlGaAs уменьшение несущей частоты на протяжении импульса длительностью $\tau = 17$ пс составило $\delta\omega = 0{,}5 \cdot 10^{12}$ рад/с, а соответствующий рост длины волны $\delta\lambda = 2$ Å. Увеличение длины волны генерации со временем происходит практически линейно со скоростью 0,1 Å/пс. Это позволяет оценить перепад концентрации электронов за время импульса $\delta\Delta n = -n' \dfrac{\delta\lambda}{\lambda}\left(\dfrac{\mathrm{d}n'}{\mathrm{d}\Delta n}\right)^{-1} = 2{,}3 \cdot 10^{17}$ см$^{-3}$, так как $\mathrm{d}n'/\mathrm{d}\Delta n = -3{,}6 \cdot 10^{-21}$ см$^{-3}$.

Импульсы с частотным сдвигом вниз позволяют избавиться в некоторой степени от дисперсионного уширения во внешних оптических компонентах (объективах, диэлектрических зеркалах и т. п.), поскольку в средах с нормальной дисперсией такие импульсы сжимаются до тех пор, пока не будет скомпенсирован их сдвиг частоты, после чего начинается уширение импульса. То же самое происходит при распространении подобного импульса через световод и волоконно-оптический кабель. Так в одномодовом кварцевом световоде с $\partial^2 n'/\partial\lambda^2 = 0{,}5 \cdot 10^{11}$ м$^{-2}$ длиной $L = \dfrac{\pi c^2 \tau}{\lambda^3 (\partial^2 n'/\partial\lambda^2)\delta\omega}\left[1 + \left(\dfrac{2\ln 2}{\tau\delta\omega}\right)^2\right]^{-1/2}$ импульс сжимается по длительности до $\tau_{min} = \tau(1 + \tau\delta\omega/2\ln 2)^{-1/2} = 6$ пс, т. е. почти в 3 раза.

Синхронизацию мод сталкивающихся импульсов можно реализовать в двухкомпонентном инжекционном лазере, в котором насыщающимся поглотителем служит пассивная область между усиливающимися частями (рис. 11.33). Если из внешнего резонатора лазера на диод поступают

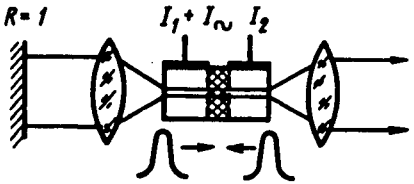


Рис. 11.33. Двухкомпонентный инжекционный лазер с внешним резонатором

два ультракоротких импульса с расстоянием между ними, равным времени прохода света между гранями диода, то после достижения первым импульсом правой грани диода, второй подойдет к левой грани. Первый импульс после отражения от грани диода, вновь усиливаясь, начнет движение в направлении, противоположном движению второго импульса. Если поглотитель находится точно в центре резонатора диода, то в поглощающей области импульсы налагаются друг на друга.

Сжатие импульса тем эффективнее, чем больше отношение показателя поглощения к коэффициенту усиления. При этом в поглощающем слое происходит непрерывное увеличение крутизны переднего фронта импульса, а в усиливающем слое — сокращение длительности за счет преимущественного усиления вершины импульса по сравнению с задним фронтом. Эффективность сжатия повышается выбором такого положения поглотителя в диоде, при котором импульсы накладывались бы друг на друга внутри него, а в усилитель проходили поодиночке.

Дополнительное сжатие временно́го профиля импульса происходит благодаря пространственно неоднородному насыщению поглотителя, обусловленному интерференцией падающих импульсов. При этом необходимо, чтобы оптическая толщина поглотителя не превосходила длины импульса и решетка населенности внутри поглотителя формировалась за время, соизмеримое с длительностью падающих импульсов. Поле дифрагировавших на пространственной решетке падающих импульсов имеет ту же фазу, что и поле прошедшей через поглотитель волны. Поэтому их интерференция усиливает прошедшие через поглотитель волны, т. е. поглотитель становится более прозрачным и соответственно более насыщенным. Это приводит к изменению формы импульсов, наблюдаемому в виде их временно́го сжатия. Переход от активно-пассивной синхронизации мод к синхронизации мод сталкивающихся импульсов в двухкомпонентном лазере позволил сократить длительность импульсов почти в четыре раза и довести ее до 0,8 пс.

Предполагается, что оптимизация двухкомпонентных лазерных диодов позволит довести длительность генерируемых импульсов до 200 фс. Естественным пределом длительности импульсов лазеров является период световых колебаний, т. е. $\tau = 3$ фс для длины волны 880 нм. Это отвечает соотношению неопределенности $\tau\mathscr{E} \geqslant \hbar$ при энергии фотона $\mathscr{E} = 2{,}3 \cdot 10^{-19}$ Дж.

Источник излучения для волоконно-оптических линий связи с высокой пропускной способностью должен не только генерировать короткие импульсы, но и обладать малой инерционностью и узким спектром излучения. В инжекционных лазерах ширина спектра генерации определяется числом продольных мод, принимающих участие в генерации, а время включения — переходными процессами. К сожалению, ширина спектра инжекционных лазеров много больше ширины спектра идеального излучателя: $\Delta\nu \approx 1/\tau$.

При модуляции тока инжекции ширина спектра генераций зависит от амплитуды модулирующего сигнала. При передаче информации короткими импульсами по одномодовому волокну необходимо учитывать уширение импульсов за счет пульсаций в нестационарном режиме. Подавление пульсаций интенсивности излучения удается обеспечить введением в активную область лазера внешнего излучения, интенсивность которого составляет $10^{-2}...10^{-3}$ от средней интенсивности стационарной генерации. Это обеспечивает сужение спектра излучения и даже установление одномодовой генерации.

§ 11.12. Инжекционные лазеры в оптической обработке информации

По мощности излучения, монохроматичности и расходимости излучения полупроводниковые лазеры значительно уступают лазерам на ионных кристаллах и стеклах, а также газовым лазерам. К преимуществам полупроводниковых лазеров относят:

1) экономичность, определяемую высокой эффективностью преобразования энергии накачки в энергию излучения;

2) быстродействие, обеспечиваемое коротким временем установления генерации ($10^{-10}...10^{-9}$ с);

3) микроминиатюрность, обусловленную огромным оптическим усилением;

4) простоту устройства, связанную с отсутствием прецизионной оптики, жесткостью монтажа, низковольтным питанием, совместимостью с интегральными схемами микроэлектроники;

5) перестраиваемость длины волны когерентного излучения под действием температуры, давления и магнитного поля.

Эти особенности открывают перспективу применения полупроводниковых лазеров для создания быстродействующих фотонных логических элементов вычислительных машин, для передачи и хранения информации.

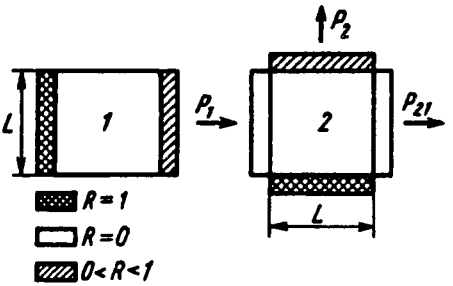


Рис. 11.34. Эффект тушения генерации лазера

Для выполнения логических операций с помощью инжекционных лазеров наряду с эффектом просветления используют *эффект тушения* генерации когерентного излучения.

Если луч P_1 одного лазера *1* проходит через активную зону второго (рис. 11.34), в последнем стимулируется излучение P_{21} в направлении падающего света. Мощность $\Delta P_2 = P_{21} - \eta P_1$ добавляется к мощности усиливаемого излучения P_1 (η — коэффициент использования света) и соответственно уменьшает инверсию населенности уровней в активной области лазера *2* и мощность P_{22} обычной лазерной моды второго лазера: $P_2 = P_{22} - \Delta P_2$. Если инверсия населенностей уровней станет ниже пороговой, лазер *2* гаснет. Этот процесс называют *тушением*, а отношение $q = \Delta P_2 / P_1$ — *коэффициентом тушения*.

Если считать коэффициент тушения постоянным, то идеальная зависимость выходной мощности второго лазера от выходной мощности первого лазера будет иметь вид прямой линии $P_2 = P_{22} - qP_1$. Однако реальная характеристика (пунктир на рис. 11.35) в силу ряда причин (в том числе благодаря спонтанному излучению) приводит к излучению второго лазера и при $P_1 > P_{22}/q$.

Простейшее *бистабильное устройство* можно составить из двух квадратных идентичных лазеров и двух отражателей (рис. 11.36, *а*). Если лазеры расположены симметрично и каждый из них может тушить генерацию другого, то их выходные мощности должны одновременно удовлетворять двум соотношениям:

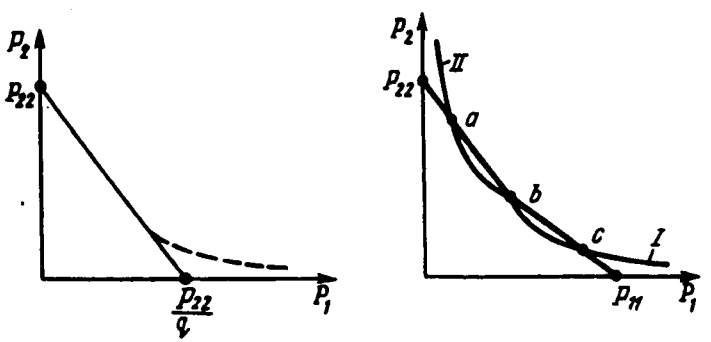


Рис. 11.35. Характеристики тушения двух лазеров

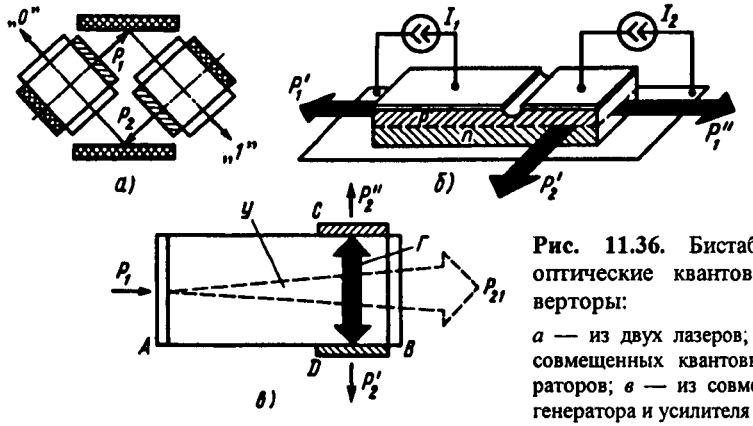


Рис. 11.36. Бистабильные оптические квантовые инверторы:

а — из двух лазеров; *б* — из совмещенных квантовых генераторов; *в* — из совмещенных генератора и усилителя

$$P_2 = P_{22} - qP_1; \text{ (I)} \qquad P_1 = P_{11} - qP_2. \text{ (II)}$$

График первого уравнения уже был рассмотрен; график второго уравнения можно получить из первого заменой координатных осей (см. рис. 11.35). Равновесным состояниям системы соответствуют три точки пересечения обеих кривых (*a*, *b*, *c*). Однако точками устойчивого равновесия являются только две из них (*a* и *c*). Эти два устойчивых состояния можно привести в соответствие с «0» или «1» двоичной системы исчисления.

Недостаток рассмотренной бистабильной схемы связан с тем, что оптическое расстояние между лазерами равно по крайней мере длине стороны лазера. Ограничение дифракционного рассеяния требует в этом случае таких размеров лазеров, которые в настоящее время неосуществимы. Эта трудность может быть преодолена в сдвоенном генераторе, изготовленном из монокристалла арсенида галлия с двумя ортогональными оптическими резонаторами с независимым питанием (рис. 11.36, *б*). Изоляция между цепями источников тока осуществляется путем пропиливания канавки в верхней металлической контактной пластине и части полупроводникового материала.

Такие приборы можно применять для реализации логических и запоминающих систем, в которых световые импульсы являются единственными носителями информации, а основной функцией электрических токов является возбуждение лазеров. В рассмотренном виде бистабильное устройство представляет собой оптический инвертор, реализующий логическую функцию НЕ.

Оптический полупроводниковый элемент, выполняющий операцию НЕ—ИЛИ, может быть получен из квантового генератора с непрерывным

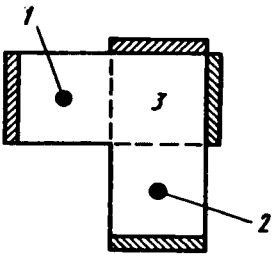


Рис. 11.37. Оптический ключ

излучением, выходные сигналы которого могут быть подавлены посредством входных сигналов от совмещенного оптического квантового усилителя (рис. 11.36, *в*). Усилитель *У* образован областью между плоскостями *A* и *B* путем их непараллельной шлифовки или нанесения поглощающих покрытий. Генератор заключен между плоскостями *C* и *D* оптического резонатора. Он занимает только часть объема усилителя. Работоспособность схемы будет обеспечена, если: а) входной световой импульс усиливается до значения, достаточного для срыва колебаний генератора; б) усиление недостаточно для самовозбуждения системы усилитель — генератор — усилитель; в) излучение достаточно для воздействия выходных сигналов по крайней мере на два других элемента (коэффициент разветвления на выходе равен не менее чем двум).

Инжекционные лазеры открыли возможность создания *оптических нейристоров*. На их основе возможно формирование узла квантово-оптического вычислительного устройства в виде монолитной интегральной схемы, состоящей из пороговых лазерных усилителей и пассивных участков. Порядок и протяженность этих соединений определяют логику устройства.

Основным элементом нейристорной интегральной схемы является полупроводниковый квантовый генератор с двумя раздельными областями инжекции. Одна из них служит для питания элементов постоянным током, а вторая обеспечивает необходимый порог генерации.

В качестве примера на рис. 11.37 показана схема диодов *1* и *2*, которые имеют пересекающиеся резонаторы с общей для обоих генераторов неинжектируемой областью *3*. Второй диод в режиме генерации может осуществлять просветление неинжектируемой области первого генератора, что приводит всю систему к порогу генерирования (путем включения добротности). В данном случае второй диод является источником внешнего оптического сигнала, а первый диод выполняет функции *оптического ключа*. Он действует на включение и выполняет логическую функцию повторения. Задержка включения снижается при повышении уровня инжекции в управляемом элементе и может достигать десятых долей наносекунды. Последовательность

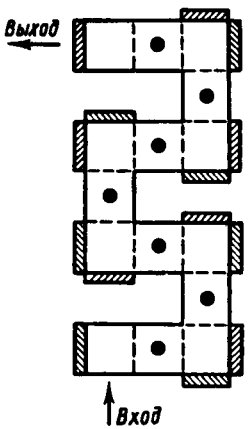


Рис. 11.38. Нейристорная цепочка из инжекционных лазеров

оптических ключей образует цепочку (рис. 11.38), которая обладает свойствами нейрона: имеет порог возбуждения, способность передавать стационарный сигнал без затухания с постоянной скоростью и *временем рефрактерности*, необходимым для восстановления первоначальных свойств цепочки после прохождения первого сигнала.

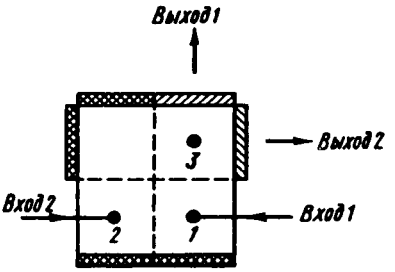


Рис. 11.39. Оптическое диодное релейное соединение

Действие оптического *релейного соединения* основано на истощении верхнего энергетического уровня в инвертируемой области *3* после излучения светового импульса (рис. 11.39). Раздельная подача сигналов на входы *1* и *2* приводит к появлению когерентного излучения на соответствующих выходах. Результаты совместного включения входных сигналов зависят от времени запаздывания между ними. Первый сигнал проходит на свой выход, а последующий вызывает появление когерентного излучения только при запаздывании на время, не меньшее рефракторного периода. Внутри рефракторного периода прохождение сигнала отсутствует. Таким образом, релейное соединение реализует функцию динамического запрета. Предельная тактовая частота релейного соединения определяет быстродействие нейристорного элемента. Она характеризуется временем восстановления пороговой инверсии и растет с увеличением тока инжекции. Длительность рефракторного периода у полупроводниковых нейристоров может доходить до 0,4 нс, что соответствует тактовой частоте 2,5 ГГц.

Интегральная *схема совпадений* (рис. 11.40) осуществляет логическую функцию конъюнкции. Самовозбуждение горизонтального генератора и сигнал на выходе появляются только в том случае, когда на входах *1* и *2* токи инжекции приводят к самовозбуждению вертикальных резонаторов и одновременному просветлению областей взаимодействия в горизонтальном резонаторе. При относительном сдвиге входных сигналов на временно́й интервал, характеризующий быстродействие схемы, сигнал совпадения отсутствует. В экспериментах минимальный интервал, налагающий запрет на выход, составил $5 \cdot 10^{-10}$ с, что соответствует тактовой частоте 2 ГГц.

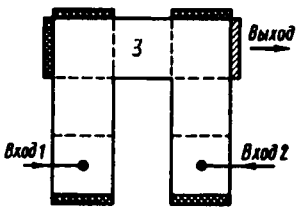


Рис. 11.40. Схема совпадений на лазерных диодах

При увеличении среднего значения тока инжекции области *3* или при увеличении ее длины сигнал на выходе включается любым из отдельных сигналов *1* и *2* или обои-

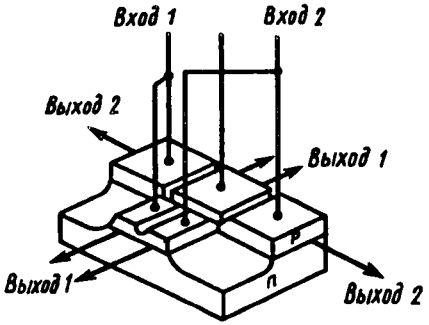


Рис. 11.41. Интегральная схема лазерного полусумматора

ми совместно. Таким образом, структура, показанная на рис. 11.40, может реализовать в рассмотренном режиме схему логического сложения, или дизъюнкции. Конструкция полусумматора представлена на рис. 11.41.

Квантовые нейристоры позволяют реализовать так называемую *нейристорную логику* и построить с ее помощью *полностью оптическую* (фотонную) вычислительную машину.

Применение инжекционных лазеров в качестве источников излучения для когерентной оптической обработки больших массивов информации даст возможность значительно уменьшить массу, габариты и энергопотребление таких систем.

Хранение информации можно осуществлять с помощью голографического запоминающего устройства. Инжекционные лазеры имеют худшую пространственную и временну́ю когерентность по сравнению с газовыми, уступают им в качестве восстановления объемных изображений. Однако для записи и считывания технических голограмм, применяемых в оптической памяти вычислительных машин, они удобны. При использовании пространственно-когерентного канала генерации с помощью инжекционных лазеров были записаны бинарные голограммы с плотностью около 10^3 бит/мм^2 при дифракционной эффективности 12%. Схема малогабаритного постоянного запоминающего устройства емкостью 10^8 бит изображена на рис. 11.42.

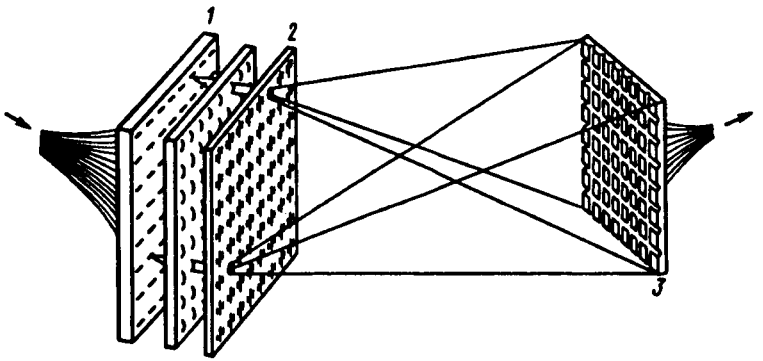


Рис. 11.42. Схема архивного запоминающего устройства:

1 — матрица инжекционных лазеров; *2* — матрица голограммы; *3* — матрица фотоприемников

Возможности использования инжекционных лазеров для голографической записи и считывания информации будут значительно расширены после промышленного освоения относительно дешевых структур с одним каналом излучения (генерация одной аксиальной моды) мощностью в десятки милливатт, а также чувствительных регистрирующих сред, обладающих высокими голографическими характеристиками в области инфракрасной области спектра. К одному из подобных инфрахроматических фотоматериалов относится И880Г со средним размером микрокристаллов AgBr 0,06 мкм с максимумом спектральной чувствительности в области длин волн 880 нм.

Началось успешное применение инжекционных лазеров и в акустооптических устройствах. Например, в двухкоординатном дифлекторе для считывания матрицы размером 32×32 голограммы использовались две последовательно установленные ультразвуковые ячейки. Эффективность двухкоординатного отклонения составила 30%, причем полезный сигнал превышал сигнал помехи примерно в 50 раз.

§ 11.13. Полупроводниковые лазеры с накачкой электронным лучом

В лазерах с возбуждением электронным пучком вторичные (неравновесные) электроны и дырки образуются в результате ионизации атомов решетки первичными электронами. Необходимость в сильном легировании здесь отсутствует.

Накачка полупроводниковых лазеров с помощью пучка ускоренных электронов не получила такого широкого распространения, как инжекционный метод возбуждения, поскольку лазеры с накачкой электронным пучком более сложны. Однако их возможности являются уникальными в части получения: 1) импульсной мощности в десятки мегаватт при высокой частоте следования импульсов и расходимости излучения в несколько минут; 2) генерации на любой длине волны в диапазоне 0,25...15 мкм; 3) пространственного сканирования луча с малоинерционным электрическим управлением; 4) квазинепрерывного излучения со средней мощностью до сотен ватт; 5) широкополосной модуляции излучения в полосе до десятков гигагерц. Первые полупроводниковые лазеры с накачкой электронным лучом были созданы в 1964 г. под руководством Н. Г. Басова.

В лазерах с поперечной накачкой пучок электронов, сформированный электронно-оптической системой, направляется на грань *ABCD* полупроводникового кристалла (рис. 11.43); параллельные грани *BCGF* и *ADHE* образуют зеркала оптического резонатора. Когда концентрация свободных носителей заряда в возбужденной области кри-

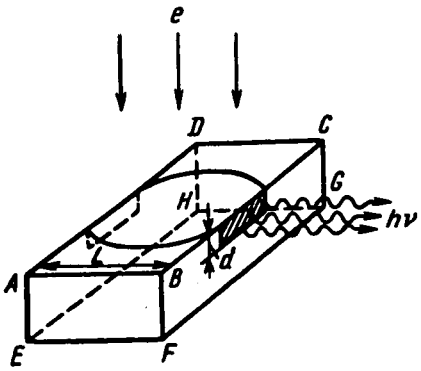


Рис. 11.43. Лазер с поперечной накачкой (d — глубина проникновения электронов)

сталла обеспечит достаточный уровень инверсии, возникает когерентное излучение, перпендикулярное плоскости зеркал и падающему электронному пучку.

Ограниченные размеры активной области в таких лазерах приводят к существенным дифракционным потерям и резкому возрастанию порога генерации в области малых энергий электронного луча. Одним из методов уменьшения дифракционных потерь является использование в активной области структуры диэлектрического волновода. Путем подбора скачка диэлектрической проницаемости можно дискриминировать поперечные моды и устранить дифракционные потери. Для этого достаточно $\delta\varepsilon'_r \approx 10^{-2}$. Причем более эффективными оказываются не основные, а высшие типы колебаний, поскольку в этом случае дифракционные потери несущественны, а усиление на проход для высших типов несколько больше, чем для основных, из-за большей длины их оптического пути. Осуществить волноводную структуру можно, например, эпитаксиальным наращиванием слоя арсенида галлия n-типа на подложку p-типа. Пороговый ток в области малых энергий электронного луча удается таким образом снизить на 1—2 порядка.

Порог генерации по плотности тока пучка накачки лежит в пределах 0,1—10 А/см^2, что соответствует удельной мощности, выделяемой в активной области $10^6...10^8$ Вт/см^3.

Геометрия активной области в виде тонкого слоя, перпендикулярного зеркалам оптического резонатора, не является оптимальной с точки зрения улучшения мощностных и когерентных свойств генерируемого излучения. Ограничивает мощность сильный нагрев рабочей области электронным пучком (при КПД лазера относительно энергии электронного пучка 10...30% бóльшая часть энергии переходит в теплоту). Поэтому генерацию удается пока осуществить только в импульсном режиме. Увеличение мощности когерентного излучения достигается лишь путем соответствующего уменьшения длительности импульса. Увеличение мощности когерентного излучения за счет удлинения резонатора ограничивает прогрессирующее увеличение потерь внутри резонатора и разрушение кристаллической структуры полупроводника полем световой волны (особенно вблизи зеркальных граней) при плотностях светового потока 10^6—10^7 Вт/см^2.

Значительно большими возможностями обладают лазеры с продольной накачкой (рис. 11. 44), в которых электронный пучок падает на кристалл в направлении, совпадающем с направлением когерентного излучения. Активная область в таком лазере определяется диаметром электронного пучка D и той частью кристалла между отражающими плоскостями $x = 0$ и $x = L$, где коэффициент усиления больше нуля.

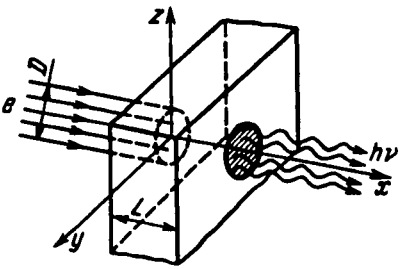


Рис. 11.44. Лазер с продольной накачкой

Высокие значения коэффициента усиления светового потока на единицу длины в полупроводниковых кристаллах ($10^3...10^4$ см$^{-1}$) позволяют на длине в несколько микрометров, сравнимой с глубиной проникновения электронного пучка, обеспечить выполнение условий генерации даже при сравнительно небольшом коэффициенте отражения зеркал.

Использование активного вещества в виде тонкого слоя, нанесенного на подложку с хорошей теплопроводностью (например, на полированную металлическую стенку вакуумной камеры), позволяет осуществить режим генерации с высокой частотой повторения или со сканированием излучения по двум координатам.

Формирование электронных пучков производится с помощью высоковольтных электронных пушек, электронно-лучевых приборов и ускорителей электронов различных типов. В качестве примера ускорителя последнего типа можно назвать резонаторные ускорители СВЧ. Они могут быть применены в тех случаях, когда использование высоковольтной аппаратуры оказывается либо нежелательным, либо невозможным.

Сканирование лазерного луча, основанное на электрооптическом или оптико-акустическом эффекте, имеет ограниченное быстродействие и большие потери света, что затрудняет получение большого числа дискретных положений в пространстве с высокой разрешающей способностью. Лазеры с электронным возбуждением, в которых используются электронно-лучевые приборы позволяют преодолеть эти трудности сравнительно просто — путем сканирования по кристаллу острофокусированного электронного пучка. Одновременно со сканированием легко можно производить широкополосную модуляцию оптического излучения практически без потерь энергии (модуляцией плотности электронного пучка).

В сканирующих лазерах с продольной накачкой оптический резонатор представляет собой плоскопараллельную полупроводниковую пластину, которая либо крепится к прозрачной подложке с хорошей теплопроводностью, либо сама должна быть достаточно толстой для

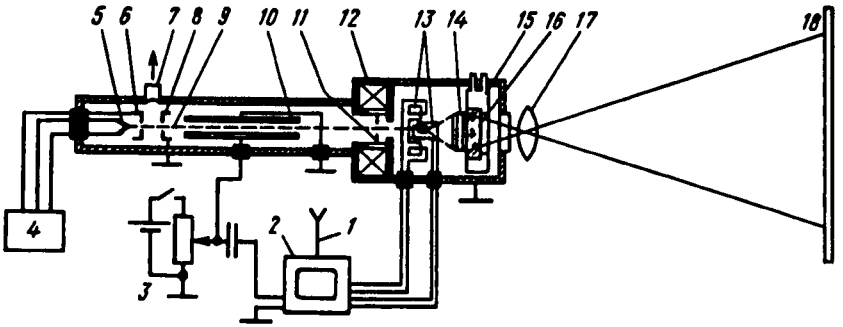


Рис. 11.45. Квантоскоп с большим экраном:

1 — антенна; *2* — телевизионный приемник; *3* — схема смещения; *4* — высоковольтный выпрямитель и выпрямитель накала катода; *5* — катод; *6* — управляющий электрод; *7* — патрубок вакуумной откачки; *8* — анод; *9* — электронный пучок; *10* — модулятор, электростатический затвор; *11* — апертурная диафрагма; *12* — магнитная линза; *13* — отклоняющая система; *14* — полупроводниковый лазерный экран; *15* — азотный криостат; *16* — сапфировая подложка; *17* — проекционный объектив; *18* — внешний экран

обеспечения механической прочности и теплоотвода. На поверхность пластинки наносят 100%-ное зеркало, через которое легко проходит электронный луч. Перемещение электронного луча по двум взаимно перпендикулярным направлениям осуществляется, как в обычной электронно-лучевой трубке (магнитной или электростатической отклоняющей системой). При этом обеспечивается сканирование в двух направлениях.

Одной из таких систем является *телевизионная трубка*, позволяющая получать изображение на экране площадью 6 м^2 (рис. 11.45). Для лазера был использован полупроводник в виде пластинки из сульфида или сульфоселенида кадмия, на которой образовывался телевизионный кадр 17×13 мм. На полупроводниковый экран с помощью магнитной линзы фокусировался электронный пучок в пятно диаметром 5...10 мкм. Перспективность подобных систем определяется их следующими преимуществами: большой яркостью свечения (до 1 МВт/см^2), преодолевающей принципиальный предел яркости нелазерных систем; высоким КПД, во много раз превосходящим КПД газовых и твердотельных лазеров (это существенно облегчает проекцию на экран с площадью поверхности более 10м^2); возможностью получения цветного изображения путем использования сочетаний различных полупроводников, например CdSe, CdS, ZnSe, а также применения хорошо разработанных в телевидении стандартных методов модуляции и сканирования электронных пучков (включение в имеющиеся телевизионные системы).

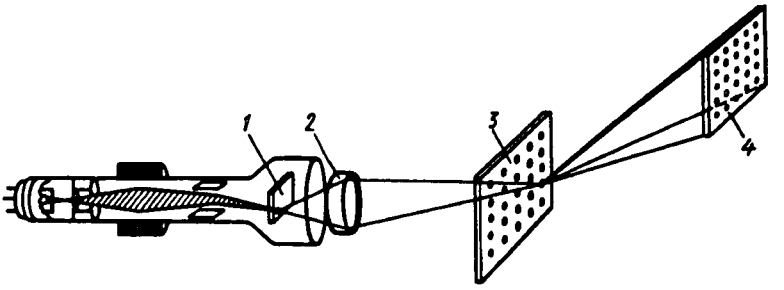


Рис. 11.46. Запоминающее устройство с лазерной электронно-лучевой трубкой:
1 — полупроводниковый лазерный экран; *2* — линза; *3* — матрица микроголограммы; *4* — матрица фотоприемников

Проекционные электронно-лучевые приборы на основе сканирующих полупроводниковых лазеров с накачкой электронным пучком получили название *квантоскопов.*

Квантоскопы позволяют наиболее эффективно осуществить отображение информации в реальном масштабе времени на экраны коллективного пользования в различного рода автоматизированных системах управления крупными объектами (например, аэродромами и пультами АЭС) и в системах автоматизированного проектирования сложных изделий машиностроения и электроники, в которых наряду с индивидуальными дисплеями требуются большие информационные экраны общего пользования, облегчающие коллективную работу проектировщиков новой техники.

Полупроводниковые лазеры с электронным возбуждением и адресацией на основе электронно-лучевой трубки позволяют сочетать высокий КПД и малую инерционность полупроводниковых лазеров с возможностью переключения луча на различные голограммы со скоростью до 10^7 положений в 1 с в *голографическом запоминающем устройстве* (рис. 11.46). Реализация одного из таких устройств обеспечила емкость памяти $3 \cdot 10^7$ бит с частотой выборки 1 МГц и потребляемой от сети мощностью 2 кВт. Мишень была изготовлена из CdS. Напряжение на трубке 60 кВ при токе 2 мА в импульсе длительностью 10 нс.

§ 11.14. Квантовая электроника и оптоэлектроника в вычислительной технике

Оптические методы обработки информации открывают новые возможности в повышении эффективности вычислительной техники. Они позволяют, в частности, производить параллельную обработку больших

массивов информации на различных уровнях, использовать картинную логику, хранить двоичный или аналоговый сигнал в голографическом виде, обрабатывать дискретные и аналоговые сигналы на одних и тех же элементах пространственных модуляторов. Реализация этих методов стала практически возможной лишь после появления оптических квантовых генераторов, обеспечивших генерацию мощного когерентного и монохроматического излучения. Поэтому лазеры являются составной частью подавляющего большинства всех проектируемых и исследуемых в настоящее время оптических вычислительных систем. Широкое внедрение их ограничивается пока несовершенством элементной базы оптоэлектроники. Тем не менее целесообразно познакомиться с некоторыми устройствами и методами обработки информации, которые обусловлены использованием фотонов в вычислительной технике и которые иллюстрируют открывающиеся перспективы.

Ассоциативное запоминающее устройство. Многие задачи вычислительной техники, связанные с обработкой больших массивов информации, могут быть решены с помощью оптических запоминающих устройств (ОЗУ) ассоциативного типа. Если в адресных ЗУ выборка информации производится путем опроса ячеек памяти с заданным номером, то в ассоциативных ЗУ поиск информации осуществляется по некоторым информационным признакам — ключевым словам. Информация извлекается из тех ячеек памяти, для которых выполняется условие близости или совпадения признака с ключевым словом запроса.

Корреляционный механизм поиска ключевых слов эффективен, когда функции автокорреляции используемых ключевых слов близки к δ-функциям, а их функции взаимных корреляций близки к нулю. В. Н. Морозовым было предложено использовать в качестве ключевых слов функции Уолша (см. § 4.14), легко реализуемых с помощью управляемых оптических транспарантов (УОТ). Формируемые ими сигналы являются ортогональными в точке отсчета, т. е. их автокорреляционные функции принимают максимальное значение в начале координат (при нулевом сдвиге), а функции взаимных корреляций в той же точке равны нулю. Это свойство ортогональных функций позволяет существенно повысить эффективность корреляционного поиска. При этом наличие сигнала в центре корреляционной плоскости оптической системы указывает на совпадение слова запроса и записанного на голограмме ключевого слова, однозначно связанного с признаком искомой информации. Равенство нулю сигнала в указанной точке при несовпадении ключевого слова со словом запроса дает возможность получить малый поисковый шум.

Ассоциативное голографическое запоминающее устройство (ГЗУ) целесообразно выполнять из двух блоков с почти идентичными оптическими схемами: блока массовой адресной голографической памяти и блока признаков.

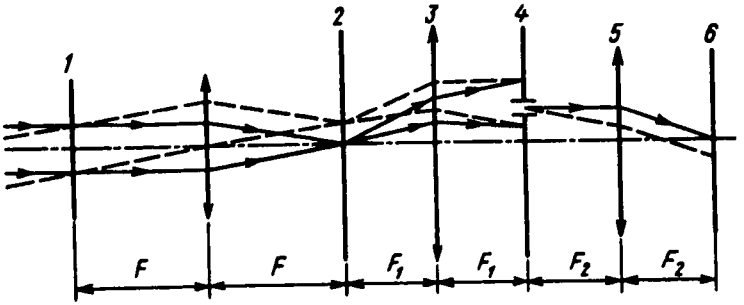


Рис. 11.47. Оптическая схема блока признаков

В блоке признаков (рис. 11.47) на строках транспаранта *1* формируются различные ортогональные последовательности, которые являются ключевыми словами. При записи фурье-голограмм транспарант освещается коллимированным световым пучком, падающим на него под некоторым углом, причем каждому адресу соответствует свой угол и свой участок регистрирующей среды *2* (все голограммы записываются на разных участках регистрирующей среды). Опорные пучки, используемые для записи различных голограмм, строго параллельны друг другу.

При восстановлении голограммы освещаются одновременно световыми пучками, прошедшими через транспарант под теми же углами, что и на стадии записи.

Если при записи голограммы было выполнено условие линейности ее регистрации, то в задней фокальной плоскости линзы *3* формируются распределения интенсивности светового потока, пропорциональные квадратам функций корреляции ключевых слов. Диафрагма *4*, установленная в этой плоскости, пропускает только центральные части корреляционных полей. Линза *5* формирует в своей задней фокальной плоскости фурье-образ корреляционного поля, пропущенного диафрагмой. В этой плоскости происходит пространственное разделение пучков, соответствующих различным адресам.

Считывание сигнала в блоке признаков осуществляет фотоприемная матрица *6*. Совмещение фотоприемников в единой интегральной схеме с логическими элементами позволяет реализовать любую логическую функцию над сигналами совпадения признаковых слов и слов запроса, в том числе операции ИЛИ, И, подсчет числа совпадающих слов и т. д. Такая матрица представляет собой однородную среду с оптическим входом, совмещенным с логической схемой обработки сигналов.

Если ключевые слова упорядочены по какому-либо параметру (например, по номеру функции Уолша) в соответствии с кодируемым признаком, то возможна организация сложного ассоциативного поиска, на-

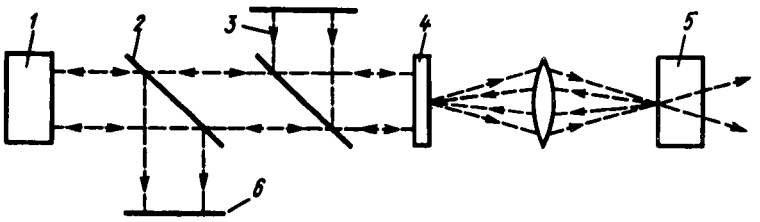


Рис. 11.48. Схема ассоциативного ЗУ нейронного типа

пример, поиск информации в окрестности требуемого признака (в том числе данных с большим или меньшим значением признака), поиск экстремума и т. д.

Одним из ограничений возможностей ассоциативного ГЗУ является число ключевых слов, формируемых в блоке признаков. Параметры существующих УОТ вряд ли позволят осуществить больше 1000 элементов в строке транспаранта. Тем не менее, этого достаточно для многих задач, например, для построения информационно-поисковых библиографических систем банков данных, поиска веществ по их молекулярным весам или спектрам и т. д. Дальнейшее увеличение числа ключевых слов (их число равно числу элементов в строке УОТ) возможно при использовании несколько одинаковых признаковых блоков в одном ЗУ.

Блок массовой голографической памяти в ассоциативном ЗУ имеет одинаковое число адресов с блоком признаков, причем адреса в соответствующих блоках совпадают. Выборка информации из ЗУ осуществляется в два этапа: сначала происходит поиск адреса в блоке признаков, а потом восстановление голограммы по найденному адресу из блока массовой памяти.

Поиск элементов по их неполному изображению позволяет осуществить оптические ассоциативные ЗУ, основанные на обращении волнового фронта. Принцип действия таких ассоциативных ЗУ можно проиллюстрировать на примере устройства (рис. 11.48), предназначенного для воссоздания (распознавания или идентификации) элемента *6*, заложенного в память *4*, после предъявления (через полупрозрачное зеркало *2*) его неполного изображения *3* на входе устройства. Для хранения информации используется набор голограмм *4*, помещенный внутри резонатора, который образуют обращающие зеркала *1* и *5*. Зеркало *5* обладает пороговыми характеристиками, позволяющими для каждой следующей итерации выделять из хранящихся в памяти те признаки, которые больше всего совпадают с признаками входного изображения.

В память *4* информация о каждом элементе вводится в виде интерференционной картины, образующейся изображением элемента и опор-

ным лучом. Для возможности различения изображений углы падения опорных лучей при записи каждый раз меняются. Все лучи, которые в опознаваемом изображении *3* соответствуют лучам одной из хранимой в пленке голограмм, складываясь, дают сильный выходной пучок; если соответствия лучей нет, выходной пучок получается слабее. Сильный выходной пучок отражается зеркалом *5*, а слабый — проходит сквозь обращающее зеркало. После второго обращения сильного пучка обращающим зеркалом *1* процесс повторяется. Последующие итерации дают возможность получить на выходе *6* полное изображение найденного (идентифицированного) элемента. Таким образом, поиск элементов в ЗУ происходит путем анализа их признаков (содержания), а не по адресу.

Ассоциативная память с обращенным волновым фронтом позволяет моделировать те элементы искусственных нейронных сетей, имитирующих работу человеческого мозга, которые пока не удалось успешно моделировать в виде СБИС. В основе модели нейтронной сети лежит математический алгоритм (Дж. Хопфилда), который сравнивает строку входных данных с хранящимися в памяти, формирует новый вариант строки данных и затем итерирует, т. е. повторяет ту же операцию, используя новый вариант в качестве входных данных. Сравнение, отбор и возвращение к началу цикла повторяются до тех пор, пока входная и результирующая строки данных не сойдутся, т. е. не станут одинаковыми. Сложность реализации обработки изображений и ассоциативной памяти на цифровых ЭВМ связана с использованием последовательности дискретных шагов. В оптической памяти обработка образной информации (массива данных) происходит одновременно по всему полю картины.

Оптоэлектронное ассоциативное ЗУ со структурой нейтронной цепи применено совместно с радиолокационной станцией для распознавания самолетов, космических кораблей и их деталей. Для нахождения изображения, наиболее точно соответствующего той или иной характеристической модели, требуется всего 10 % полного набора данных, принятых радиолокатором, что обеспечивает успешную идентификацию типа целевого самолета.

Логические и арифметические устройства. В оптоэлектронных логических элементах на основе полупроводниковых квантовых генераторов, рассмотренных в § 11.11, в качестве выходного устройства применяют обычно фотодиод. Такая полупроводниковая система обладает малой инерционностью: инжекционные лазеры обеспечивают время срабатывания до 1 пс, КПД $\approx$ 100%, а фотодиоды имеют высокую чувствительность и время фотоотклика менее 1 нс. Сначала препятствием на пути использования таких элементов была необходимость охлаждения их до температуры жидкого азота и быстрое старение (деградация). В последнее время получен непрерывный режим их работы при комнатной температуре, токах инжекции порядка десятков миллиампер и сроках

службы до десятков тысяч часов. Однако им присущи такие недостатки, как наличие двойного преобразования электричество—свет— электричество, трудности реализации систем с большим числом элементов (больше 1000) и энергопотребление, существенно большее, чем в микроэлектронике.

Для картинной обработки информации были предложены гибридные устройства, в которых основную часть операций выполняют электронные специализированные устройства, аналогичные соответствующим устройствам ЭВМ, а оптика предназначена лишь для передачи и простейшей обработки сигналов. В одном из вариантов такого *пиктографического процессора* (picture — картина) входное изображение по световодам поступает на фоточувствительные элементы. Эти элементы преобразуют световые сигналы в электрические, которые инвертируются электрическими схемами (осуществляющими операцию отрицания). Выходной сигнал инвертора поступает на светоизлучающий элемент (массив электролюминесцентных элементов). Остальные логические операции над изображениями выполняются аналогичным образом другими электронными элементами. Использование жгутов оптических волокон для передачи изображений от одного блока к другому приводит к упрощению большого числа соединений между элементами разных схем. Недостатком таких соединений является жесткая, постоянная связь между отдельными битами разных блоков, причем значения разрядов в большинстве случаев относятся к промежуточным. Изображения, обрабатываемые в таких блоках, представляют собой одноименные разряды массивов чисел. Это затрудняет обмен между изображениями, соответствующими разноименным разрядам. Выигрыш по быстродействию достигается только из-за параллелизма операций, а построенная на таком принципе вычислительная машина остается жесткой системой с постоянными логическими функциями отдельных блоков.

Значительно большими возможностями обладают логические и арифметические устройства на основе ОУТ, которые представляют собой светоклапанные устройства. Они могут быть организованы в виде матриц, каждая точка которых соответствует хранению и введению в обработку одного бита информации. Если представить обрабатываемую информацию (операнды) в парафазном коде, то ОУТ позволяет выполнить основные логические операции (логическое умножение, отрицание, эквивалентизация, импликация, сложение по модулю 2), на которые опирается решение логических уравнений.

На рис. 11.49 представлены схемы получения некоторых логических функций от двух переменных x_1 и x_2 с помощью ОУТ. Здесь управляемая поверхность расположена с правой стороны ОУТ и работает на отражении: отражает ($x = 1$) свет при наличии управляющего сигнала ($z = 1$) и не отражает ($x = 0$) его при отсутствии сигнала ($z = 0$). Управляющая поверхность расположена на левой стороне ОУТ. Световая картина, пода-

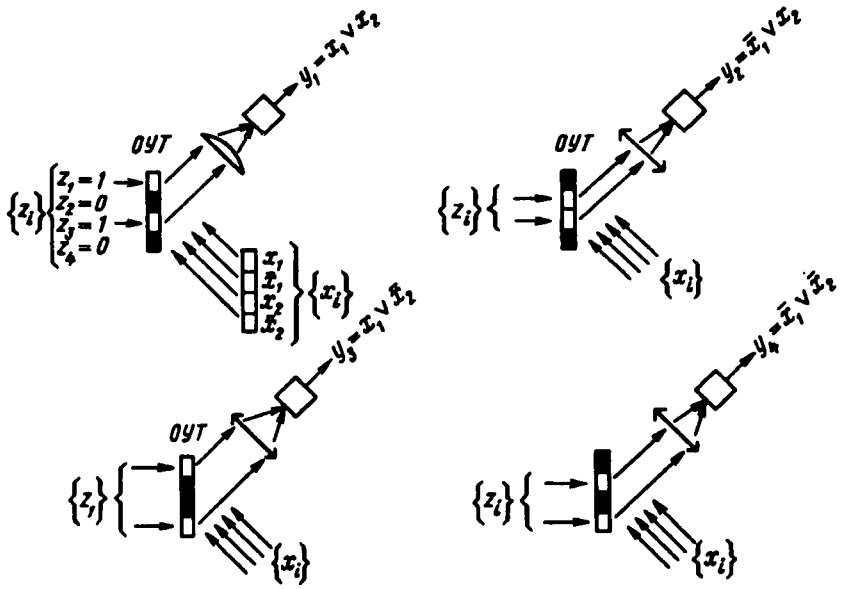


Рис. 11.49. Оптические схемы для выполнения логических преобразований

ваемая на управляющую поверхность, определяет вид дизъюнкции (логического сложения) над переменными, подаваемыми на управляемую поверхность. Реализует необходимую дизъюнкцию линза, которая собирает отраженный свет, образующий результирующую картину. Наличие яркого светового пятна в фокусе линзы означает, что дизъюнкция равна единице, а его отсутствие, что дизъюнкция равна нулю. На рис. 11.49 видно, какие сочетания $\{z_i\}$ необходимо подать на управляющую поверхность, чтобы получить операторы, обеспечивающие логические операции y_1, y_2, y_3, y_4 над входными сигналами.

Использование ОУТ с целью реализации одним и тем же оптическим *процессором широкого базиса логических операций* существенно расширяет эффективность оптоэлектронного устройства по сравнению с устройствами, в которых транспарант реализует только одну логическую функцию. Последние потребляют мощность, намного превышающую ту, которую расходуют большие электронные интегральные схемы; при большом числе транспарантов ослабление оптических сигналов в них превышает 1000. Принципиальная схема оптоэлектронного устройства, выполняющего расширенный базис логических операций, изображена на рис. 11.50.

Основными элементами, выполняющими логические преобразования, служат ОУТ и матрица гололинз ГЛ (голограмм, собирающих пре-

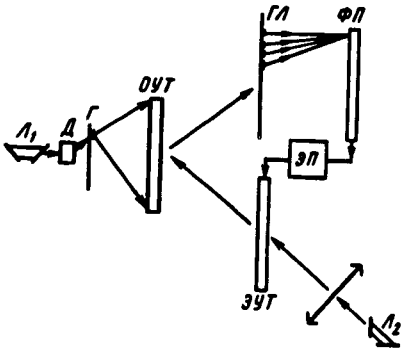


Рис. 11.50. Схема оптоэлектронного устройства с перестраиваемой логикой

ломленный световой поток в нужную точку). Управляющий оператор поступает на ОУТ из библиотеки (матрицы) голограмм *Г*. Считывание голограммы выполняет лазер $Л_1$, луч которого направляется по адресу нужной голограммы (нужной логической операции) дефлектором Д.

Обрабатываемый массив $\{x_i\}$ формируется на электрически управляемом транспаранте ЭУТ блоком электронной памяти ЭП, причем первая строка страницы переменных x_1 при парафазном коде занимает первые две строки ЭУТ (x_1 и $\bar{x}_1$), вторая строка x_2 занимает на ЭУТ следующие две строки (x_2 и $\bar{x}_2$) и т. д. Излучение лазера $Л_2$, модулированное ЭУТ, падает на управляемую поверхность ОУТ, а отраженный от ОУТ световой поток собирается на матрице фотоприемников ФП. Если имеется в виду логическое преобразование y_i функции только двух переменных x_1 и x_2, то световые пучки от первых четырех строк на ОУТ собираются на первой строке ФП.

Оптический процессор с переменными операторами, предложенный и разработанный Н. Г. Басовым и его сотрудниками, открывает принципиально новый этап в создании оптоэлектронного вычислительного комплекса высокой производительности. Его работа основана на хранении в памяти не результатов операций, как в табличном процессоре, а операторов, не зависящих от входных переменных.

Здесь используется то преимущество оптической обработки информации, что операнды (набор или матрица входных переменных) могут быть представлены в виде двумерной картины. Метод управляющих операторов реализуется благодаря тому, что при определенном представлении операндов и организации оптического канала обработки информации каждой логической или арифметической операции соответствует оператор или набор операций, зависящий только от вида операций и не зависящий от входных переменных. При этом имеется ввиду, что программа вводится через те же внешние устройства и хранится в том же блоке памяти, что и исходные данные. При этом инструкции, составляющие программу вычислений, могут быть закодированы в виде чисел и по форме ничем не отличаются от тех чисел, с которыми оперирует машина. Таким образом, при выполнении некоторой инструкции можно прочесть другую инструкцию как число, переслать ее в арифметическое устройство, про-

извести там с ней некоторые операции и вернуть на свое место в ЗУ, т. е. в свою ячейку памяти в преобразованном виде. При следующем обращении устройства управления к данной ячейке памяти за инструкцией, а не за операндом, исполняться будет уже не исходная, а преобразованная инструкция. Блок оптической обработки информации может быть реализован на основе быстродействующих цифровых оптически управляемых транспарантов.

ЗУ целесообразно строить на голографических принципах. При этом должна быть обеспечена высокая помехоустойчивость хранения операторов и их быстрое считывание, а сама двумерная информация может быть записана как в дискретной, так и в аналоговой и буквенной формах.

Высокая производительность обработки информации следует из того, что она производится по всему хранящемуся в ОУТ массиве, емкость которого для цифровой обработки может достигать 10^4 и более. Это позволяет использовать матричную и векторную формы организации вычислений, доступные в электронике только специализированным, а не универсальным ЭВМ.

Метод управляющих операторов дает возможность на несколько порядков повысить распараллеливание по командам по сравнению с реализованным в мощных ЭВМ. Действительно, емкость существующих постоянных голографических ЗУ оценивается в 10^8 бит, а объем страницы пространственно-временно́го модулятора света — в 10^4 бит, т. е. число базисных операций может составить 10^4. Это равносильно наличию такого же числа различных логических и других схем в блоках ЭВМ. Более широкий набор базисных операций позволяет находить более короткие алгоритмы вычислений, а также эффективно решать задачи и с матричной организацией и с последовательной обработкой. Общий тракт обработки информации на ОУТ делает ненужными специализированные блоки типа логических и арифметических устройств. Это не имеет аналогов в традиционных ЭВМ.

Процессор (рис. 11.51) состоит в общем случае из блоков: управляемых пучков света БУПС, записи голограмм БЗГ, носителя информации БНИ, оптически управляемых транспарантов БОУТ и буферной памяти ББП.

БУПС формирует световые пучки с заданным адресом. Блок состоит из лазера, модулятора, дефлектора, электронного блока управления ими и оптической схемы для голографической записи информации.

БЗГ формирует страницы информации, включающей в себя базисные операторы, части программ и другие числовые массивы, осуществляет запись информации в виде

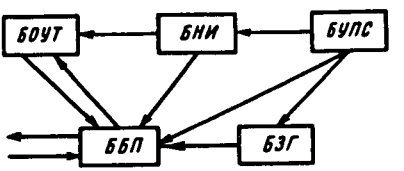


Рис. 11.51. Структурная схема оптоэлектронного процессора с переменными операторами

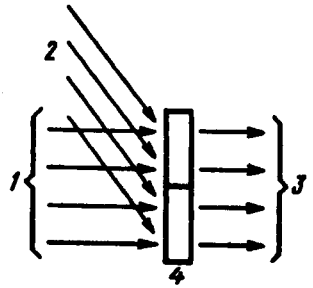


Рис. 11.52. Схема элемента ББП

голограмм на реверсивной среде в БНИ. БЗГ состоит из УОТ (формирователя страниц информации) и соответствующей оптической схемы. Считывание информации с транспаранта и запись в БНИ, стирание и перезапись голограмм в процессе работы процессора производятся с помощью БУПС. В совокупности БУПС, БЗГ и БНИ представляют собой полупостоянное оптическое голографическое ЗУ, которое служит основным накопителем информации.

БОУТ производит оптическую обработку дискретной и аналоговой информации. Он состоит из нескольких ОУТ (в плоскости изображений, восстановленных с голограмм БНИ), элементов оптической схемы, лазеров и ОУТ, позволяющих считывать информацию, производить ее преобразование и усиление яркости.

ББП осуществляет не только буферную, но и оперативную память, используя принцип ассоциативности при считывании и записи информации в голографической форме. Основную часть блока составляет носитель оптической информации, состоящий из двух одинаковых пластин реверсивной среды *4* (рис. 11.52). При записи входные данные *1* поступают на одну из этих пластин одновременно с адресом-кодом *2*, считываемым из БНИ. Они записываются в виде наложенных голограмм так, что каждому коду соответствует только одна страница информации. При этом применяется схема записи, в которой функции опорного луча выполняет кодирующий «опорный предмет», формируемый ОУТ. Использование в процессе записи некоррелируемых кодов позволяет при подаче кода (из БНИ) в процессе считывания выбирать только одну страницу информации. Наличие двух пластин регистрирующей среды дает возможность производить считывание страницы массива с одной пластины одновременно и независимо от записи информации на другую пластину. Если пластины регистрирующей среды достаточно прозрачны и на них адрес-код не подается, то возможна прямая передача входных данных *1* в выходной канал *3*.

Обработка *цифровой информации* оптическим процессором возможна в трех режимах. В режиме ввода исходная информация (числовые массивы) заносится в БНИ. В режиме обработки числовые массивы обрабатываются в БОУТ, а результат поступает в ББП или БНИ. В режиме вывода информации результаты вычислений, хранимые в БНИ и ББП, выводятся во внешнее устройство.

Обработка *аналоговой информации*, поступающей из ББП, проводится в БОУТ, а результаты передаются во внешние каналы. В гибрид-

ном варианте процессор производит одновременную обработку и цифровой и аналоговой информации.

Цифровой, аналоговый или гибридный вариант работы оптоэлектронного процессора определяется переменными операторами, содержащимися в БНИ. Эти операторы формируются управляемыми транспарантами БЗГ и в виде голограмм записываются на участках БНИ по заданным адресам. Таким образом, вместе с операторами в процессе решения задачи может быть изменена и дополнена функциональная структура процессора. Благодаря этому вычислительная машина приобретает качественно новые свойства.

Наиболее перспективным перестраиваемым логическим элементом оптического процессора является оптически управляемый транспарант с *однородной* модулирующей средой. Примером такого модулятора может быть МДП-структура с жидким кристаллом.

Фурье-оптика при нелинейной обработке сигналов. Жидкокристаллические пространственно-временные модуляторы (ПВМ) в режиме перестраиваемой решетки (РПР) позволили применить нелинейную обработку двухмерных функций, использующую преобразование интенсивности света в пространственную частоту. Модулятор состоит из тонких (тоньше 12 мкм) слоев нематического жидкого кристалла и фотополупроводника, заключенного между двумя прозрачными проводящими покрытиями. Под воздействием постоянного напряжения в жидком кристалле возникает структура фазовой решетки, период которой зависит от приложенного напряжения. В реализованных макетах частота решетки менялась в пределе от 200 до 600 мин/мм при увеличении напряжения от 20 до 50 В практически по линейному закону. В модуляторе напряжение на жидком кристалле локально управляется световым потоком, формируемым изображения и проецируемым на фотополупроводник. Преобразование интенсивности света в пространственную частоту происходит в реальном масштабе времени и параллельно во всем массиве входной информации.

Выполнение дискриминации по уровню может быть достигнуто путем расположения в фурье-плоскости щели, пропускающей только узкую полосу Δu пространственных частот u, соответствующих заданному диапазону входных интенсивностей I_1 (рис. 11.53). Перемещение щелевого фильтра позволяет непрерывно менять уровень дискриминации. Одной из областей применения этого метода нелинейной обработки информации

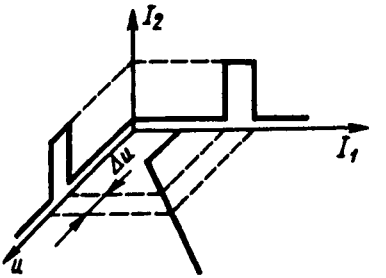


Рис. 11.53. Зависимость выходной интенсивности света от входной при РПР

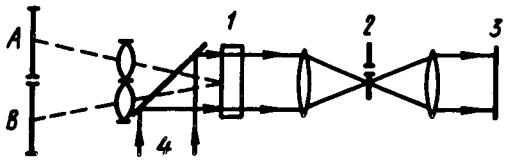


Рис. 11.54. Схема логического устройства с РПР ПВМ

является реализация оптических логических устройств. Выполнение логических операций над двухмерными двоичными входными сигналами может быть проиллюстрировано рис. 11.54. Входные двоичные сигналы A и B проецируются на слой фотополупроводника РПР ПВМ *1*. Оба некогерентных сигнала на макете формировались отфильтрованным излучением дуговой ртутной лампы. Присутствуют также оптическое смещение в виде однородной фоновой засветки, обеспечивающей ненулевые РПР-частоту и выходной сигнал при нулевом сигнале на входе $I_1 = A + B$. Считывание производится коллимированным лучом *4* гелий-неонового лазера. Выходное изображение *3* формирует фурье-фильтр *2*. В зависимости от уровня фоновой засветки (смещения) могут быть реализованы логические функции НЕ (рис. 11.55, *а*), И (рис. 11.55, *б*), ИЛИ (рис. 11.55, *в*) и исключающее ИЛИ (рис. 11.55, *г*).

Симметрия спектра РПР ПВМ позволяет реализовать полный сумматор (рис. 11.56), оперирующий во всех точках изображения одновременно. В нем отфильтрованные составляющие положительных дифракционных порядков образуют разряды суммы *4*, а отфильтрованные одновременно отрицательные порядки — выходной сигнал переноса *3*. Разрядные плоскости A и B являются входными сигналами, так же, как плоскость разрядов переноса С. Выходные сигналы формируются из считывающего луча *2*, отражающегося от РПР ПВМ *1*. Если два или более входных сигналов равны единице, то выходной сигнал переноса должен быть равен единице. Разряд суммы должен соответствовать единице, если нечетное число входных сигналов равно единице.

К перспективным можно отнести логические элементы на оптических транзисторах.

Логические и арифметические элементы на трансфазорах. В зависимости от интенсивности подведенных к трансфазору лазерных лучей он может (см. § 7.7) служить как элемент И или ИЛИ. Оптический элемент И образуется, если два падающих луча A и B подобраны так, что ни один из них сам по себе не способен переключить трансфазор, а оба вме-

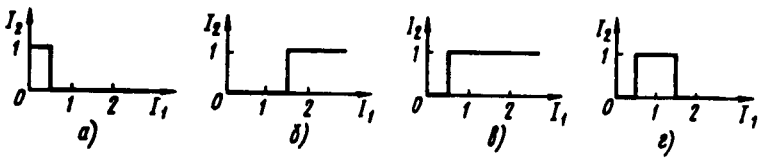


Рис. 11.55. Логические операции как двоичные нелинейные функции

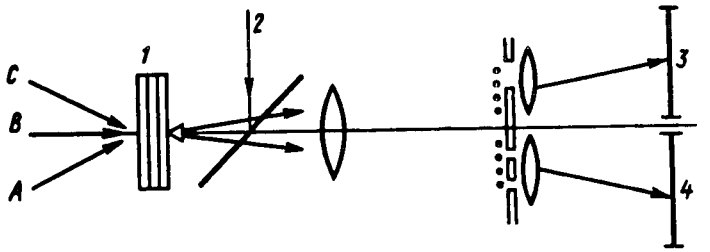


Рис. 11.56. Схема полного сумматора на РПР ПВМ

сте достаточны для его переключения (рис. 11.57, *а*). Если падающие лучи таковы, что любой из них может переключить трансфазор, то он выполняет функции схемы ИЛИ (рис. 11.57, *б*).

В элементе НЕ—И в качестве выходного сигнала используется отраженный луч, так как он является инверсией прошедшего луча — повышение интенсивности падающего луча выше порогового значения приводит к снижению интенсивности луча на выходе и наоборот (рис. 11.57, *в*).

Оптическая логическая схема НЕ может быть образована двумя связанными тонкопленочными нелинейными резонаторами из InSb использующими отраженные лучи в качестве выходных сигналов (рис. 11.58, *а*). Индивидуальные передаточные характеристики резонаторов изображены на рис. 11.58, *б*, *в*. Луч A_1 на входе первого резонатора после отражения фокусируется на входе второго резонатора, расположенного на расстоянии 500 мкм на той же тонкокристаллической пленке. Опорный луч B_1 второго элемента поддерживает его в режиме, близком к переключению. Отраженный луч B_2 служит выходным сигналом. Зависимость интенсивности луча B_2 от входного сигнала A_1 изображена на рис. 11.58, *г*, *д*. Если при увеличении интенсивности луча A_1 на входе она достигает значения, соответствующего точке переключения, выходной луч B_2 переходит на нижний уровень и будет сохраняться там до тех пор, пока луч A_2 не перейдет на нижний уровень. В реализованной схеме переключения проис-

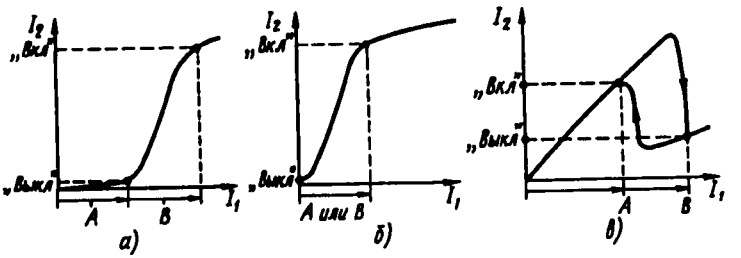


Рис. 11.57. Режимы работы трансфазора как логического элемента

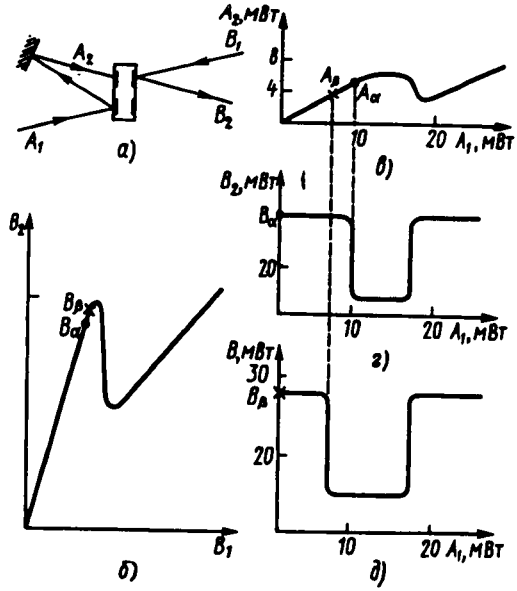


Рис. 11.58. Логическая схема на пассивных нелинейных резонаторах

ходили при мощностях входного луча 7 и 16 мВт. Напомним, что из элементов И, ИЛИ и НЕ могут быть сформированы блоки оптической вычислительной машины, выполняющей любые логические и арифметические действия.

Разработка технологии тонких пленок, прозрачных для лазерного излучения, создает предпосылки для соединения большого числа оптических интегральных схем.

Техника выращивания кристаллов и тонкопленочная технология позволяют получить тонкие монокристаллические листы большой площади. Если оптическое изображение спроецировать непосредственно на такой большой лист, каждый элемент которого представляет собой оптический переключатель, то на выходе можно получить представление изображения в цифровой форме. Затем можно выполнить его различную дальнейшую обработку, например, увеличение изображения.

Основным достоинством вычислительной машины на трансфазорах с обычной архитектурой является ее высокое быстродействие. Скорость логических схем из оптических элементов может доходить до 10^{12} оп/с, в то время, как максимальная скорость электронных переключающих устройств не превышает 10^9 оп/с.

Новые возможности в конструировании фотонного компьютера основываются на том, что один кристалл с интерферометром может выполнять несколько переключательных операций. Например, в одном нелинейном кристалле несколько лазерных пучков могут проходить параллельно независимо друг от друга. Все операции при этом могут быть различными, если соответствующим образом подобраны интенсивности входных лучей. Например, два канала могут выполнять функцию И, два — функцию ИЛИ и один канал может работать в режиме обычного транзистора.

Если при передаче модулированного сигнала с помощью светового луча происходит его расщепление, то можно выполнить одновременно

пять различных логических операций. Полученные сигналы можно снова обрабатывать либо последовательно, либо одновременно.

Свой вклад в совершенно иную организацию вычислительной машины внесут кристаллы, имеющие больше двух устойчивых оптических состояний, т. е. обладающие многократной бистабильностью. Для таких кристаллов существует много значений оптической длины пути, соответствующих целому числу полуволн в интерферометре Фабри—Перо. Каждому значению соответствует свой пик функции Эйри и максимум пропускания. Поэтому зависимость интенсивности пропускания от интенсивности падающего пучка имеет вид кривой с несколькими вертикальными ступеньками.

Если интенсивность луча смещения I_1 в оптическом транзисторе увеличивать от ступеньки к ступеньке, то, используя один и тот же слабый зондирующий луч I_3, можно получать при каждом последовательном переключении все более мощный выходной луч I_2. Так как высота вертикальных ступенек на кривой пропускания увеличивается, то коэффициент усиления K, равный приращению интенсивности прошедшего луча I_2 к интенсивности зондирующего луча будет нарастать при каждом переключении. Различные уровни пропускания можно использовать для представления значений в двоичной системе или определения состояний многозначных логических систем.

Были также обнаружены оптическая бистабильность и гистерезис при пропускании волноводной моды через тонкопленочный волновод с оболочкой, имеющей нелинейный показатель преломления. Световодные устройства позволяют использовать слабые нелинейности, поскольку нужный уровень интенсивности луча может быть сохранен на больших расстояниях. Преимуществом подобных устройств является возможность большего быстродействия из-за отсутствия характерных для резонаторов времен установления и затухания. При этом оптическая обратная связь может быть создана путем гофрировки поверхности волновода или синусоидальной модуляции показателя преломления нелинейной среды. Возможна реализация свертки пикосекундных оптических сигналов посредством нелинейного смещения в волноводе, а также аналого-цифровое преобразование на основе цепочки модуляторов Фабри—Перо на канальных оптических волноводах.

Оценка энергоемкости оптических бистабильных устройств показывает, что в диапазоне $10^{-6}...10^{-11}$ с нельзя ожидать энергии переключения, существенно меньшей энергетических затрат в полупроводниковых электронных устройствах, а при использовании джозефсоновской технологии энергия переключения значительно меньше. Однако при времени переключения от 10^{-12} до 10^{-14} с бистабильные оптические устройства не имеют конкурентов. Для осуществления логических операций в световодах на основе усиления вынужденного рас-

сеяния с быстродействием до 10^{13} переключений в 1 с на переключение затрачивается энергия $10^{-11}...10^{-12}$ Дж.

§ 11.15. Синэнергетика, лазеры и информатика

Синэнергетика в дословном переводе с греческого — это наука о совместном, согласованном действии. Основная задача синэнергетики состоит в том, чтобы вскрыть общие принципы, по которым отдельные подсистемы формируют макроскопические свойства полной системы. Ее исследования охватывают широкий круг явлений, в которых в качестве подсистем могут быть выбраны атомы, молекулы, клетки, а также компьютеры и человеческие индивидумы.

В обнаружении упомянутых общих принципов лазер сыграл фундаментальную роль. Вместе с тем синэнергетика предсказала качественно новые свойства лазерного излучения.

Примером самоорганизации системы может служить переход от излучения лампы к излучению лазера. В ансамбле двухуровневых квантовых систем статистические свойства излучения качественно меняются на пороге генерации. До порога шум в виде спонтанного излучения растет с увеличением накачки, а после порога, по мере увеличения индуцированного излучения, начинает убывать. Если ниже порога излучение состоит из коротких независимых цугов, то выше порога испускается цуг практически бесконечной длины (в результате синфазности индуцированных переходов). Поведение атомов в лазере представляет собой типичный пример самоорганизации.

Возникновение **самоорганизации в лазере** можно проследить при анализе системы уравнений (11.30)—(11.32) одномодового лазера, в которых обычно время затухания электрического поля много больше времени поперечной релаксации, определяющего затухание поляризации. Поэтому затухание электрического поля будет происходить намного медленнее поляризации, причем, согласно (11.30) временные изменения P обусловлены амплитудой E, стоящей в правой части уравнения. Это означает, что мгновенное значение амплитуды колебаний диполя определяется амплитудой напряженности поля. В этом находит проявление принцип, который в синэнергетике играет фундаментальную роль и называется *принципом подчинения*. Согласно этому принципу релаксирующие величины (например, P) принимают мгновенно свои новые значения, отвечающие новым значениям медленно изменяющихся величин (например, E), т. е. быстро релаксирующие величины являются ведомыми, а медленно релаксирующие — ведущими.

Существенным следствием принципа подчинения является значительное уменьшение числа степеней свободы. Это позволяет в случае

лазера решить систему уравнений (11.30)—(11.32) относительно напряженности электрического поля путем исключения поляризации и разности населенностей энергетических уровней, подобно тому, как это было сделано при выводе уравнения (11.47). При этом было показано, что инверсия изменяется мгновенно при изменении поля. Уравнение (11.47) является типичным уравнением, описывающим эффекты самоорганизации. Однако, оно было первым уравнением такого типа для систем, далеких от термодинамического равновесия.

Лазер стал первой системой, для которой удалось продемонстрировать возможность возникновения *неравновесных фазовых переходов*. Напомним, что возникновение макроскопической намагниченности в ферромагнетике и исчезновение электрического сопротивления в сверхпроводнике относятся к фазовым переходам в термодинамически равновесных системах. Однако, как оказалось, уравнения, определяющие свойства лазера, совпадают по форме с уравнениями, описывающими динамику изменения параметра порядка в фазовых переходах термодинамически равновесных систем. При этом аналогом температуры становится мощность накачки, при которой происходит насыщение энергетического перехода, а аналогом мощности излучения — энтропия. Дифференциальная эффективность (изменение мощности излучения, отнесенное к изменению мощности накачки) заменяет теплоемкость. Приведенная аналогия позволяет предсказать некоторые явления в лазерах, если использовать хорошо разработанную теорию равновесных систем, например, явления критического замедления и неустойчивости. В некоторых лазерных устройствах могут быть реализованы фазовые переходы, сопровождаемые возникновением петли гистерезиса и являющиеся аналогами фазовых переходов второго рода.

Закономерности синэнергетики, общие для различных научных дисциплин, дают возможность использовать аналогии процессов в системах совершенно различной природы. Эти аналогии особенно продуктивны в тех случаях, когда качественно меняется макроскопическое поведение системы. В физике лазеров примером таких качественных изменений может служить возникновение и смена различных видов неустойчивости излучения, включая нерегулярную, известную под названием *детерминированного хаоса*. Имеется ввиду состояние большого числа атомов, которые, действуя когерентно, дают хаотический свет. Это излучение ведет себя случайным образом, хотя соответствует решению уравнений (11.30)—(11.32) и никаких начальных флуктуаций не предполагает. В понятие «хаоса» не входит некоррелированное спонтанное излучение, создающее полностью случайное поле. Случайность спонтанного излучения создается флуктуациями, обусловленными его квантовой природой. Его нельзя адекватно описать упомянутыми полуклассическими уравнениями.

Аналогом хаотического лазерного излучения является гидродинамическая турбулентность, которая описывается уравнениями, аналогичными (11.30)—(11.32) и сравнительно давно рассмотрена аналитическими и численными методами. В частности, было показано, что для получения хаотического процесса нужно иметь, по крайней мере, три дифференциальных уравнения первого порядка для трех переменных. Это помогло сформулировать те условия, при которых принципиально возможно излучение хаотического света: временна́я модуляция потерь резонатора, временна́я модуляция инверсии и инжекция модулированного когерентного электромагнитного поля. Теоретические и экспериментальные исследования в гидродинамике позволили предсказать наличие в лазерах различных неустойчивостей, предшествующих достижению хаотического состояния. К ним относятся возбуждение все большего числа гармоник с разными частотами (без синхронизации частот), последовательное удвоение периода колебаний, перемежаемость стационарных и хаотических состояний при непрерывном увеличении управляющего параметра. Аналогия между уравнениями для одномодового лазера и моделью турбулентности позволила выявить такую возможность получения хаотического лазерного излучения, как увеличение потерь в резонаторе (с низкой добротностью). Результаты расчетов для одномодовых лазеров показали, что в многомодовых лазерах хаос может возникать и без внешней модуляции. Обнаружен новый режим при выходе системы из области хаоса — появление «дышащего» режима, т. е. устойчивой модуляции огибающей автопульсаций с периодом, приблизительно, в 50 раз превышающем период простых осцилляций, например, пичковых.

Способность активных систем к самоорганизации позволяет синэргетику рассматривать и как общую теорию самоорганизации в активных системах различной природы.

Предпосылкой процессов самоорганизации является поток энергии, поступающий в систему от внешнего источника и диссипирующий в термостат. Активная система не является замкнутой. При этом энтропия, служащая мерой беспорядка, может в открытой системе уменьшаться благодаря ее уходу в окружающую среду. В этом заложены основы использования **самоорганизации для обработки информации** и самообучения вычислительной сети.

Знание закономерностей самоорганизации позволяет перейти к целенаправленному конструированию искусственных активных сред и систем, процессы в которых приводили бы к образованию нужных структур. Одна из задач синэнергетики и состоит в нахождении и исследовании базовых математических моделей активных сред, исходя из типичных свойств отдельных активных элементов и законов взаимодействия между ними.

Простейшие математические модели рассматривают три типа активных элементов (автоматов): бистабильные (или триггерные), возбудимые (или мультивибраторные) и автоколебательные (автономно совершают циклические переходы через некоторую группу состояний). Моделирование возбудимой среды сетью клеточных автоматов особенно плодотворно при исследовании процессов самоорганизации в двухмерных и трехмерных средах, для которых трудоемкость решения дифференциальных уравнений оказывается чрезмерной даже для современных ЭВМ. Особенно это видно при анализе явлений в спиральных волнах, подобных вихрям в сверхпроводнике.

Одним из основных приложений закономерностей образования структур в активных средах являются задачи аналоговой обработки информации. Использование в качестве элементарной единицы обработки информации протяженных пространственных структур дает возможность резко повысить эффективность компьютера в решении проблем искусственного интеллекта.

Много веских свидетельств того, что именно аналоговые механизмы лежат в основе работы человеческого мозга. Мозг хранит и перерабатывает информацию в образах. Образ — это типичная картина (мелодия, последовательность действий), которая приобретается индивидуально, либо наследуется генетически. В обоих случаях происходит процесс обучения. Результатом обучения может быть не только распознавание изображений или последовательности событий, но и решение задач. Основной вид деятельности мозга человека — это операции с семантическими структурами (графами): их распознавание, генерация, передача, преобразование и сравнение. Каждый граф в свою очередь может рассматриваться в качестве понятий и служить элементом, входящим в состав семантической структуры следующего иерархического уровня.

Надежды на создание распределенных вычислительных сетей из простых логических элементов связаны и с молекулярной электроникой. Современный уровень технологии позволяет создавать схемы с размерами логического элемента порядка размеров полимерной молекулы и соединять их в сети на основе механизмов самосборки. Например, построен «логический вентиль» из нескольких связанных фоточувствительных органических молекул. В его работе используется перенос электронов вдоль полимерных молекул и их конфигурационные изменения. Подача и считывание информации осуществляются оптическим путем. Основная трудность на пути создания молекулярного компьютера из тысяч элементов связана с «кризисом программируемости», возникающим при необходимости выдачи единым программным блоком разнообразных индивидуальных инструкций для каждого примитивного молекулярного процессора. Преодоление этих трудностей связывают, во-первых, с созданием универсальных молекулярных компьютеров и, во-вторых, с созданием

обучающейся сети. Разработанные к настоящему времени модели обучающихся сетей часто называют «нейрокомпьютерами» или «биокомпьютерами».

Наиболее адекватной базой для будущих нейрокомпьютеров является, по-видимому, молекулярная электроника. Однако это не исключает создания обучающихся вычислительных сетей как на основе пленочной полупроводниковой технологии, так и на основе оптоэлектроники. Особенности оптических систем позволяют считать их в настоящее время реальными кандидатами для реализации аналоговых принципов обработки информации, используемых в работе мозга. К ним относят высокую степень параллельности, присущую оптическим системам: световой луч может выполнять требуемую операцию одновременно для всех элементов, формирующих двухмерную структуру, матрицу или изображение (образ). Пересекающиеся лучи не создают помех друг другу. Поэтому можно добиться в оптической системе таких плотностей соединения между логическими элементами, которые недостижимы в традиционной твердотельной электронике, но требуются в нейроподобных обучающихся схемах. Пикосекундные и фемтосекундные оптические переключатели дают возможность существенно увеличить скорость срабатывания элементов оптической системы. Преимущества оптических методов в компьютерной технологии связаны также с возможностью перехода в логических элементах к несущей частоте порядка 10^{14} Гц.

Все отчетливее проявляется тенденция использовать естественные достоинства оптики в реализации ассоциативной памяти, систем распознавания образов, обучающихся аналоговых компьютеров при создании оптического аналога нейронной сети путем использования управляемых волновых взаимодействий в распределенной нелинейной среде. Возможности реализации управляемых логических операций со световыми нелинейными образами были продемонстрированы с помощью пространственно-временны́х модуляторов на нелинейных жидких кристаллах. После перехода таких устройств на фоторефрактивные кристаллы время нелинейного отклика может быть уменьшено на шесть порядков.

Волны переключения и поперечная бистабильность в распределенной среде возникают уже в результате мелкомасштабных диффузионных поперечных взаимодействий. Нелокальная связь полей в различных точках поперечного сечения светового пучка открывает возможность реализации систем нового типа, в которых одновременно проявляются взаимодействия, имеющие существенно различный пространственный масштаб. Нелокальная связь реализована при пространственном преобразовании поля в контуре обратной связи нелинейного резонатора.

Фазовые неоднородности в нелинейных системах с диффузионными взаимодействиями приводят к возникновению волн переключения нового типа — волн переключения фазы. При этом может формироваться дви-

жущийся фазовый профиль, «проквантованный по уровню 2π». В оптически управляемом пространственно-временно́м модуляторе получен нелинейный фазовый набег порядка 2π при мощности оптического излучения, не превышающей десятков милливатт. Волне переключения фазы соответствуют бегущие выбросы интенсивности выходного поля. Таким образом, управляющее воздействие может иметь двухмерный характер, т. е. иметь вид оптического изображения.

Развитие идеи двухмерной обратной связи позволило перейти здесь от простой амплитудной бистабильности к гистерезису сигналов, гистерезису структур. Использование обращения волнового фронта («активного» зеркала) при восстановлении изображения по голограмме делает систему связи мультистабильной. Реализация одного из нескольких состояний зависит от слабого добавочного сигнала.

Нелинейная рефракция, создающая фазовую задержку в среде на уровне одного радиана, приводит к сильным поляризационным эффектам даже в простых безрезонаторных конфигурациях. Скорость установления нелинейного отклика и быстродействие поляризационных вентилей может находиться в пикосекундном и субпикосекундном диапазоне. Перспективность поляризационного кодирования информации связана с тем, что поляризационные эффекты имеют интерференционную природу, т. е. чувствуют малейшие различия фаз между ортогональными компонентами волны. Выигрыш поляризационного переключения из одного оптического канала в другой по сравнению с амплитудным взаимодействием порядка α/λ, где α — длина нелинейной среды. Кроме того, поляризационное управление светом обладает энергетическими преимуществами перед амплитудным управлением, поскольку открывает возможность построения бездиссипативных систем. Это обусловлено тем, что нелинейная гиротропия определяется действительной компонентой нелинейной восприимчивости третьего порядка, а диссипация — мнимой частью этой же восприимчивости. Их частотная дисперсия такова, что действительная часть восприимчивости спадает обратнопропорционально отстройке от нелинейного резонанса, а мнимая часть — обратнопропорционально квадрату отстройки. Выбором длины волны излучения можно найти область спектра, где поляризационное взаимодействие весьма эффективна, а поглощения практически нет. Возможности практического применения поляризационной нелинейной оптики для управления света светом и оптической связи привели к сформированию новой области исследования — поляризационной информатики.

Неполный перечень новых физических принципов оптической обработки информации и их практическая реализация показывают, что нелинейная оптика способна создать элементную базу для компьютеров

принципиально нового типа — нейронно-сетевых компьютеров, предназначенных для моделирования интеллекта.

Процессы образования и эволюции структур в нелинейных оптических системах с двухмерной обратной связью имеют много общего с процессами самоорганизации, возникающими в системах различной физической природы. Главное преимущество оптических моделей процессов самоорганизации — широкие возможности управлять этими процессами, в том числе использовать управляющие воздействия в виде образов (оптических изображений). Кроме того, здесь имеется возможность непосредственного визуального наблюдения, относительно простой реализации активных сред с большим числом компонент, широкие возможности «силового» и «параметрического» управления автоволновыми процессами. Такая система является оптическим компьютером для исследования процессов самоорганизации, автоволновой неустойчивости и хаоса.

Заключение

Элементную базу современных информационных систем составляет микроэлектроника. Требования повышения ее надежности, микроминиатюризации и быстродействия привели к поиску новых направлений ее развития, использования новых физических явлений, технологических и конструктивных решений. Результатом явилось все расширяющееся использование в информационной технике сверхвысоких частот и оптического диапазона длин волн, оптоэлектроники, акустоэлектроники, криоэлектроники, квантовой электроники и новых направлений молекулярной электроники, базирующихся на физике твердого тела и теории электромагнитного поля. В связи с этим электродинамику можно рассматривать как теоретическую основу электрофизики информационных систем. Содержание настоящей книги позволяет сделать первый шаг в изучении физических основ перспективных направлений информационной техники.

Литература

1. **Адамс М.** Введение в теорию оптических волноводов. —М.: Мир, 1984.
2. **Андрушенко Л. М., Гроднев И. И., Панфилов И. П.** Волоконно-оптические линии связи. —М.: Радио и связь, 1984.
3. **Ахманов С. А., Вислоух В. А., Чиркин А. С.** Оптика фемтосекундных лазерных импульсов. —М.: Наука, 1988.
4. **Бароне А., Д. Патерно.** Эффект Джозефсона/Пер. с англ. —М.: Мир, 1984.
5. **Барыбин А. А.** Волны в тонкопленочных полупроводниковых структурах с горячими электронами. —М.: Наука, 1986.
6. **Басов Н. Г.** О квантовой электронике. —М.: Наука, 1988.
7. **Бусырин В. И., Лярский В. Ф., Садовников В. И., Удалов Н. П.** Оптоэлектронные преобразователи на основе управляемых световодных структур. — М.: Радио и связь, 1984.
8. **Вапнэ Г. М.** СВЧ устройства на магнитостатических волнах. —М.: Изд-во ЦНИИ «Электроника», 1984
9. **Васильев А. А., Касасент Д., Компанец И. Н., Парфенов А. В.** Пространственные модуляторы света. —М.: Радио и связь, 1987.
10. **Василенко Г. И., Цибулькин Л. М.** Голографические распознающие устройства. —М.: Радио и связь, 1985.
11. **Виноградова М. Б., Руденко О. В., Сухоруков А. П.** Теория волн. —М.: Наука, 1979.
12. **Гальперович Д. Я.** Сверхпроводящие кабели связи. —М.: Радио и связь, 1986.
13. **Губанков В. Н.** Криогенная электроника//Итоги науки и техники. Сер. Радиотехника. Т. 38, 1987.
14. **В. И. Гвоздев и Е. И. Нефедов.** Объемные интегральные схемы СВЧ. —М.: Наука, 1985.
15. **Гиббс Х.** Оптическая бистабильность. Управление светом с помощью света/Пер. с англ. —М.: Мир, 1988.
16. **Голубков В. С., Евтихиев Н. Н., Пануловский Н. Н.** Интегральная оптика в информационной технике. —М.: Энергоатомиздат, 1985.
17. **Гуревич А. Г.** Магнитный резонанс в ферритах и антиферромагнетиках. — М.: Наука, 1973.
18. **Джоунс Р., Уайкс К.** Голографическая и спекл-интерферометрия. —Мир.: 1986.
19. **Додд Р., Эйлбек Д., Гиббон Д., Моррс Х.** Солитоны и нелинейные волновые уравнения/Пер. с англ. —М.: Мир, 1988.
20. **Елисеев П. Г., Богатов А. П.** Явления в полупроводниковых лазерах, связанные с нелинейной рефракцией и влиянием носителей тока на показатель преломления//Тр. ФИАН. —М.: 1986, Т. 166.
21. **Елисеев П. Г.** Введение в физику инжекционных лазеров. —М.: Наука, 1983.
22. **Ищенко Е. Ф.** Открытые оптические резонаторы. —М.: Сов. радио, 1980.
23. Интегральные схемы и микроэлектронные устройства на сверхпроводниках/Под ред. В. Н. Алфеева. —М.: Радио и связь, 1985.
24. **Зуев В. А., Попов В. Г.** Фотоэлектрические МДП-приборы. —М.: Радио и связь, 1983.
25. **Карпенков С. Х.** Тонкопленочные магнитные преобразователи. —М.: Радио и связь, 1985
26. **Кейси Х., Панин М.** Лазеры на гетероструктурах. В 2-х т. —М.: Мир, 1981.

27. **Косиевич А. М., Иванов Б. А., Ковалев А. С.** Нелинейные волны намагниченности. Динамические и топологические солитоны. —Киев: Наукова думка, 1983.
28. **Красюк Б. А., Корнеев Г. И.** Оптические системы связи и световодные датчики. —М.: Радио и связь, 1985.
29. **Кухаркин Е. С., Сестрорецкий Б. В.** Машинные методы расчета в электрофизике. —М.: МЭИ, 1986.
30. **Кухаркин Е. С., Сестрорецкий Б. В.** Диалоговая оптимизация топологии устройств в электродинамических САПР. —М.: МЭИ, 1987.
31. **Кухаркин Е. С.** Типовые задачи и примеры по инженерной электрофизике. —М.: МЭИ, 1985.
32. **Кухаркин Е. С.** Основы инженерной электрофизики. Ч.1. Основы технической электродинамики. —М.: Высшая школа, 1969.
33. **Кухаркин Е. С.** Инженерная электрофизика. Техническая электродинамика. —М.: Высшая школа, 1982.
34. **Кухаркин Е. С.** Инженерная электрофизика. Лабораторный практикум. —М.: МЭИ, 1984.
35. **Кухаркин Е. С.** Лабораторные работы по курсу инженерная электрофизика. —М.: МЭИ, 1986—1988.
36. **Кухаркин Е. С.** Оптоэлектронные приборы и методы цифровой обработки сигналов в системах реального времени//Международный форум информатизации МФИ-98: Доклады международной конференции «Информационные средства и технологии», Т. 3. —М.: изд-во «Стан-кин», 1998.
37. **Кухаркин Е. С.** Космическая информатика//Международный форум информатизации МФИ-99: Доклады международной конференции «Информационные средства и Технологии», Т 1. —М.: изд-во «Стан-кин». 1999.
38. **Лайнс М., Гласс А.** Сегнетоэлектрики и родственные им материалы. —М.: Мир, 1981.
39. **Лебедев И. В., Шнитников А. С.** Интегральная техника СВЧ. —М.: МЭИ, 1984.
40. **Лебедев И. В., Шнитников А. С.** Интегральные устройства СВЧ. —М.: МЭИ, 1985.
41. **Лихарев К. К.** Введение в динамику джозефсоновских переходов. —М.: Наука, 1985.
42. **Мидвинтер Д. Э.** Волоконные световоды для передачи информации. —М.: Радио и связь, 1983.
43. **Мировицкий Д. И., Будагян И. Ф., Дубровин В. Ф.** Микроволноводная оптика и голография. —М.: Наука, 1983.
44. **Мишустин Б. А.** Машинное проектирование СВЧ устройств. —М. МЭИ, 1978.
45. **Морозов В. Н.** Оптоэлектронные матричные процессоры. —М.: Радио и связь, 1986.
46. Нелинейные акустоэлектронные устройства и их применение./Под ред. В. С. Бондаренко. —М.: Радио и связь, 1985
47. Нелинейные электромагнитные волны./Под ред. П. Усленги. —М.: Мир, 1983.
48. **Нагаев Э. Л.** Физика магнитных полупроводников. —М.: Наука, 1979.
49. **Нефедов Е. И., Фиалковский А. Т.** Полосковые линии передачи. —М.: Наука, 1980.
50. Обращение волнового фронта лазерного излучения//Тр. ФИАН. —М.: Наука. 1986. Т. 172.
51. Оптическая голография в 2-х т./Пер. с англ.; Под ред. Г. Колфилда. —М.: Мир, 1982.
52. Оптоэлектронное ассоциативное ЗУ со структурой нейронной сети. «Электроника», 1986, Т 59, №№ 23—24.

53. Оптико-электронные методы обработки изображений/Под ред. С. В. Гуревича, Г. А. Гаврилова. —Л.: Наука, 1982.
54. Оптическая обработка изображений. —Л.: Наука, 1985.
55. Оптическая запись информации//Сб. науч. трудов —Киев: Наукова думка, 1987.
56. **Пихтин А. Н.** Физические основы квантовой электроники и оптоэлектроники. —М.: Высшая школа, 1983.
57. Применение методов фурье-оптики/Пер, с англ.; Под ред. Г. Старка —М.: Радио и связь, 1988.
58. **Пеппер Д. М.** Применения обращения волнового фронта//В мире науки, 1986, № 3.
59. Проблемы оптической голографии/Под ред. Ю. Н. Денисюка. —Л.: Наука 1981.
60. **Ривлин Л. А., Семенов А. Т., Якубович С. Д.** Динамика и спектры излучения полупроводниковых лазеров. —М.: Радио и связь, 1983.
61. Пространственные модуляторы света/А. А. Васильев, Д. Касасент, И. Н. Компанец, А. В. Парфенов. —М.: Радио и связь, 1987.
62. **Сазонов Д. М.** Антенны и устройства СВЧ. —М.: Высшая школа, 1988.
63. **Сазонов Д. М., Гридин А. Н., Мишустин Б. А.** Устройства СВЧ. —М.: Высшая школа, 1981.
64. **Свечников Г. С.** Интегральная оптика. —Киев: Наукова Думка, 1988.
65. **Свечников Г. С.** Элементы интегральной оптики. —М.: Радио и связь, 1987.
66. **Свечников С. В., Кожемяко В. П., Тимченко Л. И.** Квазиимпульсно-потенциальные оптоэлектронные элементы и устройства логико-временного типа. —Киев: Наукова думка, 1987.
67. **Селезнев В. Н.** Стимулированные светом явления в МДП-структурах//Тр. ФИАН, 1987. Т. 184.
68. **Селезнев В. Н.** Системы оптической памяти. Итоги науки и техники. Сер. Радиотехника. Т. 38, 1987.
69. **Сихарулидзе Д. Г., Чилая Г. С.** Преобразователи изображений типа МДП-электрооптический материал. —М.: Радио и связь, 1986.
70. **Скотт Э.** Волны в активных и нелинейных средах в приложении к электронике. —М.: Сов. радио, 1977.
71. Современный компьютер. —М.: Мир, 1986.
72. **Соколовский И. И., Покровский Ю. А.** Прикладная радиооптика. Теория и методы резонансной угловой фильтрации. —Киев: Наукова думка, 1986.
73. **Сонин А. С.** Введение в физику жидких кристаллов. —М.: Наука, 1983.
74. **Уласюк В. Н.** Квантоскопы. —М.: Радио и связь, 1988.
75. **Хакен Г.** Лазерная светодинамика/Пер. с. англ. —М. Мир, 1988.
76. **Ханспердже Р.** Интегральная оптика. Теория и технология. —М.: Мир,1985.
77. **Хаус Х.** Волны и поля в оптоэлектронике./Пер. с англ. —М.: Мир, 1988.
78. **Хосэгава А., Кодама Ю.** Трансформация импульса при прохождении по микроволноводу.//Тр. ТИИЭР, 1981. —М. Т. 69, № 9.
79. **Хомерики О. К.** Полупроводниковые преобразователи магнитного поля. — М.: Энергоатомиздат, 1986.
80. **Ярославский Л. П., Мерзляков Н. С.** Методы цифровой голографии. —М.: Наука, 1977.
81. **Smith S. D.** Laser, nonlinear optics and optical computers/Nature, 1985. Vol 316.
82. From optical bistability towards optical computing. The European joint optical bistability project. North-Holland. Amsterdam, New-York, Oxford, Tokyo. 1987.

Оглавление

Предисловие 3

Введение 5

Раздел первый. ЭЛЕКТРОМАГНИТНЫЕ ЯВЛЕНИЯ В ИЗОТРОПНЫХ ЛИНЕЙНЫХ СРЕДАХ 8

Глава 1. Отражение, преломление и интерференция электромагнитных волн 8
§ 1.1. Однородные и неоднородные плоские волны 8
§ 1.2. Поляризация электромагнитной волны 11
§ 1.3. Сложение волн с различной поляризацией 14
§ 1.4. Законы отражения и преломления 15
§ 1.5. Коэффициенты отражения и преломления 17
§ 1.6. Волновая матрица 19
§ 1.7. Отражение и преломление волн на границе раздела металла и диэлектрика 21
§ 1.8. Отражение и преломление волн на границе раздела прозрачных диэлектриков 24
§ 1.9. Интерферометр 27
§ 1.10. Неотражающие покрытия 31
§ 1.11. Интерференционные зеркала и фильтры 32
§ 1.12. Интерференция некогерентных волн 36
§ 1.13. Восстановление волнового фронта по интерферограмме 38
§ 1.14. Голография в информационной технике 44
§ 1.15. Голографические методы в измерительной технике 48
§ 1.16. Спекл-интерферометрия 52

Глава 2. Стоячие электромагнитные волны и объемные резонаторы 56
§ 2.1. Структура поля в прямоугольном резонаторе 57
§ 2.2. Добротность объемного резонатора 60
§ 2.3. Эквивалентная схема резонатора 62
§ 2.4. КПД резонатора 63
§ 2.5. Метод малых возмущений 65
§ 2.6. Перестройка частоты деформацией стенок резонатора 67
§ 2.7. Резонаторный метод измерения диэлектрической проницаемости 67
§ 2.8. Открытые резонаторы 69
§ 2.9. Представление поля в резонаторе в виде набора осцилляторов 76

Глава 3. Направляемые волны и волноводы 79
§ 3.1. Направляемые электромагнитные волны 80
§ 3.2. Классификация волн 83
§ 3.3. Прямоугольные волноводы 85
§ 3.4. Круглые волноводы 92
§ 3.5. Эквивалентное сопротивление волновода 96

§ 3.6. Полосковые волноводы 98
§ 3.7. Элементы и узлы микроволноводных устройств 101
§ 3.8. Диэлектрические волноводы 107
§ 3.9. Пленочные волноводы 108
§ 3.10. Оптические волокна 121
§ 3.11. Дисперсия в световодах 131
§ 3.12. Периодические структуры 138
§ 3.13. Машинные методы анализа и синтеза волноводных устройств 140

Глава 4. Излучение и дифракция электромагнитных волн 152
§ 4.1. Электродинамические потенциалы 152
§ 4.2. Элементарный электрический излучатель 154
§ 4.3. Диаграмма направленности 155
§ 4.4. Радиационное «трение» 157
§ 4.5. Спектральное представление поля излучения 159
§ 4.6. Аналогия между передачей электрических сигналов в цепях и свободном пространстве 161
§ 4.7. Импульсная характеристика свободного пространства 165
§ 4.8. Дифракция 168
§ 4.9. Дифракция на отверстиях 172
§ 4.10. Дифракция в фокусе линзы 177
§ 4.11. Оптические системы обработки информации 181
§ 4.12. Многоканальный анализ спектра 186
§ 4.13. Корреляционная фильтрация 190
§ 4.14. Математические операции, осуществляемые оптическими методами 204
§ 4.15. Автоматизация проектирования оптических устройств 208

Раздел второй. **ЭЛЕКТРОМАГНИТНЫЕ ЯВЛЕНИЯ В АНИЗОТРОПНОЙ ЛИНЕЙНОЙ СРЕДЕ 210**

Глава 5. Электромагнитное поле в немагнитных диэлектриках 210
§ 5.1. Тензор диэлектрической проницаемости 210
§ 5.2. Электромагнитная волна в анизотропной среде 212
§ 5.3. Обыкновенные и необыкновенные волны 215
§ 5.4. Двойное лучепреломление 216
§ 5.5. Квадратичный электрооптический эффект 217
§ 5.6. Линейный электрооптический эффект 220
§ 5.7. Электрооптические явления в жидких кристаллах 224
§ 5.8. Электрооптические модуляторы света 230
§ 5.9. Электроуправляемые микроволноводные элементы 236

Глава 6. Электромагнитное поле в гиротропной среде 244
§ 6.1. Магнитная проницаемость намагниченного феррита 245
§ 6.2. Электромагнитные волны в гиротропной среде 252
§ 6.3. Волны в продольно намагниченном феррите 255
§ 6.4. Волны в поперечно намагниченном феррите 257
§ 6.5. Невзаимные явления в волноводах с ферритами 258
§ 6.6. Магнитостатистические и спиновые волны 265

Раздел третий. ЭЛЕКТРОМАГНИТНЫЕ ЯВЛЕНИЯ В НЕЛИНЕЙНЫХ СРЕДАХ И УСТРОЙСТВАХ 272

Глава 7. Электромагнитные явления в сегнетоэлектриках 272
§ 7.1. Сегнетоэлектрики 273
§ 7.2. Сегнетоэлектрики в приборах СВЧ-диапазона 275
§ 7.3. Оптические свойства сегнетоэлектриков 277
§ 7.4. Фотосегнетоэлектрические явления 285
§ 7.5. Обращение волнового фронта 289
§ 7.6. Оптически управляемый пространственно-временно́й модулятор 293
§ 7.7. Самовоздействие световых импульсов 296
§ 7.8. Нелинейные эффекты в диспергирующих средах 301
§ 7.9. Акустоэлектромагнитные волны в пьезокристаллах 311
§ 7.10. Электроакустические волны в пьезополупроводниках 315

Глава 8. Электромагнитные явления в ферромагнетиках 321
§ 8.1. Магнитные пленки с плоскостной анизотропией 321
§ 8.2. Магнитные пленки с нормальной анизотропией 328
§ 8.3. Цилиндрические магнитные домены 334
§ 8.4. Доменная стенка как нелинейная волна намагниченности 345
§ 8.5. Магнитооптические явления в ферромагнитных материалах 351
§ 8.6. Дифракционная эффективность магнитооптической бинарной голограммы 355

Глава 9. Электромагнитные явления в полупроводниках 361
§ 9.1. Дрейф свободных носителей заряда в электрическом поле 361
§ 9.2. Дрейф свободных носителей заряда в магнитном поле 363
§ 9.3. Дрейф свободных носителей заряда в скрещенных электрическом и магнитном полях 364
§ 9.4. Подвижность свободных носителей заряда в поперечном магнитном поле 369
§ 9.5. Магнитодиодный эффект 374
§ 9.6. Резонансные свойства полупроводникового резистора 375
§ 9.7. Электрические домены в полупроводниках 378
§ 9.8. Логические функциональные схемы на основе эффекта Ганна 387
§ 9.9. Нарастающие волны объемного заряда 391
§ 9.10. Ограничение накопления объемного заряда 398
§ 9.11. Резонансные свойства полупроводникового материала 401
§ 9.12. Преломление и поглощение электромагнитных волн в полупроводнике 404
§ 9.13. Полупроводниковые оптические модуляторы 408
§ 9.14. Магнитооптические явления 416
§ 9.15. Эффект фотонного давления 420
§ 9.16. Внутренний фотоэффект 421
§ 9.17. Фотопроводимость 424
§ 9.18. Фотогальванический эффект 427
§ 9.19. Элементы фотопамяти на МНОП-структурах 432
§ 9.20. Оптически управляемые пространственно-временные модуляторы света на МДП-структурах 437

§ 9.21. Электролюминесценция 443
§ 9.22. Оптроны 452
§ 9.23. Интегральная оптоэлектроника 462

Глава 10. Электромагнитные явления в сверхпроводниках 482
§ 10.1. Электрон-фононное взаимодействие и куперовские пары 482
§ 10.2. Эффект Мейсснера 486
§ 10.3. Магнитные свойства сверхпроводников первого рода 490
§ 10.4. Магнитные свойства сверхпроводников второго рода 496
§ 10.5. Магнитные свойства сверхпроводников третьего рода 502
§ 10.6. Слабосвязанные сверхпроводники 507
§ 10.7. Туннелирование сверхпроводящих электронов в магнитном поле .. 513
§ 10.8. Макроскопическая квантовая интерференция 515
§ 10.9. Электродинамика джозефсоновского контакта 521
§ 10.10. Динамика флуксона 525
§ 10.11. Микроэлектронные устройства на сверхпроводниках для обработки, запоминания и преобразования информации 529

Глава 11. Электромагнитные явления в активных средах 543
§ 11.1. Квантовый осциллятор 543
§ 11.2. Квантование поля 546
§ 11.3. Взаимодействие электромагнитного излучения с веществом 549
§ 11.4. Уравнения движения для дипольного момента 554
§ 11.5. Резонансные процессы 557
§ 11.6. Квантовый усилитель 562
§ 11.7. Квантовый генератор 577
§ 11.8. Синхронизация мод 593
§ 11.9. Формирование ультракоротких импульсов 598
§ 11.10. Инжекционные лазеры 599
§ 11.11. Динамические нелинейные явления в инжекционных лазерах 622
§ 11.12. Инжекционные лазеры в оптической обработке информации 633
§ 11.13. Полупроводниковые лазеры с накачкой электронным лучом 639
§ 11.14. Квантовая электроника и оптоэлектроника в вычислительной технике 643
§ 11.15. Синэнергетика, лазеры и информатика 658

Заключение 664

Литература 665

Учебное издание

Кухаркин Евгений Степанович

ЭЛЕКТРОФИЗИКА ИНФОРМАЦИОННЫХ СИСТЕМ

Редактор *Л.В. Честная*
Художник *В.Ю. Соколова*
Художественный редактор *Ю.Э. Иванова*
Технический редактор *Н.В. Быкова*
Корректор *В.А. Жилкина*
Набор и компьютерная верстка *С.Ч. Соколовского*

ЛР № 010146 от 25.12.96. Изд. № ВТИ-97. Сдано в набор 04.10.2000
Подп. в печать 26.03.2001. Формат 60x88 1/16. Бумага газетная. Гарнитура «Таймс»
Печать офсетная. Объем: 41,16 усл. печ. л., 41,16 усл. кр.-отт., 38,97 уч.-изд. л.
Тираж 7000 экз. Заказ № 1043

ГУП «Издательство «Высшая школа», 127994, Москва, ГСП-4,
Неглинная ул., 29/14

Факс: 200-03-01, 200-06-87 E-mail: V-Shkola@g23.relcom.ru
http://www.v-shkola.ru

Отпечатано во ФГУП ИПК «Ульяновский Дом печати»
432980, г. Ульяновск, ул. Гончарова, 14