

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное
образовательное учреждение
высшего профессионального образования
«Пермский национальный исследовательский
политехнический университет»

А.И. Цаплин

ФОТОНИКА И ОПТОИНФОРМАТИКА

Введение в специальность

Утверждено
Редакционно-издательским советом университета
в качестве учебного пособия

Издательство
Пермского национального исследовательского
политехнического университета
2012

УДК 536.7: 621.036

ББК 22.3

Ц25

Рецензенты:

доктор технических наук, профессор,
заведующий кафедрой прикладной математики *В.П. Первадчук*
(Пермский национальный исследовательский
политехнический университет);

кандидат технических наук, начальник
производственно-конструкторского отдела *И.И. Крюков*
(Пермская научно-производственная приборостроительная компания)

Цаплин, А.И.

Ц25 Фотоника и оптоинформатика. Введение в специальность :
учеб. пособие / А.И. Цаплин. – Пермь : Изд-во Перм. нац. исслед.
политехн. ун-та, 2012. – 399 с.

ISBN 978-5-398-00898-2

Рассматриваются система современного высшего технического образования, особенности обучения в вузе, фундаментальные основы инженерной деятельности.

Представлены исторические этапы зарождения фотоники и оптоинформатики, определена роль фотонов как носителей информации и энергии на современном этапе. Приведены необходимые для понимания на квантовом уровне теоретические основы физики, научные и нанотехнологические основы фотоники. Рассматриваются принципы работы лазеров, оптических волокон, перспективы и тенденции дальнейшего развития компьютеров на основе фотонов.

Предназначено для студентов вузов, обучающихся по направлению бакалаврской подготовки «Фотоника и оптоинформатика», профиль «Волоконная оптика».

УДК 536.7: 621.036

ББК 22.3

ISBN 978-5-398-00898-2

© ПНИПУ, 2012

ОГЛАВЛЕНИЕ

Предисловие	6
Часть 1. Фундаментальные основы высшего образования.....	10
1. Особенности высшего технического образования	10
1.1. Современная система высшего образования и его цели	10
1.2. Особенности обучения в вузе	12
2. Фундаментальные основы творческой деятельности.....	14
2.1. Направления творческой деятельности.....	14
2.2. Практическая деятельность человека и современное естествознание.....	17
2.3. Естественно-научные основы практической деятельности человека.....	20
2.4. Эволюция вселенной и общность законов природы.....	27
2.5. Деятельность специалиста и реальность.....	33
3. Фундаментальные основы дисциплин учебного плана	40
3.1. Особенности Федерального государственного образовательного стандарта по направлению «Фотоника и оптоинформатика»	40
3.2. Математические и естественно-научные дисциплины.....	42
3.3. Профессиональные дисциплины	45
3.4. Гуманитарные, социальные и экономические дисциплины	50
3.5. Основные требования бакалаврской подготовки.....	53
Часть 2. Научные основы фотоники	56
4. Элементы квантовой физики	56
4.1. Связь фотоники и оптоинформатики с квантовой физикой	56
4.2. Этапы развития фотоники и оптоинформатики	58
4.3. Основные представления квантовой механики.....	66
4.4. Квантовая модель атома	71
4.5. Понятие о потенциальных ямах и барьерах	74
4.6. Микрочастица в прямоугольной потенциальной яме.....	77
4.7. Туннельный эффект	81
5. Элементы физики твердого тела	83
5.1. Кристаллические решетки	83
5.2. Дефекты кристаллического строения.....	89

5.3. Элементы зонной теории	93
5.4. Энергетический спектр кристалла.....	96
5.5. Понятие эффективной массы электрона	99
5.6. Экситонные эффекты	100
6. Физические основы оптики.....	104
6.1. Электромагнитная природа света	104
6.2. Основные явления волновой оптики	111
6.3. Основные явления квантовой оптики	119
7. Элементы нелинейной оптики	123
7.1. Механизмы оптической нелинейности	123
7.2. Вынужденное рассеяние света	127
7.3. Самофокусировка	128
7.4. Нелинейные эффекты в волоконных световодах	130
7.5. Оптические солитоны	133
Часть 3. Физические и нанотехнологические основы фотоники	138
8. Полупроводниковые квантовые структуры	138
8.1. Роль полупроводниковых структур в оптоэлектронике.....	138
8.2. Твердотельные гетероструктуры. Полупроводниковый гетеропереход	139
8.3. Квантоворазмерные структуры, их самоорганизация	144
8.4. Применение квантовых структур в приборах оптоэлектроники	158
9. Основы нанотехнологий получения оптических материалов	178
9.1. Исторические аспекты	178
9.2. Наноматериалы	189
9.3. Оптические метаматериалы	212
9.4. Методы формирования наноструктур	218
9.4.1. Молекулярно-лучевая эпитаксия.....	220
9.4.2. Нанолитография	224
9.4.3. Сканирующая туннельная и атомно-силовая микроскопия	226
9.5. Применение нанотехнологий в технике.....	231
10. Лазеры	236
10.1. Спонтанное и вынужденное излучение, поглощение.....	237
10.2. Принцип работы лазера	240
10.3. Схемы накачки.....	241
10.4. Свойства лазерных пучков	242
10.5. Типы лазеров.....	245
10.6. Области применения лазеров	255

11. Оптические волокна.....	270
11.1. Общие сведения.....	270
11.2. Типы оптических волокон.....	281
11.3. Материалы для изготовления оптических волокон.....	293
11.4. Технология изготовления оптических волокон.....	295
11.5. Механическая прочность оптических волокон.....	298
11.6. Принцип работы волоконного оптического гироскопа.....	300
Часть 4. Основы оптоинформатики и вычислительного	
эксперимента.....	305
12. Основы оптоинформатики.....	305
12.1. Предмет и задачи информатики.....	305
12.2. История информационных технологий.....	308
12.3. Понятие об информации.....	318
12.4. Измерение количества информации. Энтропия.....	326
12.5. Архитектура компьютера.....	334
12.6. Предельные возможности электронной компьютерной техники.....	349
12.7. Оптические системы обработки информации.....	353
13. Основы математического моделирования неравновесных теплофизических процессов в фотонике.....	373
13.1. Роль тепло- и массообмена в фотонике.....	373
13.2. Виды теплообмена. Законы молекулярного тепло- и массообмена.....	375
13.3. Перенос тепла теплопроводностью.....	378
13.4. Основы вычислительного эксперимента в теплофизике.....	382
Заключение.....	391
Список литературы.....	393

ПРЕДИСЛОВИЕ

Развитые страны сегодня находятся в состоянии перехода от «индустриального» человеческого общества к «обществу информационному», отличительная особенность которого состоит в создании и непрерывном усовершенствовании сложных «интеллектуальных сетей» – систем быстрого, эффективного и экономичного предоставления информационных услуг массовому пользователю. Увеличение объема и скорости передачи информации в высокопроизводительных интеллектуальных сетях требует разработки соответствующих технических средств, среди которых оптика и оптические методы передачи сигналов играют важнейшую роль.

Фотоника и оптоинформатика – это новое, стремительно развивающееся в России направление подготовки на базе оптики, математики и компьютерных технологий, это обработка и передача информации и энергии с помощью квантов электромагнитного поля – фотонов. Оптоволоконные системы с высокой скоростью передачи данных, голографические запоминающие устройства сверхбольшой емкости, многопроцессорные компьютеры с оптической межпроцессорной связью, в которых свет управляет светом, – вот далеко не полный перечень объектов фотоники и оптоинформатики. Для решения широкого класса задач в различных областях науки и техники – от физики и химии до биологии и медицины активно используются лазерные технологии. С помощью лазерного излучения производятся различные технологические операции, исследования, измерения и диагностика.

Решение задач получения искусственных материалов, кристаллов, имеющих рекордно низкие оптические потери при передаче информации и энергии, стало возможным с достижениями успехов в нанотехнологиях. Нанотехнология – ключевое понятие начала XXI века, символ новой, третьей научно-

технической революции. С позиций сегодняшнего дня целью нанотехнологий является создание наносистем, наноматериалов, наноустройств, способных оказать качественное воздействие на развитие цивилизации. Первая часть сложного слова нано- вообще означает одну миллиардную (10^{-9}) чего-либо. Нанотехнология – совокупность методов изготовления и обработки изделий, имеющих протяженность 1–100 нм (хотя бы в одном измерении). Нанометровый диапазон измерений размеров открывает новые свойства и подходы к изучению вещества. В этом диапазоне меняются многие физические и химические свойства и нигде так близко не сходятся физика, химия и биология. Напомним, что $1 \text{ нм} = 10^{-9} \text{ м} = 10^{-3} \text{ мкм} = 10 \text{ \AA}$. Атом имеет размер порядка 0,1 нм, неорганические молекулы ~1 нм, вирусы – от 10 до 500 нм; бактерии ~1000 нм. Десятичные кратные и дольные приставки и множители в международной системе единиц представлены в таблице.

Широкое применение в различных областях современной техники находят различные волноводные структуры. Уже сегодня волоконно-оптические технологии определяют уровень развития таких важных сфер государственной деятельности, как экономика, образование и безопасность.

Оптимизация технологических процессов получения оптических волокон, лазерной обработки материалов предполагает наряду с экспериментальными исследованиями и применение методов математического моделирования.

Целью курса «Фотоника и оптоинформатика. Введение в специальность» является ознакомление студентов с современной системой высшего технического образования, его основными задачами, организационными и методическими особенностями обучения в вузе, с документами, которые регламентируют учебу студентов, а также фундаментальными, общетехническими и профессиональными основами выбранной специальности, спецификой будущей работы выпускника, перспективами его трудоустройства.

Наименование	Обозначение русское	Обозначение международное	Десятичная запись	Обыкновенная запись
йокто	и	y	10^{-24}	0,000000000000000000000001
zepto	з	z	10^{-21}	0,00000000000000000000001
атто	а	a	10^{-18}	0,0000000000000000001
фемто	ф	f	10^{-15}	0,000000000000001
пико	п	p	10^{-12}	0,000000000001
нано	н	n	10^{-9}	0,000000001
микро	мк	m	10^{-6}	0,000001
милли	м	m	10^{-3}	0,001
сантиметры	с	s	10^{-2}	0,01
деци	д	d	10^{-1}	0,1
дека	да	da	10^1	10
гекто	г	h	10^2	100
кило	к	k	10^3	1000
мега	М	M	10^6	1000000
гига	Г	G	10^9	1000000000
тера	Т	T	10^{12}	1000000000000
Пета	П	P	10^{15}	1000000000000000
Экса	Э	E	10^{18}	1000000000000000000
зетта	З	Z	10^{21}	1000000000000000000000
йотта	И	Y	10^{24}	1000000000000000000000000

Пособие состоит из четырех частей. Первая часть посвящена фундаментальному и гуманитарному аспектам бакалаврской подготовки, без которых невозможно сформировать широкообразованного, системно мыслящего, ориентированного на многоаспектную творческую деятельность специалиста, способного с максимальной эффективностью продолжить углубленное образование в выбранном направлении. Такой подход соответствует национальной доктрине образования в Российской Федерации и макроперемемам в современном высшем образовании, связанным с переходом к экономике, основанной на знаниях.

Во второй части пособия рассматриваются научные основы фотоники с элементами квантовой физики, оптики, обсуждается и дополняется информация, полученная студентами в кур-

сах физики и химии средней школы и позволяющая осмыслить эти основы.

Третья часть посвящена научным и нанотехнологическим аспектам фотоники, путям и перспективам ее развития. Показано, что нанотехнологии – это одно из наиболее быстро развивающихся направлений получения оптических материалов. В свою очередь, достижения нанотехнологий обязаны применению устройств и систем, в которых генерируются, усиливаются, модулируются, распространяются и детектируются оптические сигналы.

Основы оптоинформатики – технической науки, занимающейся проблемами передачи, хранения и обработки информации, рассмотрены в четвертой части. Показано, что волновая и корпускулярная природа света обуславливает многочисленные преимущества фотона как носителя информации перед электроном в современных компьютерах. На примере теплофизических задач в фотонике рассмотрены основы вычислительного эксперимента.

Учебное пособие предназначено для студентов первого курса бакалаврской подготовки по направлению «Фотоника и оптоинформатика» в Пермском национальном исследовательском политехническом университете.

ЧАСТЬ 1. ФУНДАМЕНТАЛЬНЫЕ ОСНОВЫ ВЫСШЕГО ОБРАЗОВАНИЯ

1. ОСОБЕННОСТИ ВЫСШЕГО ТЕХНИЧЕСКОГО ОБРАЗОВАНИЯ

1.1. Современная система высшего образования и его цели

Система высшего образования страны включает около 1000 высших учебных заведений, из них более 500 – государственные. По статусу вузы делятся на классические университеты, технические университеты, академии и институты. По профилю – на многопрофильные и узкопрофильные, например сельскохозяйственные, медицинские и т.д. Мощную и дорогостоящую систему образования страна содержать бы не стала, если бы эта система не обеспечивала решение важнейших государственных задач, а именно:

- повышение безопасности страны (в самом широком смысле);
- подготовка специалистов для всех направлений экономики;
- повышение интеллектуального уровня населения.

Безопасность любой страны обеспечивается в основном уровнем образованности населения. Так было во все времена; особенно это важно в условиях ускоренного развития наукоемких производств, наукоемких видов техники и вооружения. Великий китайский мыслитель Конфуций около 2500 лет назад назвал образованность населения одним из условий преуспевания государства. Правительство США неизменно обосновывает все мероприятия по развитию и улучшению системы образования интересами безопасности страны. В последних документах правительства Российской

Федерации, направленных на совершенствование системы образования, необходимость его модернизации также связывается с безопасностью страны.

В соответствии с новой образовательной парадигмой у высшего технического образования две основные задачи:

- подготовка высококвалифицированного специалиста, обладающего профессиональными компетенциями;
- формирование широкообразованной личности и общих знаниевых компетенций.

Первая задача – более узкая и простая; она у каждого своя (как и специальность) и не рассчитана на всю жизнь (многие специалисты вынуждены неоднократно менять свою специальность). Вторая задача универсальна, т.е. одинакова для всех, и ее значимость не меняется в течение профессиональной деятельности специалиста. Без решения второй задачи полноценно решить первую невозможно. Одно из важнейших требований к широкообразованной личности – это творческое системное мышление и способность обеспечивать в условиях научно-технического прогресса устойчивое существование человечества на Земле. Качества широкообразованной личности и общие компетенции можно приобрести, лишь глубоко освоив фундаментальные и гуманитарные основы выбранной специальности. Компетенция – совокупность взаимосвязанных качеств человека, позволяющих ему эффективно выполнять свои профессиональные обязанности (*профессиональные компетенции*), успешно ориентироваться в жизненных и служебных ситуациях (*общие компетенции*).

Учебный план включает три блока дисциплин:

- 1) *гуманитарные, социальные и экономические* (ГСЭ);
- 2) *математические и естественно-научные* (МЕН);
- 3) *профессиональные* (ПД).

Каждый из трех блоков содержит *федеральный* и *региональный* компоненты, а также *базовую* и *вариативную* части, в том числе дисциплины по выбору студента (*элективные*).

Гуманитарные и фундаментальные знания сосредоточены в блоках ГСЭ и ЕН. Эти дисциплины изучаются на первых курсах.

Фундаментальные знания превратились в наиболее эффективную движущую силу производства. *Фундаментальные знания создаются фундаментальными науками, т.е. науками, посвященными исследованию природы. К таким наукам относятся: физика, математика, информатика, химия, биология и некоторые другие.* Инженерные теории, как правило, представляют собой модифицированные варианты теорий фундаментальных наук. Модификация состоит в том, что фрагменты общенаучных теорий, имеющие прикладное значение, преобразуются в теории, позволяющие выполнять инженерные расчеты и проекты. Поэтому любая инженерная дисциплина содержит фундаментальное ядро. Задача студента – научиться выделять из различных дисциплин фундаментальные знания, интегрировать и обобщать их в своем сознании. Без этого не может сформироваться широкообразованный специалист. Справиться с этой задачей помогают специальные учебные дисциплины, в которых интегрируются основополагающие знания. «Введение в специальность» – одна из интегрирующих дисциплин.

1.2. Особенности обучения в вузе

Основной особенностью обучения в вузе является то, что в вузе не учат, а учатся. Студент – взрослый человек, осознанно выбравший данную специальность, и поэтому обязан сам осваивать все дисциплины учебного плана, позволяющие ему стать специалистом-профессионалом. Очевидно, что человек, не проявляющий самостоятельность в учебе, не станет самостоятельным и в работе. Поэтому в университете не столь тщательно, как в школе, контролируются текущие знания студента; значительная роль отводится самоконтролю.

Отсутствие постоянного самоконтроля в учебе – основной признак того, что поступивший в вуз еще не созрел, чтобы быть студентом. В настоящем пособии вопросы для самокон-

троля приведены в конце каждой главы. Материал, необходимый для ответа на эти вопросы, выделен в тексте пособия курсивом.

Второй особенностью обучения в вузе является то, что студент за время учебы должен освоить значительно больше дисциплин, чем в школе за 10 лет. В учебном плане эти дисциплины распределены по четырем блокам, наименования которых подчеркивают содержательное различие отнесенных к ним дисциплин. Это очевидное различие мешает неискушенному студенту осознать менее очевидное, но чрезвычайно важное единство всех дисциплин учебного плана. Единство дисциплин связано с двумя обстоятельствами.

Во-первых, *все дисциплины учебного плана образуют единую систему курсов, внутренне согласованных друг с другом так, что вместе они позволяют подготовить высококвалифицированного и широкообразованного специалиста.* Принципиальная основа этой внутренней согласованности учебных дисциплин определяется Федеральным государственным образовательным стандартом (ФГОСом), а практическая реализация обеспечивается рабочими программами дисциплин.

Во-вторых, *почти все учебные дисциплины имеют общие фундаментальные основы, так как отражают различные стороны одной и той же объективной реальности.*

Документами, которые определяют описанные выше особенности обучения студента в вузе, являются ФГОС, учебный план и рабочие программы дисциплин. Эти документы рассмотрены в третьей главе пособия.

Отметим также, что, в отличие от обучения в средней школе, *учеба в вузе – это фактически начало профессиональной деятельности человека.* Во-первых, студент осваивает содержательную основу будущей профессиональной деятельности, а во-вторых, интеллектуально созревает как специалист, занимаясь конкретными исследованиями. И то, и другое обеспечивается всеми дисциплинами учебного плана. Кроме того, *успех в профессиональной карьере специалиста зависит и от способности*

студента самостоятельно работать с профессионально значимой литературой вне учебного плана.

Вопросы для самоконтроля

1. Что такое фотоника и оптоинформатика? Каковы особенности этой дисциплины?
2. Какие государственные задачи решает высшее техническое образование?
3. Какие требования к подготовке современного дипломированного специалиста соответствуют новой образовательной парадигме?
4. Какие науки и знания относятся к фундаментальным?
5. В чем состоят основные особенности обучения в вузе?
6. По какому признаку распределены дисциплины в учебном плане?
7. В чем состоят различия и единство дисциплин учебного плана?
8. Какие основные документы регламентируют учебный процесс в вузе?

2. ФУНДАМЕНТАЛЬНЫЕ ОСНОВЫ ТВОРЧЕСКОЙ ДЕЯТЕЛЬНОСТИ

2.1. Направления творческой деятельности

В соответствии с доктриной двухступенчатого высшего образования направление бакалаврской подготовки предполагает дальнейшее обучение в магистратуре. Но и этот этап формирования специалиста является промежуточным. В течение всей творческой деятельности, назовем ее инженерной, приходится осваивать новые знания.

Принципиально различающихся в своей естественно-научной основе направлений инженерной деятельности только шесть. Они связаны с *информацией, энергией, сырьем, материалами,*

изделиями и транспортом. Каждое направление подразумевает соответствующее производство: производство информации, энергии и т.д. Любое из названных производств рассматривается здесь в самом широком смысле. Производство информации включает получение, обработку и передачу информации во всех сферах деятельности человека. Производство энергии включает преобразование любых видов энергии (от механической до внутриядерной энергии) в виды, непосредственно применяемые на производстве и в быту, например в механическую или электрическую. Производство сырья включает сырье, связанное не только с геологическими, но и биологическими и другими источниками. Материалы – результат придания сырью свойств, которые обеспечивают получение из него тех или иных изделий. К изделиям отнесено то, что производится из материалов – от пуговицы до сложнейшей автоматизированной системы управления производством, от лопаты до здания, от игрушки до гидроузла и т.д. В понятие «транспорт» включается все, что обеспечивает доставку материалов, изделий и т.д. к месту дальнейшей переработки, потребления или хранения. Сюда относятся любые средства перемещения – от гужевого транспорта до фотонной ракеты.

Составные части приведенного перечня инженерной деятельности связаны между собой, так как без информации невозможно включить знания и опыт в производственную сферу, без энергии производство теряет свою естественную движущую силу, без сырья невозможно производство материалов и т.д.

Их взаимозависимость проявляется в том, что никакая из них не может существовать вне связи со всеми остальными. Это проявляется, например, в любых производственных проектах и бизнес-планах, которые обязательно учитывают издержки на информационное сопровождение, энергоснабжение, приобретение материалов и т.д. Описанные взаимосвязи наиболее ярко проявляются тогда, когда на их основе возникает конкретная инженерная специальность. Ее становление сопряжено с использованием знаний, касающихся сразу нескольких или даже всех направлений прак-

тической деятельности. Это находит отражение в учебных планах любого направления. Например, в учебном плане направления «Фотоника и оптоинформатика» имеются дисциплины, связанные с информатикой, энергетическими вопросами, материаловедением и т.д.

Число направлений подготовки, обеспечивающих инженерную деятельность, велико. Их перечень постоянно изменяется – некоторые исчезают, появляются новые. Эти изменения отражают общий научно-технический прогресс. В настоящее время технические вузы страны осуществляют подготовку инженеров примерно по 80 направлениям.

При подготовке бакалавров взаимосвязи между направлениями практической деятельности учитываются различными дисциплинами учебных блоков, в равной мере обогащающими интеллект будущего специалиста знаниями, формирующими творческое мышление высококвалифицированного специалиста. Поэтому нельзя в процессе учебы делить дисциплины на важные и неважные, как это часто делают неискушенные студенты и недостаточно опытные преподаватели. Здесь ситуация вполне аналогична методам подготовки спортсмена. Спортсменом высокого класса невозможно стать, не закалив волю, не «накачав» все группы мышц, не укрепив дыхательный аппарат, сердечно-сосудистую систему, общую выносливость организма и психологическую устойчивость. В профессиональном спорте не делят тренировки на важные и неважные. Точно так же, обучаясь в вузе, следует всесторонне «накачивать» свой интеллект с помощью всех дисциплин учебного плана.

Практическая деятельность человека связана с природой. *Информация, лежащая в основе всех направлений деятельности человека, извлекается им из природы либо непосредственно (естественными науками), либо опосредованно – путем получения вторичной информации (из первичной, фундаментальной), посредством прикладных наук и инженерной практики.*

Энергия связана с природой как общий признак, общая мера различных форм движения и взаимодействия материальных объектов.

Любое сырье есть природное вещество.

Материалы – те же вещества, преобразованные к виду, более удобному для непосредственного практического использования. Причем преобразуются они в технологических процессах, подчиняющихся законам фундаментальных наук.

Любые изделия, а также самые сложные и хитроумные устройства являются лишь комбинациями конструктивных элементов, выполненных из материалов с использованием процессов, подчиняющихся законам природы.

Транспорт использует принципы перемещения тел в пространстве, разрешенные и определяемые физическими законами.

Таким образом, все направления инженерной деятельности либо копируют природу, либо, если и создают новое, то лишь в рамках допустимого законами природы. Поэтому не существует технических специальностей, не опирающихся на законы фундаментальных наук. В частности, все дисциплины учебного плана, формирующие специалиста данного профиля, имеют общие фундаментальные основы.

К сожалению, общность фундаментальных основ дисциплин учебного плана далеко не всегда осознается студентами. Это затрудняет формирование у студента системного мышления и приобретение им широких профессиональных знаний.

2.2. Практическая деятельность человека и современное естествознание

Практическая деятельность человека многогранна. В данном разделе рассматривается лишь ее часть, относящаяся непосредственно к производству.

Современное естествознание создало научно обоснованную картину мира, которая включает материальные структуры, воз-

никшие в процессе эволюции Вселенной. *Магистральный путь эволюции – переход от простого к более сложному. Один из ее основных механизмов – самоорганизация материальных структур.* На некотором этапе эволюции природа проявляет способность создавать биологические объекты, т.е. материальные структуры, для которых характерны избирательный обмен веществом и энергией с внешним окружением, внутренняя саморегуляция, самовоспроизводство, эволюционное самосовершенствование и адаптация к окружающей среде. В процессе эволюции адаптивные способности таких структур преобразовались в сложную информационную систему сбора, переработки и рационального использования важной для жизни информации – возникла эффективно действующая нервная система. *В условиях Земли эволюция нервной системы живых организмов привела к появлению человеческого мозга и разума.* Возможности человеческого разума выходят за рамки простых потребностей поддержания жизни. Максимально ярко эти возможности проявляются в абстрактном мышлении, которое позволяет человеку познавать окружающий мир, формируя научные представления о нем. Таким образом, *человек оказался той материальной структурой, посредством которой природа проявляет свою способность к самопознанию и к осознанному самосовершенствованию.* Человек, развивая фундаментальные науки, познает законы природы и в пределах «разрешенного» этими законами создает элементы искусственной природы, призванные улучшать качество жизни и обеспечивать ему все новые возможности для более глубокого ее познания. К элементам искусственной природы относятся предметы быта, различные технические устройства, механизмы, сооружения и т.д., предназначенные для удовлетворения потребностей человека во всех сферах его жизнедеятельности. Иными словами, к искусственной природе относится все то, что производится человеком в направлениях практической деятельности.

Эволюция искусственной природы, опирающаяся на разум, науку и инженерную деятельность, также идет от простого к сложному. После удовлетворения элементарных потребностей у человека неизбежно возникают новые, более сложные. Например, потребность дистанционного общения, которая первоначально реализовывалась передачей информации мимикой, жестами, акустическими сигналами, стала затем обеспечиваться почтовыми сообщениями, телеграфом, телефоном и, наконец, бурно развивающимися в настоящее время сложными глобальными электронными системами связи с использованием широкого спектра электромагнитных волн, искусственных спутников Земли, Интернета.

Потребности общества расширяются – вплоть до желания создавать системы, подобные самому человеку как в «техническом» (роботы), так и в «интеллектуальном» (искусственный разум) отношении. В конечном счете все, что создает себе в помощь человек, должно (в области своего применения) обладать большими возможностями, чем сам человек. Здесь нет никаких принципиальных ограничений. Иными словами, в своей практической деятельности человек (в рамках законов природы) всемогущ. Он выступает в качестве творца нового направления эволюции Вселенной, создавая искусственную природу. Собственно, к этому и сводится вся его практическая и научная деятельность.

Эволюция естественной природы проявляется в виде эволюции Вселенной. В процессе эволюции природа действует «вслепую», методом проб и ошибок. Она случайным образом создает любые «разрешенные» законами природы материальные структуры. «Выживают» из них лишь наиболее стабильные в данных условиях. К таким структурам относятся, в частности, те, для которых характерна максимальная энергетическая выгодность. Все «выжившие» к настоящему времени материальные структуры и составляют современную естественную природу.

Только на стадии возникновения мыслящей материальной структуры начинает действовать новая движущая сила эволюции – творец, проектирующий каждый отдельный акт развития

искусственной природы и осуществляющий свои проекты. В этом – суть практической деятельности человека и, в частности, суть деятельности инженера. Следовательно, практическая деятельность человека есть результат эволюции Вселенной. Решающим фактором в этой деятельности является ее естественно-научная основа, которая охватывается современным естествознанием, что в равной степени относится и к инженерной деятельности в области фотоники и оптоинформатики.

Эти рассуждения станут еще более убедительными, если ознакомиться с основными элементами современной научной картины мира, включающей представления о Вселенной и ее эволюции.

2.3. Естественно-научные основы практической деятельности человека

Человек распространил свои познания Вселенной до расстояний 10^{26} м, проник в микромир до размеров 10^{-18} м и углубился в прошлое при изучении эволюции окружающего мира на 14 млрд лет. Выяснилось, что все и всегда во Вселенной сводилось к трем сущностям: материальные системы (объекты), взаимодействие объектов и их движение (развитие). Ничего иного в мире не обнаружено.

Вселенная представляет собой единую, целостную, необозримо сложную материальную систему. По этому поводу поэт сказал: «Все сущее во все века без счета верст невидимый связует мост, и не сорвать тебе цветка, не стронув звезд» (Френсис Томпсон, XIX в.). Описанная взаимосвязь затрагивает любые формы проявления материи, включая духовную, культурно-гуманитарную, социальную и т.п. сферы. Не является исключением и сфера инженерной деятельности, а также подготовка к этой деятельности, т.е. обучение в вузе. Вот почему кажущаяся разнородность учебных курсов относится лишь к частностям; на самом деле практически все дисциплины обладают глубоким естественно-научным единством. Это единство проявляется во многих аспектах.

Начнем с общего для всех научных, технических и гуманитарных сфер подхода к изучению реальности, который называется моделированием. Попытка изучать Вселенную как единую, целостную материальную систему, какой она в действительности является, бесперспективна. Поэтому во всех науках любые сложные системы и процессы стараются разложить на простейшие составляющие и каждую изучать в отдельности. Но и на этом пути сохраняется непреодолимый для познания реальности уровень сложности материальных систем, если не заменить каждый выделенный объект (процесс, взаимодействие) его идеализированной копией – так называемой моделью. *Моделирование – универсальный метод познания во всех сферах научной, инженерной, социальной, экономической и гуманитарной деятельности человека, а также в образовании.*

В ходе любой познавательной деятельности человек создает модели объектов и процессов природы, а затем последовательно совершенствует эти модели с целью все более полного и точного отражения реальности. Описанный процесс познания бесконечен и постепенно приближает нас к постижению абсолютной истины.

Моделирование возникло на поздней стадии эволюции мозга с появлением абстрактного мышления и впервые стало массово применяться в физике. Поэтому в одном из определений физики утверждается, что физика – это искусство моделирования. По мнению академика Н.Н. Моисева, «ничего другого, по своей целостности и логике сравнимого с системой моделей в физике, человечество еще не придумало». Моделирование широко применяется и в других науках, а также в инженерном деле. Методы моделирования явно или опосредованно рассматриваются и используются практически во всех дисциплинах. Поэтому *идея моделирования является для всех дисциплин учебного плана мощным объединяющим фактором и создает одну из важнейших методологических основ интеграции гуманитарной, естественно-научной и профессиональной составляющих инженерного образования.*

Мысленно разделить окружающий нас целостный мир на некие части и изучать их путем построения соответствующих моделей можно по-разному. Однако *объективная реальность такова, что сама «навязывает» наукам о природе логически непротиворечивую стройную иерархическую систему моделей материальных структур.* Именно эта система и определяет разделение науки на отдельные направления и их неразрывную естественно-научную взаимосвязь. В этом еще одна причина естественно-научной общности учебных дисциплин.

Иерархическая система материальных структур полно отображает устройство нашей Вселенной в том приближении, на которое способны фундаментальные науки в настоящее время.

На рис. 2.1 представлен блок неорганических материальных структур. *Каждая следующая материальная структура в представленной системе включает предыдущие. Однако свойства более сложной структуры качественно отличаются от свойств составляющих ее частей или свойств простой их совокупности.* Например, атом обладает химическими свойствами (валентностью) в отличие от его составляющих – ядра и электронов, которым валентность не присуща. Макротела обладают теплопроводностью, электропроводностью, твердостью в отличие от отдельных атомов, из которых состоят макротела и т.д.

Каждой материальной структуре соответствует своя частная фундаментальная наука (физика, химия, геология и т.д.) или несколько таких наук. Значительная часть практической деятельности человека связана с макротелами, но используются также поля, потоки частиц, ядра атомов, атомы, молекулы и тела, имеющие размеры порядка 1...100 нм. На этом основаны новейшие технологии: плазменная, лазерная, нанотехнологии и т.д. Например, в ультрасовременных лазерных технологиях используются потоки фотонов, возбужденных квантовыми системами – атомами, молекулами активной среды, в которой преобразуются различные виды энергии в энергию когерентного электромагнитного излучения оптического диапазона.

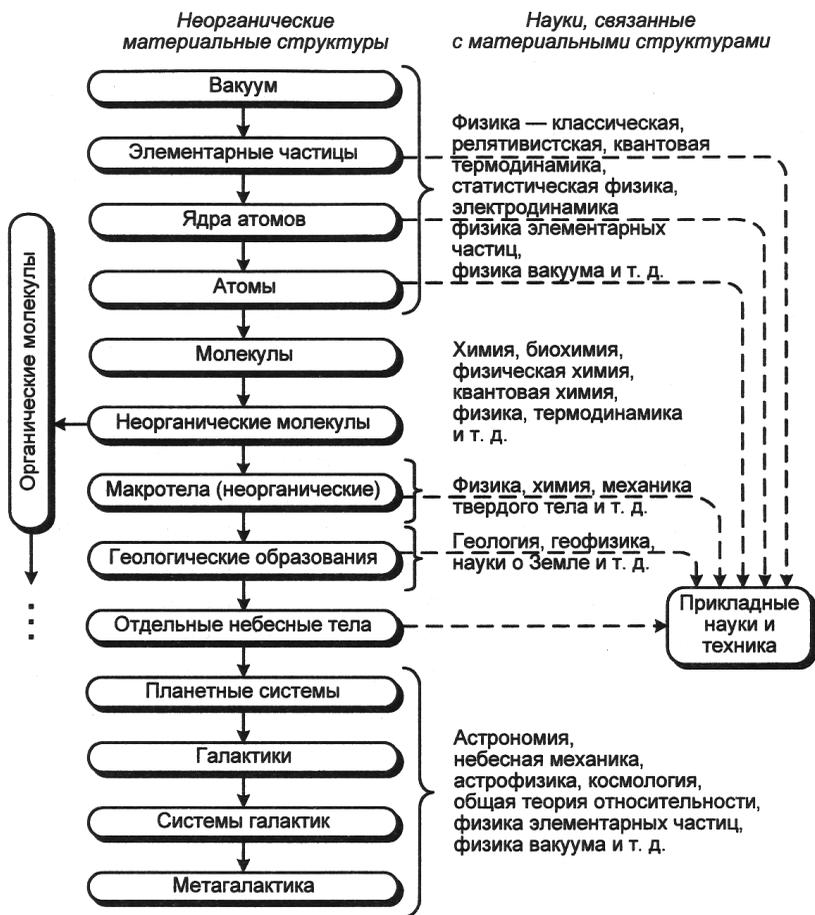


Рис. 2.1. Иерархическая система неорганических материальных структур

Таким образом, почти все материальные системы находят техническое применение. Исключением пока являются физический вакуум и космические объекты. Однако изучение вакуума и космических объектов позволяет глубже познавать свойства многих других материальных структур и процессов, опосредованно влияя на развитие практики. Достаточно упомянуть, что

первоначально термоядерные процессы в макромасштабах были обнаружены в космических объектах (звездах) и лишь позднее стали широко использоваться в земных условиях (в ядерной энергетике).

Изложенное отображено на рис. 2.1 прямоугольником «Прикладные науки и техника», а пунктирными линиями указаны связи различных направлений техники с фундаментальными науками, т.е. с общей естественно-научной основой.

Все направления практической деятельности человека связаны с фундаментальными науками, как кисти винограда с лозой. Кисть может зародиться лишь там, где возникнет почка, что соответствует месту зарождения в недрах науки практически значимого результата. Если кисть символизирует отдельное направление практической деятельности, то ягоды – отдельные разделы этого направления; в образовании это – отдельные направления, специальности и специализации. Биохимические процессы в ягодах, за счет которых они созревают, – аналог механизмов собственного развития техники (рационализация, изобретательство, проектирование, разработки инженерных наук).

Прекращение подпитки новыми фундаментальными знаниями неизбежно приведет к остановке в развитии данной отрасли. Так, развитие ламповой электроники, исчерпав к середине XX в. основной естественно-научный ресурс, практически прекратилось. Лишь открытия в области физики полупроводниковых кристаллов обеспечили электронике новый импульс развития. Так возникло современное техническое направление – микро-, а затем и нанoeлектроника.

На рис. 2.2 представлен блок иерархической системы материальных структур, зародившихся в процессе самоорганизации больших совокупностей органических молекул. На рисунке видно, что науки о живом также связаны с объективной реальностью, т.е. имеют естественно-научную основу. Не являются исключением и сфера культуры, все виды искусства, религии и т.д.

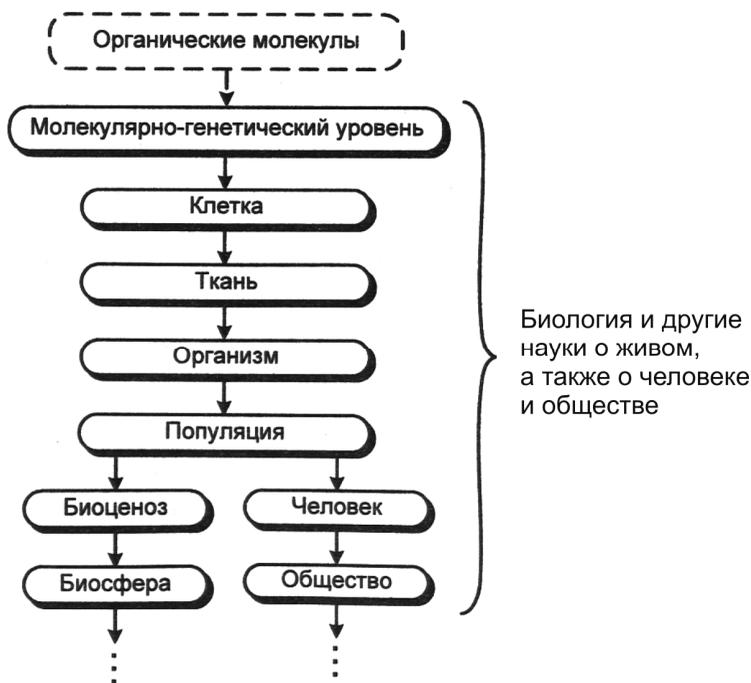


Рис. 2.2. Иерархическая система органических материальных структур

На рис. 2.3 представлен блок иерархической системы материальных структур, связанных с деятельностью человека. В эту иерархическую систему включено все, что создано человеком за всю историю в культурной, научной и технической сферах. Конечно, строгую систему структур искусственной природы создать невозможно, и на рис. 2.3 обозначены только отдельные направления основ функционирования различных производств. Так, в основу производства энергии могут быть положены: «мышечная» энергия, гидроэнергия, тепловая, ядерная, солнечная и другие виды.

Представленные на рис. 2.1, 2.2 и 2.3 блоки иерархий неживого, живого и мыслящего и созданного человеком образуют основу всего, с чем человек сталкивается в науке, культуре и технике.

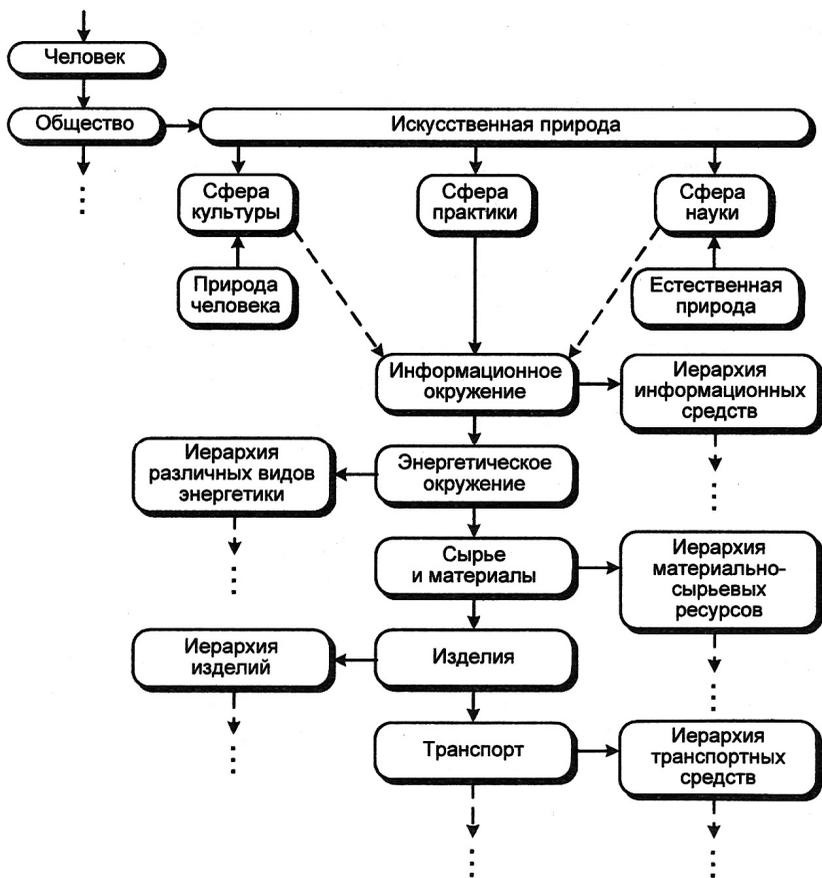


Рис. 2.3. Иерархическая система искусственных материальных структур

Естественно-научные основы практики ярко представлены эволюцией Вселенной, из которой следует ее единство и целостность, а также единство законов, которым подчиняются любые процессы в неживой и живой природе.

2.4. Эволюция Вселенной и общность законов природы

Все рассмотренные материальные структуры возникли в процессе эволюции Вселенной, подчиняющейся единым законам природы. В пределах разрешенного этими законами протекает и практическая деятельность человека. *Законы природы – еще один общий естественно-научный стержень всей практической деятельности человека.*

Рассмотрим лишь некоторые основные этапы эволюции Вселенной, отраженные на рис. 2.4 и связанные с возникновением структур неживой материи. По оси ординат отложено время t (от момента зарождения Вселенной) сначала в секундах, а затем в годах вплоть до настоящего времени (~ 14 млрд лет). Масштаб искажен таким образом, чтобы было удобно описывать происходящее на различных этапах эволюции. По оси абсцисс от точки «0» откладывается вправо и влево радиус Вселенной R (в метрах). Взаимное соответствие временных и геометрических характеристик выдерживается только для момента возникновения ($R = 0$) и для настоящего времени ($R = 10^{26}$ м); сплошные кривые, соединяющие начальный и конечный размеры Вселенной, произвольны и лишь условно отображают ее постоянное расширение. Временной уровень AA' соответствует современному периоду существования Вселенной, а стрелка на прямой AA' указывает направление увеличения константы взаимодействия, характеризующей относительную «силу» фундаментальных взаимодействий. Например, электромагнитное взаимодействие (константа взаимодействия – 10^{-2}) значительно «сильнее» гравитационного (10^{-39}) и «слабее» ядерного (10). Вдоль левой вертикальной пунктирной прямой отложены значения температуры, которую Вселенная имела в соответствующие моменты своей эволюции.

В соответствии с этой схемой отметим основные этапы эволюции Вселенной, описываемые *теорией Большого взрыва*. Эту теорию, основанную на модели горячей Вселенной, дополняет

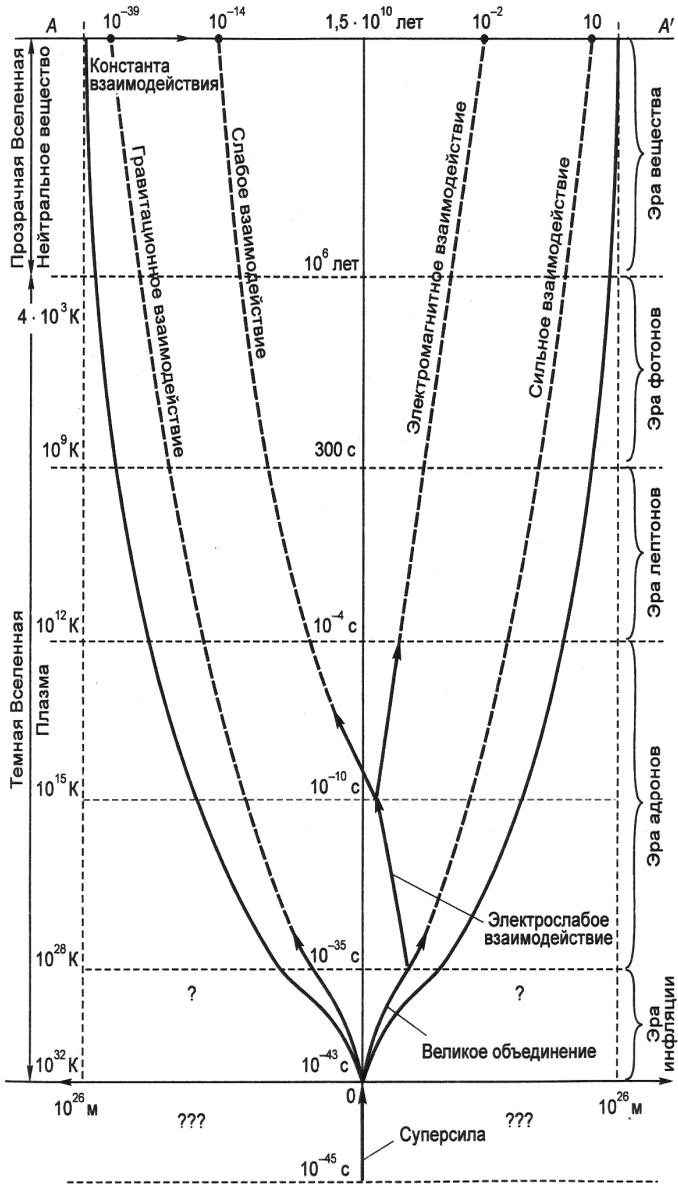


Рис. 2.4. Условная схема эволюции Вселенной

теория инфляции (раздувания) Вселенной. В теории инфляции возникновение Вселенной связывают с физическими процессами, происходящими в вакууме; она объединяет процессы в микро- и мегамирах. По современным представлениям, вакуум – сложная все заполняющая динамическая квантовомеханическая система, постоянно и повсеместно порождающая виртуальные частицы и античастицы. Некие весьма сложные процессы в вакууме, названные инфляционными, лежат в основе зарождения Вселенной.

Вселенная возникла примерно 14 млрд лет назад из первоначально чрезвычайно малого объема и стала очень быстро расширяться. Этот период в истории Вселенной называется эрой инфляции, продолжающейся от 10^{-43} до 10^{-35} с. Вопросительные знаки на рис. 2.4 (внизу) отражают отсутствие для этого периода эволюции общепризнанной теории. Весьма вероятно, что до эры инфляции существовал лишь один вид фундаментального взаимодействия между элементарными частицами, заполнявшими Вселенную. Это взаимодействие названо Суперсилой. К моменту 10^{-43} с от Суперсилы отделилась гравитация (левая пунктирная кривая на рис. 2.4).

Гравитация существует в качестве отдельной фундаментальной силы и в наше время. Оставшиеся объединенными три взаимодействия (сильное, электромагнитное и слабое) получили название Великого объединения.

К моменту 10^{-35} с закончилась эра инфляции, расширение замедлилось, начался период, описываемый теорией Большого взрыва.

Великое объединение к моменту 10^{-35} с разделилось на сильное (ядерное) взаимодействие (правая пунктирная кривая на рис. 2.4) и на электрослабое. Ядерное взаимодействие существует и в наше время.

К моменту 10^{-10} с электрослабое взаимодействие распалось на слабое и электромагнитное. Это утверждение уже получило строгое теоретическое обоснование.

Итак, в процессе эволюции Вселенной единое взаимодействие (Суперсила) распалось на четыре фундаментальных взаимодействия. Слабое взаимодействие пока не находит использования в технике. Все силы, учитываемые в инженерном деле, сводятся к трем фундаментальным видам взаимодействия: гравитационному, электромагнитному и ядерному. *Универсальность фундаментальных сил, используемых в практической деятельности человека, – одна из основ естественно-научного единства фундаментальных и инженерных дисциплин.*

Рассмотрим, как в процессе эволюции Вселенной формировались материальные структуры неживой природы.

В эру инфляции (10^{-43} – 10^{-35} с) возникла первопричина появления в будущем всего вещественного. Дело в том, что в ту эру (рис. 2.5) при температурах 10^{32} – 10^{28} К материя существовала в виде излучения, элементарных частиц и античастиц. На последующих этапах эволюции частицы и античастицы попарно аннигилировали (взаимно уничтожились), превращаясь в излучение. Так как античастиц было на 10^{-7} % меньше, чем частиц, то некоторая доля частиц (10^{-9} % от исходного их количества) сохранилась до наших дней, обеспечив образование современной вещественной природы. Античастицы аннигилировали практически полностью, поэтому антивещества в космических масштабах в природе нет.

Затем (10^{-35} – 10^{-4} с) наступает эра адронов, дальнейшее расширение Вселенной происходит уже в соответствии с теорией Большого взрыва. Температура Вселенной убывает от 10^{28} до 10^{12} К. В конце этой эры кварки объединились в адроны, к которым относятся, в частности, протоны и нейтроны. Так образовались частицы, из которых состоят ядра атомов. Однако в эру адронов объединения нейтронов и протонов в ядра атомов произойти не могло, так как температура еще была слишком высокой ($T > 10^{12}$ К).

В эру лептонов (10^{-4} –300 с) происходит реакция между протонами и нейтронами, в результате которой отношение чис-

ла оставшихся нейтронов к числу протонов стало равным примерно 0,15. К моменту 300 с температура расширяющейся Вселенной упала до 10^9 К, и возникли условия для образования ядер изотопа водорода (${}^2_1\text{D}$) и гелия (${}^3_2\text{He}$, ${}^4_2\text{He}$). Но это уже начало эры фотонов, которая длилась примерно от 300 с до 10^6 лет. До момента 10^4 лет нейтроны были израсходованы полностью на образование гелия (~25 %). Оставшиеся протоны проявились в дальнейшем как ядра атомов водорода (~75 %). Примерно это же соотношение между содержанием гелия и водорода сохранилось в среднем во Вселенной и в наше время.

После образования ядер легких элементов (~300 с) вещество еще длительное время (~ 10^6 лет) представляло собой плазму. Высокая температура не позволяла существовать нейтральным атомам. После снижения температуры примерно до 4000 К электроны стали удерживаться вблизи ядер, образуя атомы водорода и гелия. Активность взаимодействия фотонов с веществом ослабла. Вселенная, ранее непрозрачная, стала прозрачной. Началась эра вещества (эра звезд), продолжающаяся и сейчас.

Изложенное показывает, что в эру адронов, лептонов и фотонов, т.е. до эры вещества, материальные системы усложнились и разнообразились путем объединения элементарных составляющих во все более сложные конструкции. Аналогичный подход в практической деятельности человека называется технологией по принципу «снизу вверх» (создание сложного из более простых частей). В химии так получают молекулы из атомов, различные материалы – из молекул; механики таким же образом создают сложные изделия из отдельных конструкционных элементов; электронщики получают сложные электронные системы из схемных элементов и т.д.

В начале эры вещества Вселенная состояла из газообразного водорода и гелия, почти равномерно распределенных в пространстве. Вокруг первоначально весьма слабых неоднородностей вещества стали образовываться (за счет гравитационных сил) огромные газовые сгущения. Со временем они превра-

тились в галактики, которые, в свою очередь, распались на протозвезды. Сжимаясь, протозвезды нагревались до возникновения собственного свечения и таким образом становились звездами.

В итоге первоначально сплошная, почти однородная и предельно хаотизированная плазменная среда Вселенной превратилась за счет гравитационного (и в меньшей степени – электромагнитного) взаимодействия в обособленные звездные скопления – галактики. Механическое взаимодействие материальных частей, ускоряющихся при гравитационном стягивании к центрам галактик и звезд, трансформировалось в поступательное и вращательное движение галактик, звезд и других небесных тел. Возникли новые процессы самоорганизации, которые привели сначала к ядерному превращению звездного водорода в гелий. Затем следовало превращение в более тяжелые элементы (вплоть до железа), а затем и в еще более тяжелые элементы (вплоть до урана) – при взрывах звезд, находящихся в неустойчивом состоянии («взрывы сверхновых»). Так появился «строительный материал» для возникновения вещества.

Таким образом, в начале эры вещества новые структуры (скопления галактик, галактики, звезды и другие небесные тела) возникали из более обширных образований и превращались в их части. В инженерном творчестве подобный процесс связывается с технологией по принципу «сверху вниз». Так изготавливают канцелярские кнопки из цельного листа металла, элементы интегральных микросхем в поверхностном слое монолитного полупроводникового кристалла и т.д.

Изучение Вселенной показывает, что фундаментальные законы природы одинаковы во всей Вселенной и на всех этапах ее эволюции. Те же законы лежат в основе техники и обеспечивают научно-технический прогресс. *Универсальность и единство законов природы определяют также естественно-научную общность фундаментальных основ всех дисциплин учебного плана.*

2.5. Деятельность специалиста и реальность

Естественная природа развивается на основе процессов самоорганизации. С этой точки зрения человек предстает как особая движущая сила самоорганизации, проявляющаяся на определенном уровне развития материальных структур. С его появлением связано создание и развитие искусственной природы. У истоков искусственной природы стоит человек, ставящий и реализующий определенные цели. В области производства – это инженер. Связи объективной и субъективной реальностей с инженерной деятельностью представлены на схеме (см. рис. 2.5).

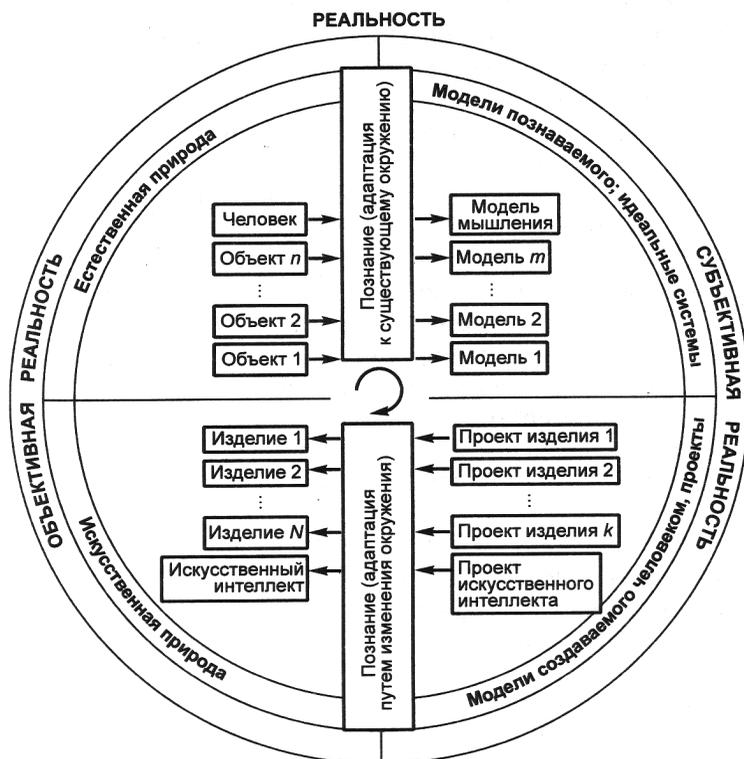


Рис. 2.5. Связи объективной и субъективной реальностей с инженерной деятельностью

Реальность подразделяется на объективную (иногда ее называют действительностью) и субъективную. Объективную реальность можно разделить на естественную (верхняя левая часть схемы) и искусственную (нижняя левая часть).

Естественная объективная реальность – это весь мир самопроизвольно возникших материальных систем, взаимодействий, движений. В своей видимой части – это метагалактика (наша Вселенная). Назовем для краткости все, что «населяет» естественную объективную реальность, объектами. Термином «объект» здесь обозначены все три материальные сущности: объект, взаимодействие, движение.

Наша Вселенная непрерывно эволюционирует. Это проявляется в возрастании разнообразия и сложности материальных структур. В процессе эволюции сложилась иерархия в области неживых объектов (см. рис. 2.1). Ветвь эволюции в направлении от простых неорганических к сложным органическим системам (см. рис. 2.2) привела к образованию мыслящих материальных систем; на Земле – это человек. На схеме, изображенной на рис. 2.5, «мыслящий объект» обозначен верхним прямоугольником слева.

Любая материальная система отражает изменением своего состояния наличие и свойства других материальных систем, с которыми она взаимодействует. Строго говоря, в мире все взаимодействует со всем. *Поэтому отражение является всеобщим свойством материи, а мышление, возникшее в процессе эволюции материи, представляется высшей формой проявления этого свойства.*

Посредством мышления происходит самопознание естественной объективной реальности. В процессе познания наше мышление создает некие образы материальных объектов. *Мысленный образ объекта – это его модель. Субъективная реальность «заселена», в частности, моделями материальных объектов.* Каждому объекту может быть сопоставлено несколько моделей. Например, при развитии наших представлений об атоме использовались мо-

дели Демокрита, Томсона, Резерфорда, Бора и квантово-механическая модель. Поэтому субъективный мир «заселен» более плотно, чем известный естественный мир.

Модели объектов естественного мира возникают в процессе познания природы фундаментальными науками, искусством, религиями, бытовым мышлением. Правая верхняя часть схемы на рис. 2.5 содержит наши представления о том, что существует в естественном объективном мире в настоящее время и что существовало в прошлом. Прошлое моделируют история, теория эволюции Вселенной и другие науки. Для практической деятельности (в частности, инженерной) важно умение домысливать нечто дополнительное к существующему и создавать модели желаемого будущего. Например, в экономике это бизнес-план, в военном деле – тактическая разработка боевой операции, в инженерии – проект нового изделия.

Инженерные проекты, т.е. мысленные образы новых устройств, изделий, механизмов, технологических процессов, занимают свою часть субъективной реальности. Проекты могут материализоваться в схемах, чертежах, вещественных копиях будущих изделий и т.д. Однако во всех случаях суть проекта заключена в инженерной мысли. На рис. 2.5 проекты, как модели будущего, «заселяют» ту часть субъективной реальности, которая представлена справа снизу. Проекты порождаются инженерным творчеством на основе критики недостатков существующего и разработки путей их преодоления.

Субъективная и объективная реальности соприкасаются, во-первых, в процессе познания человеком природы (вертикальный прямоугольник в верхней части схемы) и, во-вторых, в процессе производственной деятельности инженера (прямоугольник в нижней части схемы на рис. 2.5). Производственная деятельность инженера порождает искусственную объективную реальность (искусственную природу) при воплощении проектов в изделия (левая нижняя часть схемы). В настоящее время искусственная природа «заселена» громадным количеством изде-

лий, и их число постоянно возрастает в результате инженерной деятельности. Они постепенно вытесняют объекты естественной природы из непосредственного жизнеобеспечивающего окружения человека.

Истоки инженерной деятельности находятся в недрах естественной объективной реальности, в фундаментальных законах природы. Ни один проект не может быть осуществлен, если он противоречит этим законам. Создание моделей в процессе познания естественной объективной реальности позволяет человеку осознать, каким образом можно лучше адаптироваться к окружающему миру и видоизменить его с целью повышения качества своей жизни. Такое осознание возникает в той части субъективной реальности, которая является продуктом фундаментальных наук. На следующей стадии творческой деятельности человека (при создании моделей желаемого будущего) абстрактная познавательная деятельность превращается в конкретную созидательную. Зарождается новый этап эволюции природы – этап создания искусственной объективной реальности.

Сравнение особенностей развития естественной и искусственной природы показывает, что эти два процесса имеют как существенно различающиеся, так и схожие признаки.

И инженер, и природа могут создавать только то, что не противоречит законам естествознания. И в первом, и во втором случае сохраняется лишь то, что выдерживает конкурентный отбор. Общее направление развития – «от простого к более сложному». Ограничений в создании интеллектуальных материальных структур у человека, как и у природы, нет. Разница в том, что природа в этом уже преуспела, а человек находится лишь в начале пути. Перечисленное – основа для оптимизма человека как творца своего будущего, источник созидательного порыва для новых поколений инженеров. Фундаментальные науки демонстрируют будущему специалисту неисчерпаемое обилие примеров для подражания в недрах естественной природы. Созданное природой (особенно на этапе биологической эво-

люции) по многим показателям опережает созданное человеком. Достаточно вспомнить сложнейшие биологические информационные системы (нервная, гормональная, генетическая), возникшие на Земле более 3 млрд лет назад. Инструментальная информатика пока не может создавать подобные системы.

Осмыслить различия в особенностях эволюции естественной и искусственной природы также важно для будущего инженера. В природе отсутствует феномен морального и физического износа. Переход от простых материальных структур к более сложным структурам не превращает простое в ненужные отходы эволюции. Все возникшее на различных этапах эволюции входит в состав Вселенной как неотъемлемые части единой целостной системы. Напротив, все созданное человеком устаревает, перестает использоваться, превращается в постоянно накапливающиеся отходы искусственной эволюции. К этим отходам добавляются изделия разового использования, а также неизбежные побочные продукты технологических процессов. Важнейшим отходом становится и выделяемая в процессе деятельности человечества энергия.

В итоге возникает техногенная перегрузка сферы жизнедеятельности человека бесполезными или даже вредными отходами искусственной эволюции. На планете Земля допустимый уровень указанной перегрузки имеет определенный критический предел, и этот предел фактически достигнут. Поэтому инженер XXI века, создавая новое, должен способствовать снижению количества техногенных отходов.

Во Вселенной нет процессов, приводящих к общим потерям энергии, происходят лишь процессы ее перераспределения и взаимного превращения из одного вида в другой. Человек вынужден пока использовать в основном ограниченные невозобновляемые энергетические ресурсы. Поэтому современный инженер в любой сфере деятельности должен максимально снижать энергопотребление и по возможности переходить на возобновляемые источники энергии.

Наконец, при естественной эволюции не обнаруживаются какие-либо пределы, за которыми просматривался бы сценарий саморазрушения и гибели всей системы в целом. Развитие же искусственной природы на Земле имеет критические пределы. Некоторые ученые убеждены, что пределы возможного для эволюции человечества уже преодолены. Если не перейти на оптимальное, научно обоснованное регулирование развития техники и экономики, то в XXI в. нас ожидают глобальные катастрофы (необратимые изменения климата, существенное сокращение видового разнообразия в биосфере, недопустимые потери минеральных ресурсов и т.д.), а вслед за этим – социальные катаклизмы. *Современный инженер, будучи профессионалом в области некоторого конкретного производства, должен учитывать сложную цепь взаимосвязей практической деятельности человека и глобальной стабильности существования человеческого общества на Земле.* Инженерная прослойка в обществе является достаточно активной и многочисленной. От культуры и мировоззренческих установок, преобладающих в сообществе инженеров, во многом зависит формирование рационального типа поведения общества в целом. Очевидно, что узкоспециальное инженерное образование совершенно не отвечает этой социальной роли инженера.

Инженер в своей деятельности часто копирует природу, в которой реализуются два универсальных «подхода» при создании новых систем из уже существующих. *В процессе эволюции природа создает новую систему либо из более простых и компактных элементов (по так называемому принципу «снизу вверх»), либо из более сложных и крупных по принципу «сверху вниз».* По принципу «снизу вверх» возникли, например, молекулы из атомов, макротела из молекул, залежи монолитного песчаника из совокупности песчинок, планеты из мелких небесных тел. По принципу «сверху вниз» образовались галька и песок из монолитных горных пород, капли дождя из однородных объемов водяного пара, звезды (как обособившиеся части галактик).

Аналогичным образом работает и инженер. По принципу «снизу вверх» он синтезирует молекулы из отдельных атомов и вещество в макрообъемах (из молекул). По принципу «сверху вниз» инженер создает из крупного макротела более мелкие изделия. Например, исходным «макротелом» в производстве интегральной микросхемы является монокристаллический слиток кремния. Его разделяют на тонкие пластины, из которых получают небольшие кристаллики – чипы, на которых специальными методами (литография) выделяются еще более мелкие части, являющиеся активными, пассивными или соединительными элементами интегральных микросхем. В итоге материал слитка объемом более 10^4 см^3 используется для создания схемных элементов объемом менее 10^{-14} см^3 . Теперь этот метод совершенствуется и переносится в технологию наноструктур.

В обычных технологиях чаще используется принцип «сверху вниз». В нанотехнологии, вероятно, более широко будет использоваться принцип «снизу вверх», т.е. от атомарных объектов к объектам наноразмеров. При этом особую значимость приобретают процессы самоорганизации атомов в необходимые наноструктуры.

Вопросы для самоконтроля

1. Какие основные направления инженерной деятельности вы знаете?
2. В чем состоит иерархическая связь между различными видами производств?
3. Как связаны различные виды производств с природой?
4. Охарактеризуйте основные направления и механизмы эволюции естественной и искусственной природы.
5. Каков результат эволюции нервной системы живых организмов?
6. Что представляет собой человек как особая материальная система?
7. Чем обусловлена и в чем проявляется взаимосвязь между различными учебными дисциплинами?
8. Опишите иерархические системы материальных структур.

9. Как связаны фундаментальные науки с материальными структурами?

10. Как связаны общие естественно-научные основы различных направлений инженерной деятельности с законами природы?

11. Каким образом возникли химические элементы и структуры иерархической системы неорганической природы?

12. Что такое реальность? Что охватывают понятия «объективная» и «субъективная реальность»?

13. Что такое инженерный проект?

14. Охарактеризуйте сферы соприкосновения объективной и субъективной реальностей.

15. Сравните особенности развития естественной и искусственной природы.

16. Каковы основные источники накопления техногенных отходов жизнедеятельности человека? Каковы задачи инженера в этой сфере?

17. Охарактеризуйте роль инженера в обеспечении глобальной стабильности существования человеческого общества.

18. В чем суть двух основных вариантов возникновения в природе новых систем по принципу «снизу вверх» и «сверху вниз»? Опишите аналогичные варианты технологий в инженерной практике.

3. ФУНДАМЕНТАЛЬНЫЕ ОСНОВЫ ДИСЦИПЛИН УЧЕБНОГО ПЛАНА

3.1. Особенности Федерального государственного образовательного стандарта по направлению «Фотоника и оптоинформатика»

В основе подготовки бакалавров лежит Федеральный государственный образовательный стандарт (ФГОС). Для каждого направления подготовки имеется свой ФГОС. Он определяет:

– нормативный срок подготовки;

- перечень компетенций, которыми должен обладать выпускник;
- требования к структуре основных образовательных программ бакалавриата;
- перечень гуманитарных, социальных и экономических, математических и естественно-научных, профессиональных дисциплин федерального уровня;
- сроки освоения основной образовательной программы;
- требования к условиям реализации основных образовательных программ бакалавриата; учебно-методическому, материально-техническому и кадровому обеспечению учебного процесса, к организации практик;
- требования к профессиональной подготовке выпускника и его итоговой государственной аттестации.

ФГОС определяет также область, объекты и виды профессиональной деятельности бакалавра. Сюда относятся: научно-исследовательская, проектно-конструкторская, производственно-технологическая, экспертная, организационно-управленческая.

Область профессиональной деятельности бакалавров включает науку и технику, связанную с использованием светового излучения (или потока фотонов) в элементах, устройствах и системах, в которых генерируются, усиливаются, модулируются, распространяются и детектируются оптические сигналы; оптические устройства и технологии передачи, приема, обработки, хранения и отображения информации.

При этом объектами профессиональной деятельности могут быть фундаментальные и прикладные научно-исследовательские разработки в области фотоники и оптоинформатики; элементная база, системы и технологии интегральной, волоконной и градиентной оптики; элементная база лазеров; систем, обеспечивающих оптическую передачу, прием, обработку, запись и хранение, преобразование и отображение информации на основе наноразмерных и фотонно-кристаллических структур; оптические компьютеры, системы искусственного интеллекта, устройства на основе когерентной оптики и голографии.

Даже приведенный неполный список того, что определяет ФГОС в подготовке бакалавра, демонстрирует основополагающее значение этого документа. На его базе разрабатываются учебные планы, содержащие помимо дисциплин федерального компонента, региональные дисциплины, дисциплины по выбору студентов.

В итоге учебный план содержит набор дисциплин, распределенных по четырем блокам. Дисциплины учебного плана, за малым исключением, имеют общие фундаментальные основы. Все естественно-научные и технические дисциплины объединены общностью материальных основ и фундаментальных законов природы, которым подчиняются любые материальные процессы. Проиллюстрируем это утверждение на примере курсов МЕН- и ПД-блоков дисциплин, рассматривая только дисциплины федерального компонента.

3.2. Математические и естественно-научные дисциплины

К блоку естественно-научных дисциплин относятся: математика, физика, химия, информатика, экология, квантовая физика, физика твердого тела. Общая естественно-научная основа этих дисциплин очевидна и непосредственным образом отражена в ФГОСе. Например, если рассматривать обобщенно, то действующий ФГОС требует дать студентам в курсах физики, химии и экологии представления о Вселенной в целом, ее эволюции; фундаментальном единстве естественных наук; дискретности и непрерывности; соотношении порядка и беспорядка в природе; динамических и статистических закономерностях; вероятности как объективной характеристике природных систем; принципах симметрии и законах сохранения; соотношении эмпирического (опытного) и теоретического в познании и т.д. По каждой из перечисленных в ФГОСе дисциплин дается также более конкретный перечень специфических законов природы, с которыми должны быть ознакомлены студенты. У дисциплин различных блоков обнару-

живается общность многих законов и моделей, используемых по характерному для каждой дисциплины назначению. Например, в физике и химии рассматриваются модели атомов, молекул и более сложных структур; законы молекулярной физики, термодинамики, активационные процессы, элементы квантовой физики и т.д.

Естественно-научная общность физики, химии и экологии объясняется тем, что все они с различных сторон описывают один и тот же «объект» – природу. Принципиальные различия между перечисленными дисциплинами состоят лишь в том, что каждая из них описывает природу со своих позиций или «свою» составляющую природы.

Для математики и информатики ситуация не столь очевидна. Однако и эти дисциплины по своему происхождению, сути и результатам относятся к естественно-научным. Математика возникла как прикладная наука, непосредственно связанная с окружающей природой и разнообразными видами деятельности человека. Например, в первой известной математической энциклопедии, «изданной» 4000 лет назад в Вавилоне в виде 44 глиняных табличек, содержатся только практические задачи (по земледелию, орошению, торговле и т.п.). В более поздние времена математика превратилась в весьма разветвленную систему крайне абстрактных теорий. Даже сами математики долгое время были убеждены в самодостаточности своей науки и полной независимости ее развития от объективной реальности. Наиболее ярко это убеждение проявилось в абсолютизации аксиоматического подхода при формализованном построении логически замкнутых математических теорий. В этих теориях сначала формулируют ограниченное число основных положений (аксиом), а затем путем строгих математических или логических выводов получают остальное содержание данной теории.

Однако в 30–40-е гг. XX столетия была доказана недостаточность аксиоматического метода (теоремы К. Гёделя). Согласно этим теоремам во всякой формализованной математической системе обнаруживаются утверждения, истинность которых нельзя ни

доказать, ни опровергнуть на основе тех аксиом, которые выбраны для логически непротиворечивого построения данной теории. Остается путь использования иных, не связанных с основополагающими аксиомами, истин. Это могут быть и истины, полученные эмпирическим путем.

Математика связана с познанием природы. Многие ее теории оказываются адекватными моделям, используемым при описании природных процессов. Например, движение небесных тел описывается теорией, в которой используется геометрия конических сечений (окружности, эллипса, параболы). Именно эта математическая теория оказалась адекватной тем физическим явлениям, которые описываются небесной механикой. Таких примеров много. Вот почему математику можно отнести к фундаментальным наукам, которые существенно облегчают познание окружающего нас мира. Она оказывается крайне абстрактным, но весьма полезным отражением реальности. Часто математики «рисуют» правильный формальный образ того, что еще никем не наблюдалось. Например, один тип дифференциальных уравнений оказался точным абстрактным портретом электромагнитных волн, распространяющихся в свободном пространстве, что обнаружилось лишь после открытия и экспериментальных исследований этих волн. Возможность «забегания» математиков вперед в познании реальности связана, вероятно, с тем, что мозг человека есть часть этой реальности и функционирует по ее законам. Поэтому логически безупречные абстрактные построения математиков не могут полностью выходить за рамки реальности.

Аналогичные рассуждения применимы и к информатике. Следует учитывать также, что любые теории, используемые в информатике, в конечном счете оперируют понятием «информация», которое всегда отражает реальность. Сверх того, информатика имеет инструментальную базу, создаваемую и функционирующую на основе фундаментальных законов природы. Например, нанотехнологии получения материалов для хранения

и передачи информации в виде потока фотонов включает различные стадии, опирающиеся главным образом на законы физики и химии. Функционирование оптического компьютера определяется законами квантовой физики, электродинамики, статистической физики, физики твердого тела.

В природе информационные процессы существовали всегда, а примерно 3,8 млрд лет назад, когда на Земле появилась жизнь, возникли и стали совершенствоваться сложные информационные системы: генетическая, нервная, гормональная, функционирующие на основе фундаментальных законов природы. Таким образом, информатика имеет, несомненно, общую естественно-научную основу с остальными дисциплинами рассматриваемого блока. Кроме того, информатика, как и математика, делает процесс познания природы строгим и все более динамичным. Со своей стороны, природа «подсказывает» информатике (и особенно на ее нанoeлектронном уровне) пути дальнейшего развития.

В основе фундаментальной подготовки лежат не только общие законы *физики*, но и ее специальные разделы – *физика твердого тела* и *квантовая физика*.

Итак, естественно-научные основы и фундаментальная общность дисциплин естественно-научного блока очевидны.

3.3. Профессиональные дисциплины

Профессиональных дисциплин более шестнадцати. Практически все они полностью базируются на естественно-научных дисциплинах. Рассмотрим подробно только дисциплины федерального компонента учебного плана. Будем обращать внимание лишь на явные связи с фундаментальными дисциплинами.

Дисциплина «Общая электротехника», всецело базируется на разделе физики «Электромагнетизм». В курсе физики изучают основы и физическое содержание электродинамики, ее связь с другими разделами физики и техникой, а в электротехнике электродинамика используется для разработки методов расчета электрических и магнитных полей в различных электротехниче-

ских устройствах. При разработке частных методов расчета в электротехнике применяют основные законы макроскопической электродинамики (выраженные уравнениями Максвелла) и их следствия (теорема Гаусса, уравнение Пуассона, закон электромагнитной индукции Фарадея, уравнения Кирхгофа и т.д.). Методы расчета усилителей и генераторов гармонических сигналов в значительной степени основаны на физической теории колебаний. Использование того или иного раздела фундаментальной науки при построении соответствующей общепрофессиональной дисциплины – весьма распространенный подход в техническом образовании. Такой подход непосредственным образом отражает фундаментальные основы общепрофессиональных дисциплин. В результате изучения этих дисциплин обучающийся должен знать методы анализа электрических и магнитных цепей постоянного и переменного тока; физические процессы в электрических машинах постоянного и переменного тока, их типы и основные характеристики; физические основы электроники; типовые элементы электроники, микроэлектроники, нанoeлектроники, основы цифровой электроники и микропроцессорной техники; основные компоненты элементной базы современных электронных приборов.

Дисциплина «Метрология, стандартизация и сертификация» построена на основе изучения законодательных актов, использования физических принципов измерений геометрических, электрических, оптических, структурных и других характеристик макро-, микро- и нанообъектов. Методы численной обработки результатов исследований основаны на математической теории случайных процессов. Отметим также, что метрология зародилась и первоначально развивалась в недрах физики, а затем распространилась на другие науки и технику.

Дисциплина «Оптическое материаловедение» представляет собой прикладную версию той части физики твердого тела, которая теоретически «обслуживает» нанотехнологию. Связи «состав–структура–свойства», которые вскрываются в физике твердого тела, позволяют разрабатывать оптимальные техноло-

гические процессы, используемые для получения оптических материалов с заданными эксплуатационными характеристиками, изучать особенности их применения в фотонике и оптоинформатике, основы современных технологий синтеза оптических кристаллов, стёкол и керамик, методы исследования физико-химических свойств оптических материалов. Связь рассматриваемой дисциплины с естественными науками заключается в том, что сам технологический процесс всегда представляет собой сочетание химических и (или) физических процессов.

Дисциплина «Безопасность жизнедеятельности» (БЖД) рассматривает различные процессы в системе «человек – среда обитания»: физиологию труда и безопасность жизнедеятельности; негативные факторы техносферы, их воздействие на человека; критерии безопасности технических систем – отказ, вероятность отказа, качественный и количественный анализ опасностей; средства снижения вредного воздействия технических систем.

Законы одинаковы во всех частях Вселенной и на всех этапах ее долгой эволюции. Указанный факт установлен фундаментальными науками и положен в основу одной из концепций современного естествознания – концепции о единстве законов, управляющих всеми процессами в этом мире. Самоорганизация и саморазвитие любых сложных систем также подчиняется единым законам. Самоорганизация сложных систем обеспечивается проявлением в системе положительных и отрицательных обратных связей. Первые выводят систему из состояния равновесия, а вторые – приближают к нему. Эти же взаимосвязи лежат в основе многих естественных процессов в системе «человек – среда обитания». Следовательно, если не учитывать правовые, нормативно-технические и организационные вопросы, то в остальном и дисциплина БЖД основана на фундаментальных науках.

Дисциплина «Инженерная и компьютерная графика» также базируется на дисциплинах математического и естественно-научного блока – математике и информатике. Обучающийся должен знать основы инженерной графики, задачи

геометрического моделирования, методы и средства компьютерной графики.

В дисциплинах «Теория информации и информационных систем», «Архитектура вычислительных систем» изучаются информационные характеристики источников сообщений и каналов передачи информации, основные понятия теории кодирования информации и особенности основных видов кодов, методы преобразования, обработки и анализа сигналов, принципы организации современных архитектур вычислительных систем. Изучаются принципы построения и функционирования процессора, оперативной памяти и внешних устройств, классификация и основные особенности конвейерных и параллельных архитектур.

Объектом изучения дисциплины «Основы фотоники» являются законы излучения, закономерности высоких интенсивностей излучения, характеристики современных источников излучения, принципы работы и характеристики современных лазеров и усилителей света, физические эффекты, принципы, элементы и устройства для управления светом в оптических материалах и волноводных структурах, устройство, принцип действия и характеристики современных типов фотоприемных устройств фотоники.

В дисциплине «Основы оптоинформатики» рассматриваются основные принципы и технологии передачи информации оптическими методами, физические возможности оптических информационных технологий, современные достижения в области оптоинформатики, принципы и технологии оптической записи, хранения и считывания информации, основные принципы построения фотонно-кристаллических структур и устройств на их основе, принципы построения оптических систем искусственного интеллекта.

Дисциплина «Оптическая физика» предполагает изучение не только общих закономерностей записи и воспроизведения волновых полей, но и прикладных аспектов кодирования, сжатия, хранения и отображения информации, необходимой для

диагностики в различных прикладных областях фотоники и оптоинформатики.

В учебный план подготовки бакалавра включены дисциплины «Введение в фотонику и оптоинформатику», «Нелинейная оптика», «Технология искусственного интеллекта», «Материалы и технологии интегральной и волоконной оптики», «Наноматериалы и нанотехнологии», «Волноводная фотоника», «Лазерные, нелинейные и регистрирующие среды».

Эти дисциплины основаны на фундаментальных законах природы. В области любой техники человек является творцом нового. Эта роль выполняется через проектирование и воплощение проектов в реальность. Поэтому многие специальные курсы явно или опосредованно содержат информацию, необходимую для освоения методов проектирования. Для разработчиков нового знание фундаментальных основ функционирования проектируемых объектов и процессов приобретает непосредственную профессиональную значимость. Человек – творец нового – в отличие от мифического Творца не всемогущ. *В своей творческой деятельности он должен неукоснительно подчиняться фундаментальным правилам «запрета».* Этими правилами являются законы природы. Игнорирование хотя бы одного из этих законов приводит к неосуществимости проекта. Например, не удалось осуществить ни один из многих тысяч проектов вечного двигателя 1-го и 2-го рода. Их авторы действовали вопреки двум правилам «запрета»: «Вечный двигатель первого рода невозможен», «Вечный двигатель второго рода невозможен». Первый запрет связан с первым началом термодинамики, т.е. с законом сохранения энергии, а второй – со вторым началом термодинамики. Следовательно, приступая к проектированию, необходимо выяснить, не противоречит ли задуманное какому-либо из законов природы. В этом может преуспеть только специалист, знающий эти законы.

Кроме того, проектировщик, создавая принципиально новое, создает практически новую «популяцию» искусственной природы.

Всякая популяция в естественной и искусственной природе имеет громадное число жизненно важных для нее связей с окружающим. Эти связи должны быть оптимизированы, иначе результат проектирования не приобретет необходимой конкурентоспособности (по техническим, экономическим или другим характеристикам) и будет вытеснен другими техническими новинками. Помимо этого, объекты современного проектирования имеют, как правило, достаточно сложную структуру и многочисленные внутренние взаимосвязи. Эти взаимосвязи проектировщик должен в полном объеме понимать и учитывать. Исчерпывающий учет внешних и внутренних взаимосвязей проектируемого объекта по силам лишь специалисту, обладающему достаточно развитым системным мышлением. Системное мышление наиболее эффективно формируют фундаментальные науки. Это связано с тем, что именно они имеют дело со сложнейшими системами, включая ноосферу, галактики и всю Вселенную. Поэтому знание фундаментальных дисциплин является особенно важным для создателей принципиально нового в технике.

Применяя, проектируя и изобретая, бакалавр должен проявлять высокий уровень творческого мышления, умения решать конкретные задачи. Поэтому в план подготовки включена дисциплина «Учебно-исследовательская работа студентов» (УИРС).

3.4. Гуманитарные, социальные и экономические дисциплины

Гуманитарные, социальные и экономические дисциплины входят в 1-й блок дисциплин и обычно изучаются на первых двух курсах. Часто в сознании студента эти дисциплины резко отделяются от профессиональных, и даже математических и естественно-научных дисциплин. В основе этого лежат те же причины, которые привели к размежеванию между «лириками и физиками» и возникновению так называемой проблемы «двух культур».

В античной науке этой проблемы не было. На ранних этапах развития науки исследования различных аспектов проявля-

ний природы были объединены и проводились одними и теми же людьми. Обострение этой проблемы связано с *дифференциацией науки, которая распалась на множество направлений, концентрировавшихся вокруг двух обширных сфер познания – естественно-научной и социогуманитарной.*

В естественных науках объектом изучения является неживая природа и биологические аспекты живой природы. Социогуманитарные науки изучают человека и созданные им системы (языки, социальные системы, экономику, науку, право, искусство, религию и т.д.). Однако в наше время все отчетливее осознается *общность всего в природе. Все имеет единую первичную материальную основу, подчиняющуюся одним и тем же фундаментальным законам природы, изменяется в рамках общего эволюционного процесса. Все чаще методы исследования, созданные в естественных науках, используются в социогуманитарных областях. Современная научная картина мира становится единой.* Она включает достижения не только естественных, но и социогуманитарных наук. Все явственнее дает себя знать взаимное проникновение (конвергенция) естественно-научного и гуманитарного знания. *Все большую роль гуманитарные науки играют в формировании у специалиста понимания своей ответственности за устойчивое развитие человечества на Земле, за недопущение глобальных техногенных катастроф.*

Интеграция гуманитарных и естественно-научных знаний лежит в основе подготовки широкообразованных специалистов.

Федеральный компонент учебного плана подготовки бакалавра по направлению «Фотоника и оптоинформатика» содержит 9 социогуманитарных дисциплин. Некоторые из них непосредственным образом связаны с современным естествознанием, общими фундаментальными законами природы. К таким дисциплинам относятся «Основы логической культуры и риторики», «Социология», «Экономика» и «Философия».

Дисциплина «Основы логической культуры и риторики» описывает личностные аспекты рационального (сознательного)

и иррационального (интуитивного) поведения человека в социальном окружении. Однако человек – это материальная система, возникшая в процессе эволюции Вселенной. Особенность указанной системы в том, что посредством нее материя проявляет свою способность к самопознанию. В процессе самопознания этой системы, т.е. человека, возникла психология. Поэтому, строго говоря, психология и педагогика такие же естественные науки, как и физика, химия, биология. В отличие от перечисленных наук, психология и педагогика изучает более сложную систему и такие процессы, описание которых не поддается пока строгому математическому моделированию.

Социология описывает еще более сложную материальную систему, состоящую из громадного количества отдельных личностей с необозримо большим количеством неопределяемых взаимосвязей. Поэтому строгие математические теории в социологии также отсутствуют. Однако уже имеются обоснованные макромоделли некоторых социальных процессов и разрабатываются математические методы их анализа. Сближение социологии с естественными науками становится все более явственным.

Экономика имеет дело со сложнейшей материальной системой, включающей все сферы производства, перемещения и распределения товаров и услуг. В этой системе процессы «переноса» финансов, товаров и рабочей силы, процессы самоорганизации отдельных подсистем в микро- и макроэкономике весьма напоминают аналогичные процессы в неживой природе. Однако, в отличие от неживой природы, в экономике могут возникать дополнительные движущие силы, искажающие естественные рыночные процессы. Сюда относятся политические решения, внешние силовые воздействия (например, военные) и т.д. Несмотря на сложность экономических систем, уже существуют и эффективно используются в экономических прогнозах достаточно строгие математические модели.

Философия связана с осмыслением общих методов познания неживой, живой и мыслящей природы. Этим она приобща-

ет себя к естественным наукам. Особенно важна философия для развития специалиста с широким взглядом на окружающий мир, свое место и свою роль в этом мире.

Еще пять дисциплин рассматриваемого блока «Менеджмент», «История», «Экономика предприятия», «Культурология», «Основы патентоведения и защиты интеллектуальной собственности» имеют связь с современным естествознанием через социологию, отдельные частные аспекты которой они развивают в специфических направлениях.

Наконец, дисциплины «Иностранный язык», «Физическая культура», с естествознанием связаны лишь косвенно. Для языка эта связь оказывается значимой, например, при создании искусственных синтезаторов речи, для физической культуры – при формировании концепции здорового образа жизни, разработке теории движений человека в различных видах спорта и т.д.

3.5. Основные требования бакалаврской подготовки

При обучении в вузе важно иметь в виду, что современный рынок труда предъявляет все более высокие требования к профессиональной мобильности специалиста. Например, по прогнозам, выпускник технического вуза США к 2020 г. будет вынужден на протяжении профессиональной карьеры изменять род своей деятельности до 5 раз. Очевидно, что всякая вынужденная перемена рода профессиональной деятельности – весьма трудная проблема для узкого специалиста. Только для человека, обладающего глубокими фундаментальными знаниями, освоение конкретной специфики новой инженерной профессии не составит большого труда. Это объясняется общностью фундаментальных основ практически всех технических специальностей. В процессе обучения по направлению «Фотоника и оптоинформатика» студент получит весьма глубокие знания в области фундаментальных наук и приобретет высокую профессиональную мобильность.

В течение четырех лет студент вуза вынужден изучать много различных курсов. Реально никто не способен сохранить

в своей памяти содержание такого числа дисциплин, но и не в этом цель образования. Это хорошо понимают как сами педагоги, так и специалисты, эффективно использовавшие плоды образования в своей профессиональной деятельности и жизни. Например, один из крупнейших ученых прошлого столетия Макс Планк, заложивший основы квантовой физики, утверждал: «Образование – это то, что остается, когда все изучавшееся забудется». Остается же (должно оставаться!) развитое мышление, позволяющее адаптироваться ко всем переменам в науке и технике и эффективно участвовать в научно-техническом прогрессе.

Понятие «профессиональное мышление» многогранно. Однако его основные характерные признаки можно свести к трем: критичность, творчество, системность. *Высокая критичность помогает раньше конкурентов вскрыть потребность в новации, творчество позволяет ее своевременно предложить, а системность мышления обеспечивает учет всех значимых внутренних и внешних взаимосвязей предложенного*, что гарантирует надежность и конкурентоспособность новой разработки.

Крупнейший американский инженер-изобретатель Томас Эдисон говорил: «Величайшая задача цивилизации – научить человека мыслить». Кратчайший путь решения этой задачи – познакомить студента с наиболее эффективными методами мыслительной деятельности человека, опирающимися на фундаментальные основы тех учебных курсов, которые он изучает в вузе.

Именно приобретая фундаментальные знания, можно наиболее результативно развить мышление и интегрировать в своем сознании естественно-научные основы изучаемых дисциплин. А это как раз та база, на которой формируется широкообразованная личность и высококвалифицированный специалист, способный не только поспевать за научно-техническим прогрессом, но и участвовать в его развитии, а также легко изменять при необходимости сферу своей профессиональной деятельности. Именно поэтому в последнее время увеличивается число фирм, предпочитающих принимать на работу специалистов с широким фундаментальным образованием. Следует также иметь в виду,

что в современных условиях успешность специалиста обеспечивается в конечном счете не только его квалификацией, удостоверенной дипломом, но и особым качеством личности, которое называют компетентностью. Компетентность – это ярко выраженная способность специалиста проявлять знания, умения и навыки в своей профессиональной деятельности (в том числе при решении принципиально новых задач, к решению которых вуз не готовит). В вузе студент имеет дело с десятками учебных дисциплин, каждая из которых содержит примеры решений творчески ориентированными специалистами и таких проблем, которые ранее считались неразрешимыми. Подобные примеры студент должен использовать для приобретения собственных профессиональных знаний.

Вопросы для самоконтроля

1. Какой документ определяет квалификационную характеристику, сроки, последовательность и содержание подготовки бакалавров? Охарактеризуйте этот документ.

2. Какие дисциплины относятся к блоку «Математические и естественно-научные дисциплины»? Охарактеризуйте естественно-научную основу каждой из них.

3. Как связаны профессиональные дисциплины с дисциплинами естественно-научного блока? Приведите примеры.

4. Какую роль в творческой деятельности бакалавра играет знание фундаментальных законов природы?

5. Как связаны гуманитарные, социальные и экономические дисциплины с фундаментальными науками? Приведите примеры.

6. Какие процессы отражают дифференциацию и интеграцию различных наук?

7. Что такое «проблема двух культур», как изменялась со временем ее острота?

8. Какова связь между профессиональной мобильностью специалистов и фундаментализацией образования?

9. Какие основные признаки характеризуют профессиональное мышление?

ЧАСТЬ 2. НАУЧНЫЕ ОСНОВЫ ФОТОНИКИ

4. ЭЛЕМЕНТЫ КВАНТОВОЙ ФИЗИКИ

4.1. Связь фотоники и оптоинформатики с квантовой физикой

Фотоника – область науки и техники, связанная с использованием светового излучения (или потока фотонов) в элементах, устройствах и системах, в которых генерируются, усиливаются, модулируются, распространяются и детектируются оптические сигналы; *оптоинформатика* – выделившаяся область фотоники, в которой создаются оптические устройства и технологии передачи, приема, обработки, хранения и отображения информации.

Фотон (от греческого *phōtós* – свет) – элементарная частица квантового электромагнитного поля с нулевой массой покоя, его скорость равна скорости света. Представление о фотоне возникло в ходе развития квантовой теории и теории относительности. Термин «фотон» был введен Г. Льюисом (*G. Lewis*) в 1929 году. В результате развития квантовой механики стало ясно, что свет обладает не только волновыми, но и корпускулярными свойствами. Понятие кванта-излучения было введено М. Планком (*M. Planck*) в 1900 году. Развивая идею Планка, А. Эйнштейн ввел гипотезу световых квантов, согласно которой световые кванты – реальные элементарные частицы, подчиняющиеся тем же кинематическим законам, что и частицы вещества. Фотон, как и другие элементарные частицы, проявляет как волновые, так и корпускулярные свойства. В электростатическом поле атомного ядра фотон с энергией больше 1 МэВ может превратиться в электрон и позитрон (процесс *рождения пар*). А при столкновении электрона и позитрона может произойти их *аннигиляция* с выделением γ -квантов.

Часть спектра электромагнитного излучения, ограниченная с одной стороны рентгеновскими лучами, а с другой – микроволновым диапазоном радиоизлучения, называется *оптическим излучением*. Наука о зрительных восприятиях – раздел физики, в котором изучаются оптическое излучение, его распространение и явления взаимодействия света и вещества, называется *оптикой* (от греческого *optikē*).

Как и любая другая частица, свет является носителем информации (от лат. *information* – разъяснение, осведомление). С середины XX века понятие «*информации*» стало общенаучной категорией, что было связано с введением количественной меры информации, разработкой теории информации, всеобщим распространением компьютеров, становлением *информатики* – науки о статистических процессах передачи информации в технических, природных и социальных системах. Введенное К. Шенноном (*C. Shannon*) представление о количестве информации, содержащемся в том или ином сообщении, тесно примыкает к понятию *энтропии*. Связь между этими понятиями становится особенно содержательной, если учесть, что, как и в термодинамике, получение любой информации неизбежно связано с затратами энергии.

С появлением компьютеров (от англ. *compute* – вычислять) информатика приобретает новый импульс развития. Компьютеры, построенные на электроде как носители информации (электронно-вычислительные машины – ЭВМ), имеют ограниченные возможности из-за свойств электрона, недостаточной плотности записи информации. Фотон как носитель информации имеет перед электроном преимущества как по скорости переноса информации, так и по плотности ее записи. Поэтому *оптоинформатика* открывает новые возможности в технологии передачи, приема, обработки, хранения и отображения информации.

Фотоника и оптоинформатика зародились в недрах квантовой физики, базируются на достижениях технологий, манипулирующих на уровне атомных размеров (нанотехнологии). По-

этому учебный план по специальности «Фотоника и оптоинформатика» содержит много дисциплин, являющихся разделами квантовой физики.

Курс «Введение в специальность» для данной специальности может быть полноценно изложен лишь на основе определенных сведений из квантовой физики. Некоторые из них содержатся в курсах физики и химии средней школы. Эти сведения (с некоторыми добавлениями) излагаются в настоящей главе.

4.2. Этапы развития фотоники и оптоинформатики

Появление в научной литературе термина «*фотоника*» связано с выходом в свет в 1967 году книги академика А.Н. Теренина (1896–1967) «Фотоника молекул красителей». Тремя годами раньше по его инициативе на физическом факультете Ленинградского государственного университета создается кафедра биомолекулярной и фотонной физики, которая с 1970 года называется кафедрой фотоники. А.Н. Теренин определил фотонику как «совокупность взаимосвязанных фотофизических и фотохимических процессов». В мировой науке получило распространение более позднее и более широкое определение фотоники, как раздела науки, изучающего системы, в которых носителями информации являются фотоны. В этом смысле термин «фотоника» впервые прозвучал на 9-м Международном конгрессе по скоростной фотографии (Denver, USA, 1970).

Оптоинформатика – это область фотоники, быстро развивающееся научно-техническое направление в России и в мире, связанное с исследованием, созданием и эксплуатацией новых материалов, устройств и технологий, направленных на передачу, прием, обработку, хранение и отображение информации на основе материальных носителей – фотонов.

Слово *информатика* происходит от французского слова *Informatique*, образованного в результате объединения терминов *Information* (*информация*) и *Automatique* (*автоматика*), что выражает ее суть как науки об автоматической обработке инфор-

мации. Исторически информатика как техническая наука, систематизирующая приемы создания, хранения, воспроизведения, обработки и передачи данных средствами вычислительной техники, а также принципы функционирования этих средств и методы управления ими, получила импульс развития с появлением электронно-вычислительных машин (ЭВМ).

ЭВМ принято делить на поколения. Это деление связано не только со временем появления различных их типов на мировом рынке, но и с особенностями архитектуры и принципов построения программного обеспечения. Формальным критерием отнесения того или иного типа ЭВМ к какому-либо поколению является та элементная база, на которой построены логические схемы ЭВМ.

ЭВМ *первого поколения* представляли собой солидные инженерные сооружения, начиненные тысячами радиоламп. По современным меркам это были совершенно ненадежные устройства, которые могли давать сбой чуть ли не каждые полчаса. Их быстродействие составляло от нескольких сотен операций в секунду до двух-трех десятков тысяч. Эпоха ЭВМ первого поколения длилась с 1945 года до середины 50-х годов. Для сравнения приведем некоторые данные. Одной из первых ЭВМ первого поколения в нашей стране была БЭСМ-1 (быстродействующая электронная счетная машина). Она могла выполнять до восьми тысяч операций с плавающей запятой в секунду (рекордное по тем временам быстродействие), объем оперативной памяти составлял всего лишь 1024 слова (приблизительно 5 Кбайт).

В 1954 году в качестве элементной базы радиоэлектронной аппаратуры стали использоваться транзисторы и диоды – так называемые твердотельные элементы. Впервые они были применены в радиоприемниках, но в скором времени стали использоваться для создания логических схем ЭВМ. Это ознаменовало начало эпохи компьютеров *второго поколения*. Для машин второго поколения быстродействие достигало уже нескольких сотен тысяч операций в секунду, а объемы оперативной памяти – нескольких десятков Кбайт.

Архитектура машин первого и второго поколений прямо отражала принцип фон Неймана: *строго последовательное выполнение команд, хранимых в строго линейной общей памяти для команд и данных*. По современной терминологии такие машины работали в однопользовательском режиме.

В качестве элементной базы ЭВМ *третьего поколения* использовались интегральные схемы (ИС), выпуск которых начался в начале 60-х годов. В ИС на одной пластине кристалла кремния можно было размещать достаточно сложные логические схемы, состоящие из нескольких транзисторов и диодов. В состав этих объединяющих схем входили не только активные логические элементы, но и соединительные проводники, сопротивления, емкости. Интегральные схемы воплощают законченные функциональные элементы, такие как полный сумматор, хранящий регистр, счетчик и т.д.

Для ЭВМ *третьего поколения* помимо повышения быстродействия и надежности характерны такие особенности их использования, как мультипрограммный режим, дистанционный доступ с использованием удаленных терминалов, разделение времени центрального процессора между многими пользователями и задачами. Появилось семейство программно-преемственных машин, идея преемственности впервые была реализована фирмой IBM в семействе IBM-360.

Эпоха машин *четвертого поколения* связана с выпуском больших ИС в начале 70-х годов. Характерной чертой данного периода развития средств вычислительной техники является формирование двух наиболее интенсивно развивающихся направлений: массового выпуска персональных ЭВМ (ПЭВМ) и выпуска супер-ЭВМ сериями в десятки и сотни штук.

В первом направлении стремление упростить взаимоотношения пользователя с ПЭВМ привело к созданию многооконного графического интерфейса, некоторой внешней оболочки, в которой текстовое меню в основном заменено графическими символами – «иконками», или пиктограммами, а нажатие кла-

виш заменено манипулированием «мышью». Графический интерфейс с широкими возможностями реализован в многочисленных версиях *Windows*-оболочек.

Второе направление развития, связанное с суперЭВМ, характеризуется появлением крупных серий вычислительных систем с *параллельной обработкой данных*, обладающих огромной производительностью, исчисляемой сотнями миллионов и десятками миллиардов операций в секунду.

Распараллеливание обработки данных, предоставляемое структурной организацией суперЭВМ, потребовало разработки совершенно новых алгоритмов организации вычислительного процесса, новых языков параллельного программирования и новых численных методов, т.е. принципиально новых средств математического и программного обеспечения.

О машинах *пятого поколения* специалисты заговорили в начале 80-х годов. К началу разработки проекта машин пятого поколения уже существовали экспертные системы, базы знаний, системы машинного логического вывода, языки логического программирования, системы распознавания сложных образов, т.е. все то, что именуется искусственным интеллектом. Это позволило общаться с ЭВМ на языках, близких к естественным, заставить машины решать сложные логические задачи, по уровню не уступающие тем, которые решаются высококвалифицированными специалистами в их профессиональных областях, превратить визуализацию и озвучивание в неотъемлемый элемент интерактивного режима.

Начало XXI века знаменует появление машин *шестого поколения*. Эти машины должны обслуживать новейшие информационные технологии, которые позволят объединить все компьютеры в единое информационное пространство, которое предоставит возможность с помощью персональных компьютеров общаться друг с другом, получать оперативную информацию любого интересующего содержания и вида, обращаться к базам данных и знаний, распределенных по информационным центрам всего

мира. Иными словами, вычислительная техника вместе с ее программным обеспечением должна заменить собой телефон, телефакс, информационно-справочное бюро, собственный архив и средства массовой информации. Главный акцент делается на разработке единого глобального информационного пространства, доступ к которому должен быть обеспечен большинству людей.

Настоящее время характеризуется быстрым увеличением тактовой частоты процессоров, которая преодолела гигагерцевый рубеж, и позволила на одном процессоре достигать гигафлопной производительности (10^9 операций в секунду). Несмотря на столь высокие скоростные характеристики одиночного процессора, продолжает бурно развиваться линия суперЭВМ с массовым параллелизмом. Пущены в эксплуатацию гиганты с пиковой производительностью в десятки терафлопс (10^{12} операций в секунду), проектируются вычислительные системы пентофлопной (10^{15} операций в секунду) производительности.

Основу машин массового производства составляют микропроцессоры (МП). Они также делятся на поколения, но это деление основано на других критериях, хотя и связанных со сроком появления их на рынке, однако в меньшей степени, чем в вышеприведенной классификации больших ЭВМ. Основу такого деления для МП составляет размер их разрядной сетки. Говорят о 8-разрядных МП, о 16-разрядных, 32-, 64- и 128-разрядных микропроцессорах.

Этот формальный критерий отражает рост функциональных возможностей МП. Рост разрядности сопровождается ростом объемов оперативной памяти, доступной МП, резким увеличением адресного пространства, расширением возможностей по способам адресации, ростом логических возможностей арифметического устройства по выполнению логических и арифметических операций. Например, 64-разрядный процессор, изготовленный на одном кристалле (чипе), содержит несколько миллионов логически активных элементов. Современное поколение МП работает на частотах, превышающих гигагерцевый рубеж, имеет развитый внутренний параллелизм и логически весьма сложную архитектуру.

Однако казавшиеся неисчерпаемыми возможности развития электронной вычислительной техники, быстродействие которой начиная с 70-х годов удваивалось практически каждые полтора года, исчерпывается. Это связано с *фундаментальными ограничениями возможностей микроэлектроники*.

В информационных системах ключевое место стала занимать оптоэлектроника. Это обусловлено не только тем, что более 90 % информации, которую получает человек, составляет видеoinформация. В широком спектре получения, обработки, хранения, передачи и отображения информации ведущее место занимают приборы оптоэлектроники, включая индикаторные системы, формирователи сигналов изображения, волоконно-оптические линии передачи информации, преобразователи солнечной энергии и др. Современная оптоэлектроника характеризуется большим разнообразием серийно выпускаемой продукции (рис. 4.1), в которой в стоимостном выражении доминирует индикаторное направление.

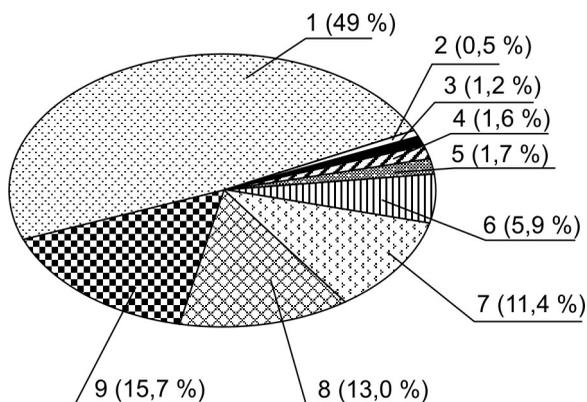


Рис. 4.1. Структура мирового рынка производства приборов оптоэлектроники в денежном исчислении (%): 1 – индикаторные компоненты; 2 – пассивные оптические компоненты; 3 – оптические соединители; 4 – солнечные батареи; 5 – волоконные линии; 6 – гибридные приборы; 7 – фотоприёмники; 8 – оптические волокна; 9 – прочие компоненты

Качество перспективной оптоэлектронной продукции зависит от успехов технологии создания новых высокоэффективных полупроводниковых материалов и структур на их основе, включая квантово-размерные структуры (наноструктуры с квантовыми ямами, квантовыми нитями и квантовыми точками).

Вместе с тем, несмотря на значительные успехи оптоэлектроники, существует точка зрения, что ее состояние на сегодняшний день можно сравнить с уровнем развития электроники до изобретения транзистора. Как правило, в настоящее время оптоэлектронная продукция продается в основном в виде дискретных элементов. Резкое расширение оптоэлектронного рынка ожидается после того, как произойдут существенные изменения в функциональных характеристиках приборов.

К таким качественно новым оптоэлектронным изделиям можно отнести оптоэлектронные компьютеры с архитектурой, подобной человеческому мозгу; стереоскопические системы, имитирующие характеристики человеческого зрения с автоматическим распознаванием движущихся объектов.

С дальнейшим развитием технологии оптоэлектронных датчиков становятся реальными обработка и отображение не только видимых, но и рентгеновских, ультрафиолетовых и инфракрасных изображений, а также визуализация ультразвуковых полей.

Главным недостатком оптоэлектронных устройств являются значительные потери энергии при ее обмене между электронными и фотонными приборами. Стремление уменьшить эти потери приводит к информационно-телекоммуникационным технологиям только на оптических принципах. Несмотря на то, что на данный момент еще не представляется возможным столь же эффективно управлять фотонами, как электронами, уже повсеместно применяются волоконно-оптические системы связи, скорость передачи информации по которым тысячекратно превышают предельные скорости в электрических линиях. По существу, без оптоволоконных систем связи немыслим современный Интернет.

Создаются *оптические системы сверхплотной записи информации*. Пока далекие от совершенства, но созданы и элементы цифровых оптических компьютеров. Их архитектура более причудлива и богата дополнительными возможностями, отсутствующими у электронного аналога. Но самым многообещающим вкладом оптики в информатику может стать создание оптического компьютера и искусственного интеллекта.

Чтобы использовать уникальные возможности оптики в информационных технологиях, необходимо совершенствовать технологии создания наноразмерных устройств генерации, детектирования оптических сигналов, а также оптических логических элементов, управляемых светом. Элементарная оптическая ячейка должна потреблять энергии меньше, чем элемент микро-чипа, быть интегрируемой в большие массивы и иметь возможность связи с большим числом подобных элементов. Создание такой ячейки является основной задачей оптоинформатики на ближайшие десятилетия.

Отметим ближайшие цели, которые могут быть достигнуты в системах оптоинформатики:

- оптические телекоммуникационные системы со скоростью передачи данных до 40 терабит в секунду по одному каналу (бит – единица количества информации в двоичной системе);

- оптические голографические запоминающие устройства сверхбольшой емкости до 1,5 терабайт на диск стандартных размеров;

- многопроцессорные компьютеры с оптической межпроцессорной связью;

- оптический компьютер, в котором свет управляет светом. Максимальная тактовая частота такого компьютера может составлять 10^{12} – 10^{14} Гц, что на 3–5 порядков выше существующих электронных аналогов;

- фотонные кристаллы – новые искусственные кристаллы, имеющие гигантскую дисперсию и рекордно низкие оптические потери (0,001 дБ/км, Белл – единица логарифмического уровня

интенсивности света P_2 относительно начального уровня P_1 , $x = \lg(P_2/P_1)$, $x = 1$ Б, если $P_2 = 10P_1$).

Телевидение сверхвысокого разрешения, трехмерная компьютерная графика для систем виртуальной реальности и другие продукты становятся с развитием оптоинформатики доступными для многих потребителей.

Возникновение и развитие фотоники и оптоинформатики стало возможным благодаря фундаментальным достижениям в области современной квантовой физики. Чтобы разобраться в принципе действия оптоэлектронных и оптических устройств, необходимо изучить основы квантовой механики. Поэтому следующие параграфы пособия посвящены повторению и некоторому расширению элементарных представлений о квантовой природе окружающего нас мира, полученных из курса физики средней школы.

4.3. Основные представления квантовой механики

Квантовая механика – раздел теоретической физики, изучающий законы движения микрочастиц и их систем. В отличие от классической механики квантовая механика учитывает двойственную корпускулярно-волновую природу частиц и квантуемость, т.е. скачкообразное, дискретное изменение многих характеристик.

Квантовая механика – один из наиболее плодотворных в прикладном отношении разделов физики. На ней базируются многие современные отрасли техники. Сюда относятся микро-, опто- и наноэлектроника, лазерная технология, солнечная, атомная и термоядерная энергетика и т.д. Это объясняется тем, что законы квантовой механики в той или иной степени определяют свойства и поведение материальных структур: от физического вакуума до макротел. В мире элементарных частиц, атомов и наноструктур законы квантовой механики играют решающую роль. Они определяют также электрические, оптические и магнитные свойства кристаллов.

В школьной программе по физике и химии квантовые представления связываются с дискретностью излучения и поглощения света, фотоэффектом, корпускулярно-волновым дуализмом частиц, соотношением неопределенностей Гейзенберга, строением атома, атомными спектрами, принципом Паули, структурой электронных оболочек и валентностью атомов, особенностью расположения атомов химических элементов в периодической системе Менделеева, свойствами атомного ядра и элементарных частиц, физикой полупроводников и полупроводниковых приборов.

В данном пособии мы повторим упомянутый выше учебный материал и расширим его до объема, который необходим для осознанного восприятия основных идей фотоники и оптоинформатики.

Первые представления о квантовой природе микрообъектов и процессов возникли при изучении закономерностей теплового излучения. Было показано, что строгая теория теплового излучения, доказанная экспериментально во всем диапазоне длин волн, может быть построена лишь в предположении, что свет излучается веществом дискретно, т.е. отдельными порциями – квантами. Это предположение называется *гипотезой Планка*, который показал также, что энергия кванта излучения E пропорциональна его частоте ν :

$$E = h\nu, \quad (4.1)$$

где $h = 6,62 \cdot 10^{-34}$ Дж·с – постоянная Планка.

Позднее было установлено, что постоянная Планка связана с широким кругом физических явлений в микромире, для которых существенна квантованность величин с размерностью действия. Действие – физическая величина, имеющая размерность произведения энергии на время и являющаяся одной из важнейших характеристик движения систем в классической и квантовой механике. Поэтому постоянная Планка относится к фундаментальным физи-

ческим константам. За работы в области теплового излучения (1900) Планк был удостоен Нобелевской премии (1918).

При построении теории внешнего фотоэффекта Эйнштейн (1905) предположил, что свет не только испускается и поглощается квантами, но и распространяется как поток особых частиц (фотонов), несущих на себе дискретную порцию энергии, равную $h\nu$. На основе квантовых представлений о свете Эйнштейн объяснил не только фотоэффект, но и другие физические явления, которые не поддавались описанию с точки зрения электромагнитной теории света (Нобелевская премия, 1921).

Тем не менее теоретические и экспериментальные основания электромагнитной теории столь неоспоримы, что волновая теория света осталась непоколебимой. Она была лишь дополнена убедительно обоснованной квантовой теорией света. Было признано, что свет имеет двойственную (корпускулярно-волновую) природу, а модуль импульса фотона $p = |\vec{p}|$ может быть выражен через соответствующую длину волны λ :

$$p = h/\lambda, \quad (4.2)$$

где $\lambda = c/\nu$, а ν находится для данного кванта света из формулы Планка (4.1).

В квантовой механике формулу (4.2) обычно записывают в векторной форме:

$$\vec{p} = \hbar \vec{k}, \quad (4.3)$$

где $\vec{k} = \frac{2\pi}{\lambda}$, $\hbar = \frac{h}{2\pi} \approx 1,05 \cdot 10^{-34}$ Дж·с.

Величина \vec{k} , называемая *волновым вектором*, широко используется в квантовой механике. Направление волнового вектора совпадает с направлением вектора импульса \vec{p} фотона, т.е. волновой вектор направлен в сторону распространения световой

волны. Модуль волнового вектора $|\vec{k}|$ называется *волновым числом*. Величина \hbar (так же, как и h) называется постоянной Планка.

Итак, в начале XX века была установлена двойственная, корпускулярно-волновая природа частиц света – фотонов. *Де Бройль предположил* (1923), что волновые свойства присущи не только фотонам, но и любым частицам вещества. При этом длина волны, соответствующая любой частице, называется *волной де Бройля* и определяется формулой (4.2):

$$\lambda = \frac{2\pi\hbar}{p} = \frac{h}{p} = \frac{h}{mv}, \quad (4.4)$$

где m – масса частицы, v – ее скорость.

Гипотеза о корпускулярно-волновом дуализме приобрела универсальный характер и была использована Э. Шрёдингером для получения *основного уравнения квантовой механики – уравнения Шрёдингера* (Нобелевская премия, 1933).

Восприятие корпускулярно-волнового дуализма и квантовых представлений об объектах и процессах в микромире оказалось столь трудным, что даже творцами квантовой механики принималось не без колебаний. Вот характерное высказывание Шрёдингера: «Если мы собираемся сохранить эти проклятые квантовые скачки, то я вообще жалею, что имел дело с атомной теорией». Однако жизнь показала, что именно квантовая механика позволила с удивительной полнотой и точностью описать происходящее в микромире. Она получила разностороннее экспериментальное подтверждение и широчайшее применение на практике. В фотонике, например, волновые свойства частиц используются для получения схемных элементов, имеющих наноразмеры.

Основное уравнение классической механики – *второй закон Ньютона* – дает возможность получить уравнение движения материальной точки, которое позволяет установить точные значения координат этой точки x в каждый момент вре-

мени t . Иными словами, основной закон классической механики отражает детерминированное поведение макротела.

Поведение микрочастиц носит недетерминированный, вероятностный характер, что связано с их волновыми свойствами. Поэтому уравнение Шрёдингера записано относительно функции, которая позволяет определять *вероятность* нахождения микрочастицы в том или ином объеме пространства в некоторый момент времени t . Эта функция называется волновой или пси-функцией. Она обозначается греческой буквой Ψ . Квадрат модуля этой функции $|\Psi(x)|^2$ обозначает плотность вероятности нахождения микрочастицы в точке с координатой x в момент времени t . Если умножить $|\Psi|^2$ на малый элемент объема ΔV , включающий рассматриваемую точку пространства, то получим вероятность нахождения частицы в данном объеме.

Волновые свойства микрочастицы делают невозможным описание ее поведения одновременным заданием точных значений координаты и скорости.

Этот факт математически описывается *соотношением неопределенностей Гейзенберга* (1927) (Нобелевская премия, 1932)

$$\Delta p_x \Delta x \geq \frac{\hbar}{2}, \quad (4.5)$$

где Δp_x – значение неопределенности компоненты импульса частицы по оси x , а Δx – неопределенность координаты частицы в тот же момент времени.

Соотношение неопределенностей имеет место также для энергетических и временных переменных:

$$\Delta E \Delta t \geq \frac{\hbar}{2}, \quad (4.6)$$

где ΔE – неопределенность энергии системы в рассматриваемом квантовом состоянии, Δt – время нахождения системы в этом состоянии.

Соотношение неопределенностей позволяет осмыслить широкий спектр явлений микромира и оценить значения физических характеристик, относящихся к этим явлениям. Соотношение неопределенностей относится к фундаментальным положениям физики, оно получило всестороннее экспериментальное подтверждение.

4.4. Квантовая модель атома

Атом можно описывать посредством двух взаимодополняющих моделей: пространственной и энергетической. Пространственная модель атома отображает его трехмерную объемную структуру и то, как в пределах этой структуры распределены плотности вероятности нахождения электронов атома. Совокупность распределенных вблизи ядра электронов (точнее – вероятностей их нахождения там) называют электронным облаком. В простейшем случае оно сферическое (например, для водорода в невозбужденном состоянии), но чаще имеет сложную конфигурацию. В курсе химии средней школы приведены условные изображения внешней формы электронного облака для электронов, находящихся в различных квантовых состояниях. Для атома водорода уравнение Шрёдингера позволяет получить строгое математическое описание геометрических особенностей электронного облака. Однако во всех случаях его наглядное представление невозможно, так как распределение вероятности нахождения электрона в облаке является довольно сложным. Поэтому при описании атома часто используют упрощенную модель атома – модель Бора (1913) (Нобелевская премия, 1922). Эта модель допускает изображение атома в виде центрального ядра и электронов, движущихся вокруг него по определенным орбитам.

В своей теории Бор принял за основу планетарную модель Резерфорда, но дополнил ее двумя постулатами, отражающими квантовую природу атома. Согласно первому постулату электрон в атоме может двигаться не по любым, а лишь по осо-

бым, стационарным орбитам, не излучая при этом электромагнитной энергии. Каждая орбита характеризуется квантовым числом $n = 1, 2, \dots$, а также квантованной энергией:

$$E_n = -\frac{m_e e^4}{32\pi^2 \epsilon_0 \hbar^2} \frac{1}{n^2}, \quad (4.7)$$

где m_e и e – масса и заряд электрона; ϵ_0 – электрическая постоянная ($\epsilon_0 = 8,85 \cdot 10^{-12}$ Ф/м).

Согласно второму постулату Бора переход электрона с одной стационарной орбиты на другую сопровождается испусканием (или поглощением) атомом кванта электромагнитной энергии:

$$h\nu = E_i - E_k \quad \text{или} \quad \hbar\omega = E_i - E_k, \quad (4.8)$$

где i и k – квантовые числа для соответствующих стационарных орбит; $\omega = 2\pi\nu$ – круговая частота.

Формулу (4.7) и вывод о квантованности энергии атома водорода можно получить путем решения уравнения Шрёдингера для электрона в потенциальном кулоновском поле ядра. Потенциальная энергия электрона U в атоме водорода

$$U(r) = -\frac{e^2}{4\pi\epsilon_0 r}, \quad (4.9)$$

где r – расстояние электрона от ядра.

Энергетическая модель атома водорода (рис. 4.2, а, б) включает представление потенциальной энергии электрона в атоме как функции координат (4.9) и систему энергетических уровней стационарных квантовых состояний атома (4.7).

Из рис. 4.2, а видно, что чем меньше r (чем ближе электрон к ядру), тем больше по абсолютной величине потенциальная энергия $U(r)$. Точке А на рис. 4.2, а соответствует первая стационарная боровская орбита с главным квантовым числом $n = 1$ и энерги-

ей $|E_1| = 13,6$ эВ. Радиус этой орбиты $r_1 = 0,53 \cdot 10^{-10}$ м = 0,053 нм. Согласно квантовой механике никаких орбит нет. Электрон движется вблизи ядра, «посещая» с разной вероятностью все точки пространства. В точке с координатой $r = r_1$ вероятность нахождения электрона в невозбужденном атоме водорода (при $n = 1$) максимальна.

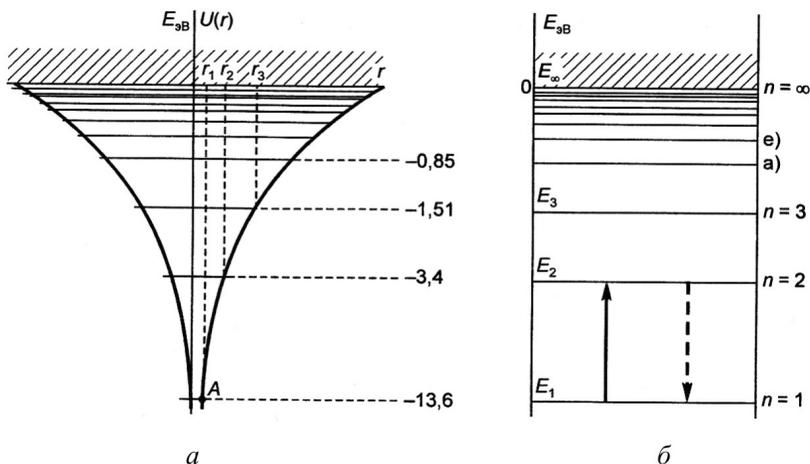


Рис. 4.2. Энергетическая модель атома водорода

В стационарном состоянии ($n = 1$) электрон может находиться бесконечно долго и самопроизвольно выйти из этого состояния не может. Однако если ему сообщить дополнительную энергию, $E_2 - E_1 = 10,2$ эВ, то электрон способен перейти в первое возбужденное состояние ($n = 2$, $E_2 = -3,4$ эВ). На рис. 4.2, б такой переход символизирует сплошная стрелка. В возбужденном состоянии атом способен находиться в среднем в течение 10^{-8} с, а затем самопроизвольно (спонтанно) переходить в стационарное состояние (пунктирная стрелка на рис. 4.2, б). Освобожденная энергия преобразуется в энергию спонтанного излучения: $h\nu = E_2 - E_1 = 10,2$ эВ. *Спонтанное излучение атома* может распространяться в любом направлении. Оно характерно для теплового излучения тел. Воздействие внешнего фотона той же час-

тоты ($h\nu = 10,2$ эВ) на возбужденный атом способно вызвать переход электрона с верхнего энергетического уровня на нижний. Такое излучение называется *вынужденным*. Направление, частота и фаза вынужденного излучения совпадают с направлением, частотой и фазой движения первичного фотона. Вынужденное излучение лежит в основе работы *квантового генератора (лазера)*.

Если электрону в стационарном состоянии ($n = 1$) сообщить энергию $E_1 - E_\infty = 13,6$ эВ, то произойдет ионизация атома водорода. Электрон окажется свободным. *Энергия свободного электрона не квантована, она изменяется непрерывным образом (верхняя, заштрихованная часть на рис. 4.2, б).*

4.5. Понятие о потенциальных ямах и барьерах

Энергетическая модель атома позволяет, как видно из рис. 4.2, а, объяснить некоторые его свойства, например механизм поглощения и испускания атомами электромагнитного излучения. Аналогичным образом *энергетические модели любых квантовых систем могут плодотворно использоваться для анализа свойств этих систем. Энергетический подход опирается на закон сохранения энергии и поэтому является универсальным. Он применяется во всех разделах физики.* Рассмотрим его применение в классической механике.

Предположим, что некоторое тело массой m движется без трения по поверхности, профиль которой представлен на рис. 4.3 кривой $h(x)$. Потенциальная энергия этого тела $U = mgh$, где g – ускорение силы тяжести.

Так как U и h пропорциональны друг другу, то подбором масштаба всегда можно совместить графики функций $U(x)$ и $h(x)$, что и осуществлено на рис. 4.3. При решении задачи о движении тела в потенциальном поле можно использовать лишь кривую $U(x)$. Подобные кривые называются *потенциальными кривыми*.

Участок графика вблизи точки C называется *потенциальной ямой*, а вблизи точки D – *потенциальным барьером*. На рис. 4.3 представлена плоская кривая $U(x)$. С помощью подобных кривых

можно рассматривать одномерные задачи. Понятие о потенциальных ямах и барьерах остается справедливым для двух- и трехмерных задач.

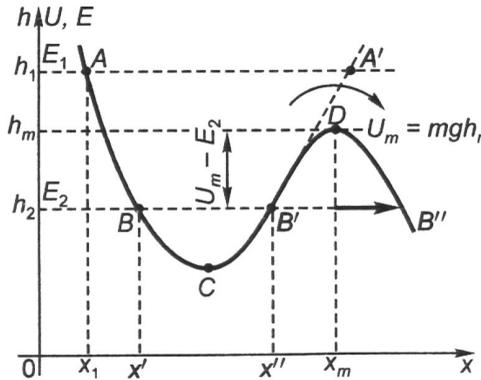


Рис. 4.3. Зависимость потенциальной энергии тела от координаты $U(x)$, совмещенная с профилем поверхности $h(x)$, по которой тело движется

Если трение в рассматриваемой системе отсутствует, то полная энергия E при движении тела m (см. рис. 4.3) сохраняется и можно записать

$$E = U + E_k = mgh + \frac{mv^2}{2}. \quad (4.10)$$

Метод потенциальных кривых позволяет (при заданных E , $U(x)$ и m) определить значение кинетической энергии, скорость тела, действующую на него силу и его ускорение в каждой точке x , а также область значений x , в пределах которой тело может двигаться. Подобные задачи рассматривались в курсе физики средней школы. Для дальнейшего важно уяснить, какую роль играют потенциальные барьеры и потенциальные ямы в поведении частиц, подчиняющихся законам квантовой механики. Тело, подчиняющееся законам классической механики, спо-

собно преодолеть потенциальный барьер, если его полная энергия превышает потенциальную энергию mgh_m на «вершине» барьера (точка D). Например, если тело без начальной скорости (и трения) начинает свое движение под действием силы тяжести от точки A , где его полная энергия $E_1 = mgh_1 > mgh_m$, то оно преодолеет потенциальный барьер и окажется в области значений x правее точки x_1 . Если же тело начинает движение от точки B , где его полная энергия $E_2 = mgh_2 < mgh_m$, то вправо за точкой C оно сможет переместиться только до точки $x = x''$ и, нарушая закон сохранения энергии, не сможет преодолеть потенциальный барьер D . Его движение вдоль оси x будет ограничено точками x' и x'' . Между этими точками тело будет совершать незатухающие колебания (в отсутствие трения). Все точки в указанном интервале равнодоступны для движущегося тела.

Совершенно иначе ведет себя тело, подчиняющееся законам квантовой механики. Для него имеется отличная от нуля вероятность проникнуть за барьер и при условии, когда полная энергия тела меньше высоты потенциального барьера. Такой квантовомеханический эффект называется туннельным. С точки зрения классической физики он нарушает закон сохранения энергии. При этом «дефицит» энергии тела в точке x_m $\Delta E = U_m - E_2$. Однако в соответствии с соотношением неопределенностей (4.6) такой дефицит энергии допустим в течение времени $\Delta t = \hbar / (2\Delta E)$. Если за это время частица туннелирует сквозь барьер, то закон сохранения энергии не нарушается. После туннелирования (в точке B'') у частицы сохраняется та же энергия, что и до туннелирования (в точке B').

Туннельный эффект в микромире достаточно распространен. Он лежит в основе действия туннельного диода, α -радиоактивности, термоядерного синтеза легких элементов. Используется он и в фотонике.

Тело, подчиняющееся законам классической физики, движущееся в потенциальной яме (см. рис. 4.3), имеет непрерывный энергетический спектр, т.е. его энергия E изменяется не-

прерывным образом. Энергетический спектр частицы в потенциальной яме нано- и атомарных размеров дискретен. Этот квантовый эффект лежит в основе функционирования многих наноструктур.

Конфигурации реальных потенциальных ям и барьеров в микромире зависят от геометрических особенностей тех полей, которые образуют эти ямы. Например, потенциальная яма для электрона, движущегося в электрическом кулоновском поле ядра атома водорода, представлена на рис. 4.2, а.

4.6. Микрочастица в прямоугольной потенциальной яме

Поведение микрочастицы в потенциальной яме строго описывается с помощью уравнения Шрёдингера. Это уравнение является дифференциальным; для его решения необходимо знать, как зависит потенциальная энергия микрочастиц от координат, т.е. должна быть задана функция $U(x)$. Здесь и далее будем предполагать, что эта функция не зависит от времени (микрочастица находится в стационарном поле). Для выяснения основных особенностей поведения микрочастицы в потенциальной яме достаточно рассмотреть яму простейшей прямоугольной формы (рис. 4.4, а). Ширина ямы обозначена l , глубина – U^0 . Ширина барьеров, ограничивающих потенциальную яму, бесконечна. Глубина ямы может изменяться от некоторого конечного значения до бесконечности (когда яму можно считать неограниченно глубокой). Ширину ямы будем варьировать от микро- до макро-размеров.

Решение уравнения Шрёдингера дает возможность определить энергетический спектр рассматриваемой микрочастицы, т.е. полный набор значений ее энергии E , и волновую функцию $\Psi(x)$, квадрат модуля которой $|\Psi(x)|^2$ является плотностью вероятности обнаружить микрочастицу в точке x .

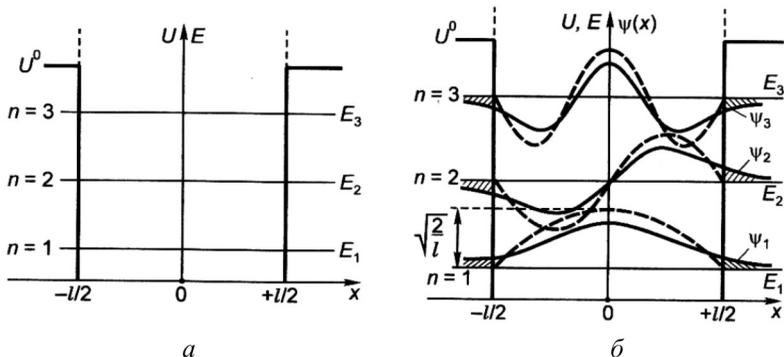


Рис. 4.4. Одномерная потенциальная яма: *a* – энергетический спектр микрочастицы в потенциальной яме (приведены три энергетических уровня E_1, E_2, E_3); *б* – волновые функции Ψ_n для трех состояний ($n = 1, 2, 3$) микрочастицы в потенциальной яме (пунктир для бесконечно глубокой ямы)

Для микрочастицы в *одномерной прямоугольной потенциальной яме* энергия оказывается квантованной и может быть представлена приближенным соотношением

$$E_n = -\frac{\pi^2 \hbar^2}{2ml^2} n^2, \quad (4.11)$$

где $n = 1, 2, 3, \dots$ – квантовое число; m – масса частицы.

Из формулы (4.11) следует, что энергетический спектр микрочастицы в потенциальной яме дискретен, и расстояние между соседними энергетическими уровнями (n и $n - 1$)

$$\Delta E_{n,n-1} = (2n-1) \frac{\pi^2 \hbar^2}{2ml^2}. \quad (4.12)$$

Можно показать также, что на ширине ямы l укладывается примерно целое число полувольт де Бройля:

$$l \approx n \frac{\lambda_n}{2}. \quad (4.13)$$

Для бесконечно глубокой ямы формулы (4.11)–(4.13) становятся точными. В таком случае волновая функция выражается через тригонометрические функции: через косинусы при нечетном n ($n = 1, 3, 5$):

$$\left(\Psi_1 = \sqrt{\frac{2}{l}} \cos \frac{\pi}{l} x, \quad \Psi_3 = \sqrt{\frac{2}{l}} \cos \frac{3\pi}{l} x \right)$$

и через синусы при четном n :

$$\left(\Psi_2 = \sqrt{\frac{2}{l}} \sin \frac{2\pi}{l} x, \dots \right),$$

что и представлено на рис. 4.4, б пунктирными кривыми. На рисунке видно, что в рассматриваемом случае амплитуда волны де Бройля в точках $x = \pm l/2$ обращается в нуль. Следовательно, при $U^0 \rightarrow \infty$ микрочастица не может проникнуть внутрь барьера и выйти за пределы ямы. Если глубина ямы конечна, то амплитуда волны де Бройля в точках $x = \pm l/2$ не обращается в нуль при любых n и имеет продолжение внутри барьера.

На рис. 4.4, б и 4.5 это представлено сплошными кривыми в заштрихованных областях. Иными словами, микрочастица в потенциальной яме конечной глубины U^0 может проникать в глубь барьера при энергиях E , меньших U^0 , что противоречит закону сохранения энергии и не наблюдается в макром мире.

На рис. 4.5 представлена плотность вероятности обнаружения микрочастицы, обладающей энергиями E_1 , E_2 и E_3 в точке с координатой x , что выражается как $|\Psi_1(x)|^2$, $|\Psi_2(x)|^2$ и $|\Psi_3(x)|^2$.

Обсудим теперь особенности поведения микрочастицы в потенциальной яме.

Во-первых, энергетический спектр такой частицы дискретен, а ее минимальная энергия не равна нулю ($E_1 > 0$); энергия E_1 называется нулевой и часто обозначается E_0 . Наличие нулевой

энергии свойственно любым квантовым системам: физическому вакууму, кваркам в адронах, нуклонам в ядрах атомов, электронам в атомах, атомам в молекулах и кристаллах. Из формулы (4.11) следует, что при $l \rightarrow \infty$ и/или $m \rightarrow \infty$, $E_1 = E_0 \rightarrow 0$, т.е. нулевая энергия становится равной нулю, что характерно для макросистем, подчиняющихся законам классической физики. Если $l \rightarrow \infty$ и/или $m \rightarrow \infty$, то при любом конечном n расстояния между соседними энергетическими уровнями частицы в яме стремятся к нулю (4.12), т.е. дискретный энергетический спектр преобразуется в сплошной, что свойственно макросистемам.

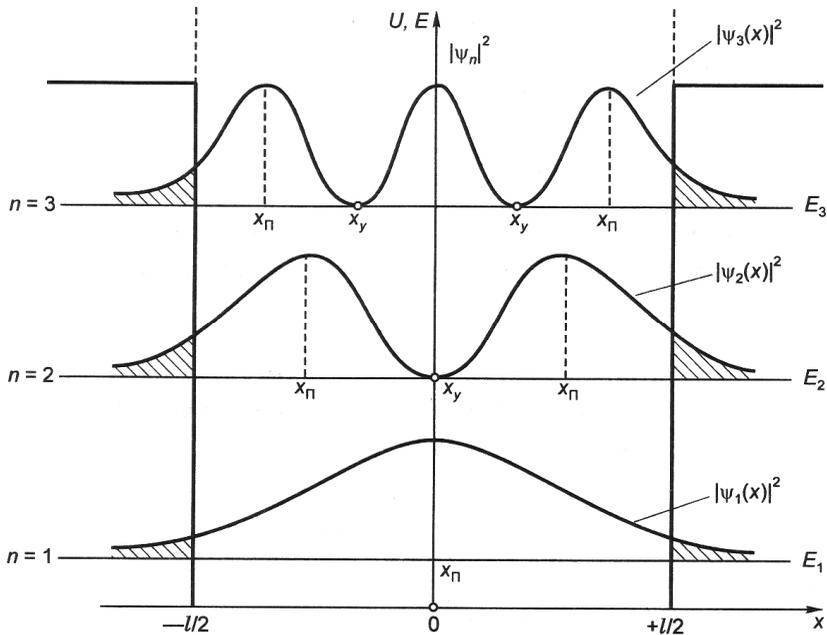


Рис. 4.5. Плотность вероятности $|\Psi(x)|^2$ обнаружить микрочастицу в различных точках x для случая потенциальной ямы конечной глубины

Наконец, из рис. 4.5 следует, что вероятности обнаружить микрочастицу в различных точках внутри ямы существенно неодинаковы. Имеются точки, вероятность «посещения» которых частицей максимальна. Они называются пучностями (x_n). Имеются и такие точки, в которых частица не бывает, – узлы (x_y). Такое поведение совершенно не свойственно макрочастицам.

Для дальнейшего особенно важно то, что вероятность проникновения микрочастицы в область барьера ($|x| > l/2$) не равна нулю, а лишь постепенно убывает с увеличением расстояния от границы барьера (заштрихованные области на рис. 4.5). Если ширина барьера не бесконечна, то имеется отличная от нуля вероятность проникновения микрочастицы за пределы барьера (туннельный эффект). Туннельный эффект лежит в основе действия многих схемных элементов нанoeлектроники. Поэтому рассмотрим особенности этого эффекта более подробно.

4.7. Туннельный эффект

Если потенциальный барьер имеет прямоугольную форму, как это показано на рис. 4.6, то вероятность проникновения микрочастицы массой m за пределы потенциального барьера можно выразить простой формулой

$$D = D_0 e^{-\frac{2}{\hbar} \sqrt{2m(U_0 - E)} \cdot d}, \quad (4.14)$$

где D – коэффициент прозрачности потенциального барьера, равный отношению числа удачных попыток частицы проникнуть за барьер ΔN к общему числу таких попыток N ($D = \Delta N/N$); d – ширина барьера; U^0 – его высота; $(U^0 - E)$ – «дефицит» энергии частицы внутри барьера; D_0 – предэкспоненциальный множитель.

Из формулы (4.14) следует, что вероятность D туннельного эффекта уменьшается с увеличением массы частицы, толщины потенциального барьера и дефицита энергии ($U^0 - E$). Например, коэффициент прозрачности потенциального барьера D_e толщиной

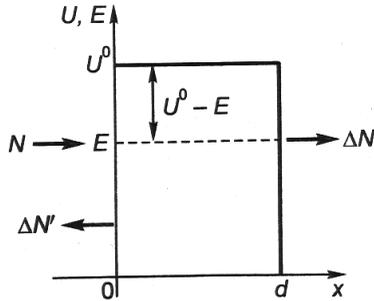


Рис. 4.6. Прямоугольный потенциальный барьер:
 $\Delta N'$ – число неудачных попыток микрочастицы
 преодолеть барьер ($N = \Delta N + \Delta N'$)

$d = 0,1$ нм при $U^0 - E = 1$ эВ равен для электрона 0,4, а для протона $D_p = 10^{-19}$ ($m_e = 9,1 \cdot 10^{-31}$ кг; $m_p = 1,67 \cdot 10^{-27}$ кг). Если толщина барьера $d = 0,4$ нм, то $D_e = 0,01$.

Вопросы для самоконтроля

1. Какие этапы в развитии фотоники и оптоинформатики вы знаете?
2. Перечислите этапы развития электронно-вычислительных машин.
3. Охарактеризуйте дискретную полупроводниковую электронику.
4. В чем состоит фундаментальное ограничение возможностей электронно-вычислительной техники?
5. В чем состоит гипотеза Планка? Приведите и поясните формулу Планка для кванта излучения.
6. В чем состоит двойственность природы света? Приведите и поясните формулу, связывающую импульс и длину световой волны.
7. В чем состоит гипотеза де Бройля? Приведите и поясните формулу де Бройля.

8. Какую роль играет уравнение Шрёдингера? В чем смысл функции, относительно которой записывается уравнение Шрёдингера?

9. Приведите и поясните соотношение неопределенностей Гейзенберга.

10. Что отражают пространственная и энергетическая модели атома?

11. Опишите основы теории Бора.

12. Опишите и поясните энергетическую модель атома водорода. Какое излучение атома называется спонтанным, а какое – вынужденным?

13. В чем состоит энергетический подход при решении задач о движении частиц?

14. Что такое потенциальная яма, потенциальный барьер? В чем состоит туннельный эффект?

15. Чем отличается энергетический спектр частицы, подчиняющейся квантовой механике, от спектра классической частицы, находящейся в потенциальной яме?

5. ЭЛЕМЕНТЫ ФИЗИКИ ТВЕРДОГО ТЕЛА

5.1. Кристаллические решетки

Многие свойства твердых тел зависят от их характерных размеров. При изучении объемных материалов микроскопические детали усредняются. В традиционных областях физики – механике, электромагнетизме, оптике, имеющих дело с макромасштабами, изучаются образцы *на макроскопическом уровне* с размерами от миллиметров до километров. Свойства таких материалов – это усредненные характеристики – плотность и модуль Юнга в механике, электрическое сопротивление и диэлектрическая проницаемость в оптике, намагниченность в электромагнетизме.

Поведение материалов можно изучать и на *мезоскопическом уровне*, неоднородности свойств на котором проявляются,

например, в пределах зерна кристалла. Следующие, более низкие, уровни размеров: *микроскопический* (мкм), интервал от 1 до 100 нм характеризует *нанометровый диапазон*. Ниже него лежит атомный масштаб порядка 0,1 нм, еще ниже – ядерный, порядка фемтометра (10^{-15} м).

Большинство твердых тел имеют кристаллическую структуру с атомами, расположенными упорядоченным образом. У них присутствует так называемый дальний порядок, так как упорядоченность распространяется на весь кристалл. В противоположность этому аморфные материалы, такие как стекло и другие, не имеют дальнего порядка, зато имеют ближний порядок, так как локальное окружение каждого атома подобно окружению других таких же атомов, но этот порядок не сохраняется на сколько-нибудь значительных расстояниях. В жидкостях также наблюдается ближний порядок при отсутствии дальнего порядка. У газов нет ни дальнего, ни ближнего порядка.

В твердых телах многие фазы являются кристаллами, т.е. имеют дальний порядок. При нормальных условиях все твердые металлы, а также большинство природных минералов относят к категории кристаллических материалов. Наличие дальнего порядка обуславливает повторяющуюся картину расположения атомов в пределах всего кристалла. Минимальный объем, повторяющийся в кристалле во всех трех измерениях, трансляцией которого можно полностью воспроизвести структуру кристалла, называют *элементарной ячейкой*. Совокупность элементарных ячеек представляет собой *кристаллическую решетку*.

Во всех материалах существует межатомное взаимодействие, которое приводит к образованию атомных связей. При сближении между атомами начинают взаимодействовать электронные оболочки атомов, т.е. возникают силы отталкивания. На рис. 5.1 показано, как меняется энергия притяжения ($E_{\text{прит}}$) и энергия отталкивания ($E_{\text{отт}}$) при сближении противоположно заряженных ионов (электрически заряженных частиц, образующихся при отрыве или присоединении одного или нескольких электронов к атому).

Если просуммировать обе составляющие энергии, то получается известная зависимость полной энергии взаимодействия (сплошная линия) от расстояния между ионами. Кривая полной энергии имеет минимум, определяющий значение межатомного расстояния a . Здесь E_∞ – потенциальная энергия ионов при их расположении на бесконечно большом удалении друг от друга; разность $E_\infty - E_{\min}$ соответствует энергии (теплоте) сублимации, поскольку именно такая энергия должна быть подведена для полного разделения двух атомов или ионов, что соответствует переходу вещества из кристаллического состояния в парообразное минуя плавление.

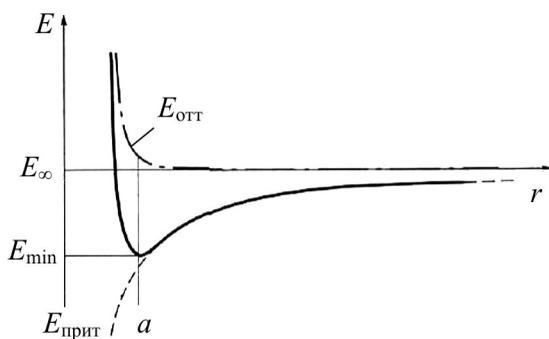


Рис. 5.1. Изменение энергии E взаимодействия противоположно заряженных ионов в зависимости от расстояния r между ними

Межатомное взаимодействие может быть описано моделью твердых шаров определенного радиуса (рис. 5.2, *а*). Более точной является пружинная модель межатомных расстояний, изображенная на рис. 5.2, *б*. При превышении равновесного расстояния пружины растягиваются, и атомы находятся под воздействием сближающей силы. Если расстояние между атомами уменьшается, то создается отталкивающая сила сжатой пружины. Пружинная модель более точно описывает колебания атомов, тепловое расширение, энергию активации и другие факторы.

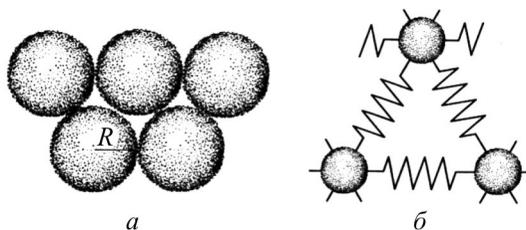


Рис. 5.2. Моделирование связи атомов: a – модель твердых шаров; R – радиус шара; b – пружинная модель

Возможны различные способы упорядоченного расположения или компоновки атомов в кристалле. Эти компоновки называют *решетками Бравэ*.

Для кубической решетки Бравэ возможны три системы упаковки атомов. Простая кубическая упаковка, в которой атомы занимают восемь вершин кубической элементарной ячейки, показана на рис. 5.3, a . Простая кубическая решетка является неплотной,

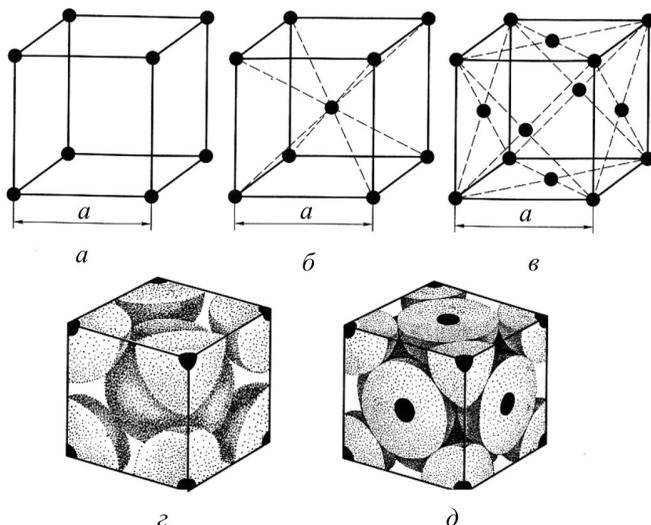


Рис. 5.3. Кристаллические решетки кубического типа (a – e) и заполнение атомами элементарных ячеек плотноупакованных кубических решеток (z , d): a – простая кубическая; b , z – объемно-центрированная (ОЦК); c , d – гранецентрированная (ГЦК)

так как в ней свободны от атомов центр куба и центры граней. Более плотные кубические упаковки обеспечиваются размещением дополнительных атомов по одному из этих вариантов, и их соответственно называют *объемно-центрированной (ОЦК)* и *гранецентрированной (ГЦК) кубическими решетками*. Элементарные ячейки таких решеток показаны на рис. 5.3, б, в, а заполнение атомами этих ячеек в модели твердых шаров приведено на рис. 5.3, г, д. Межатомное расстояние в ячейке a называют *периодом кристаллической решетки*.

Простейшие кристаллы состоят из атомов только одного элемента. В соединениях из двух элементов атомы кристаллизуются в решетки, сдвинутые относительно друг друга по диагонали куба. Например, для соединений ZnS (цинковая обманка) каждый атом цинка (светлый кружок на рис. 5.4) находится в центре тетраэдра из атомов серы (темные кружки), и, наоборот, каждый атом серы имеет в качестве ближайших соседей четыре атома цинка.

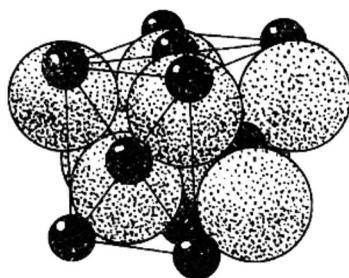


Рис. 5.4. Упаковка больших атомов серы и меньших атомов цинка в структуре цинковой обманки ZnS (сфалерита). Каждый атом расположен в центре тетраэдра из атомов другого типа

Особенностью кристаллов является *анизотропия* – это зависимость свойств от направления. Обусловлена она различием упаковки частиц в разных кристаллографических плоскостях. Наиболее сильно она проявляется в кристаллах со структурами, обладающими малой симметрией.

Анизотропия свойств кристаллов проявляется при использовании монокристаллов. В поликристаллических материалах, состоящих из множества мелких различно ориентированных кристаллов, анизотропия не проявляется. Если же в процессе обработки поликристаллических материалов происходит тек-

стирование их структуры (т.е. ориентированное расположение однотипных участков структуры), то эти материалы также обладают анизотропией.

Жидкие кристаллы – это жидкости с упорядоченной молекулярной структурой. К ним относят некоторые органические вещества, состоящие из удлинённых (одномерных) или плоских (двумерных) молекул. Они занимают промежуточное положение между кристаллами и обычными жидкостями: обладают текучестью (как обычные жидкости) и анизотропией свойств (как кристаллы). Такие свойства и строение эти вещества имеют в определенном интервале температур, выше которого они представляют собой изотропные жидкости (рис. 5.5, *а*).

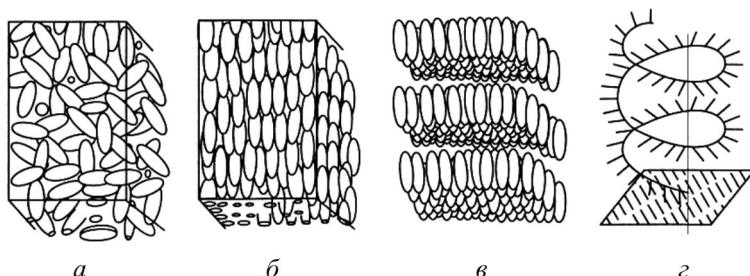


Рис. 5.5. Схемы строения молекулярной жидкости (*а*) и жидких кристаллов: нематических (*б*), смектических (*в*), холестерических (*г*)

Упорядочение возможно в одном направлении, когда молекулы выстроены в цепочку (рис. 5.5, *б*). Кристаллы этого типа относят к группе *нематических*, у которых оптическая ось совпадает с осью жидкого кристалла. В кристаллах второго вида (*смектических*) молекулы образуют параллельные слои (рис. 5.5, *в*), легко смещающиеся относительно друг друга. Кристаллы третьего вида (*холестерические*) имеют наиболее сложное строение: их молекулы размещаются по пространственной спирали (рис. 5.5, *г*).

Структура жидких кристаллов легко изменяется под действием давления, электрического поля, с изменением температуры,

что позволяет легко управлять их структурой и свойствами и делает эти материалы незаменимыми для изготовления особо чувствительных индикаторов.

5.2. Дефекты кристаллического строения

В реальных кристаллах всегда имеются дефекты. Дефекты кристаллического строения подразделяют по геометрическим признакам на точечные, линейные, поверхностные и объемные. Размеры точечного дефекта близки к межатомному расстоянию. У линейных дефектов длина на несколько порядков больше ширины; у поверхностных дефектов мала толщина, а ширина и длина больше ее на несколько порядков. Объемные дефекты имеют значительные размеры во всех трех направлениях.

К самым простым *точечным дефектам* относят *вакансии* и *межузельные атомы* (рис. 5.6). Вакансия представляет собой пустой узел кристаллической решетки; межузельным атомом называют атом, перемещенный из узла в позицию между узлами. Вакансии и межузельные атомы появляются в кристаллах при любой температуре выше абсолютного нуля из-за тепловых колебаний атомов.

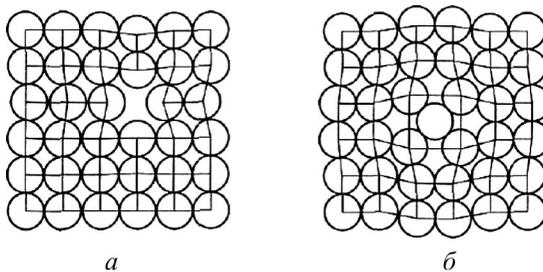


Рис. 5.6. Точечные дефекты в кристаллической решетке: *a* – вакансия; *б* – межузельный атом

Пересыщение точечными дефектами возникает при резком охлаждении после высокотемпературного нагрева, при пластическом деформировании и при облучении нейтронами. В послед-

нем случае концентрация вакансий и межузельных атомов одинакова: выбитые из узлов решетки атомы становятся межузельными, а освободившиеся узлы становятся вакансиями. С течением времени избыток вакансий сверх равновесной концентрации исчезает на свободных поверхностях кристалла, порах, границах зерен и других дефектах решетки. Места, где исчезают вакансии, называют *стоками вакансий*. Убыль вакансий объясняется их подвижностью и непрерывным перемещением в решетке. Соседний с вакансией атом может занять ее место и оставить свободным свой узел, в который затем переходит другой атом.

Чем выше температура, тем больше концентрация вакансий и тем чаще они переходят от узла к узлу. Вакансии являются самой важной разновидностью точечных дефектов; они ускоряют все процессы, связанные с перемещениями атомов: диффузию, спекание порошков и т.д.

Точечные дефекты вызывают местное искажение кристаллической решетки, распространяющееся обычно на несколько соседних атомных слоев. Искажение вокруг межузельных атомов в плотноупакованных решетках значительно больше, чем вокруг вакансий.

Линейные дефекты имеют малые размеры в двух измерениях и большую протяженность в третьем измерении. К ним относятся краевые и винтовые дислокации.

Краевая дислокация (рис. 5.7, а–в) представляет собой локализованное искажение кристаллической решетки, вызванное наличием в ней «лишней» атомной полуплоскости или экстраплоскости. Наиболее простой и наглядный способ образования дислокации в кристалле – сдвиг на одно межатомное расстояние одной части кристалла относительно другой.

Плоскость, в которой произошел сдвиг, называют *плоскостью скольжения*, а линия пересечения экстраплоскости с плоскостью скольжения является линией дислокации, вдоль которой и локализуется искажение кристаллической решетки.

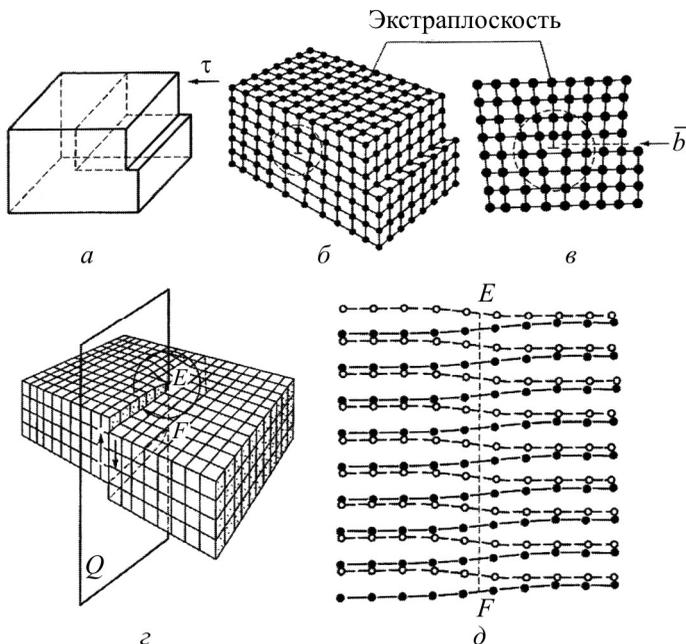


Рис. 5.7. Краевая ($a - в$) и винтовая ($г, д$) дислокация: a – сдвиг, создавший краевую дислокацию; τ – сдвиговое напряжение, стрелками показано направление сдвига; $б$ – пространственная схема краевой дислокации; $в$ – схема расположения атомов у краевой дислокации: \bar{b} – вектор Бюргерса; $г$ – пространственная модель образования винтовой дислокации EF в результате неполного сдвига по плоскости Q ; $д$ – расположение атомов в области винтовой дислокации

Размеры дефекта перпендикулярно линии дислокации невелики и не превышают пяти межатомных расстояний. Дислокационные линии не обрываются внутри кристалла, они выходят на его поверхность, заканчиваются на других дислокациях или образуют замкнутые дислокационные петли.

Одним из параметров, характеризующих поведение дислокации, является *вектор Бюргерса* \bar{b} , показывающий величину и направление сдвига в процессе скольжения. Он характеризует

степень искажения кристаллической решетки вокруг дислокации (упругая энергия искажения пропорциональна квадрату вектора Бюргерса).

Кроме краевых дислокаций, различают еще *винтовые дислокации*. На рис. 5.7, *г* показана пространственная модель винтовой дислокации. Это прямая линия *EF* (рис. 5.7, *д*), вокруг которой атомные плоскости изогнуты по винтовой поверхности. Винтовая дислокация образована неполным сдвигом кристалла по плоскости *Q*. В отличие от краевой дислокации винтовая дислокация параллельна вектору сдвига.

Вокруг дислокаций на протяжении нескольких межатомных расстояний возникают искажения решетки. Энергия искажения является одной из важнейших характеристик дислокации любого типа. Критерием этого искажения служит вектор Бюргерса. В краевой дислокации вектор Бюргерса перпендикулярен линии дислокации, а в винтовой – параллелен ей.

Дислокации образуются при кристаллизации металлов, а также в процессах пластической деформации и фазовых превращениях. Дислокации присутствуют в кристаллах в огромном количестве ($10^6 \dots 10^{12} \text{ см}^{-2}$) и обладают легкой подвижностью и способностью к размножению.

Поверхностные дефекты малы только в одном измерении. Они представляют собой поверхности раздела между отдельными *зернами* или *субзернами* в поликристаллическом материале. Субзерна размерами 0,1...1,0 мкм разориентированы относительно друг друга, границы между ними имеют дислокационный характер и представляют собой дислокационные стенки.

Объемные дефекты – это поры, трещины, усадочные дефекты и т.п. Они образуются при кристаллизации, фазовых превращениях, деформации и других процессах.

5.3. Элементы зонной теории

При связывании атомов и формировании твердого тела энергетические уровни отдельных атомов расширяются и образуют зоны с щелями между ними. Если изолированные атомы имеют совпадающие схемы энергетических уровней, то при сжатии атома в кристаллической решетке энергетические уровни деформируются, смещаются и расширяются в зоны, образуя *зонный энергетический спектр* (рис. 5.8). При этом уровни внешних *валентных* электронов расширяются, а уровни внутренних электронов расщепляются слабо.

Электроны могут иметь значения энергии, только лежащие внутри какой-либо зоны (на рис. 5.8 заштрихованы). Таким образом, из энергетического спектра электронов можно выделить три зоны: *валентную*, образованную из энергетических уровней внутренних электронов свободных атомов и полностью заполненную электронами, в нее попадают электроны, связывающие кристалл в единое целое; зону *проводимости* (свободную зону), частично заполненную электронами, образованную из энергетических уровней «коллективизированных электронов; *запрещенную* зону с шириной энергетической щели ΔE . Зона проводимости расположена по энергиям выше валентной зоны и отделена от нее запрещенной зоной.

В соответствии с взаимным расположением зон все твердые тела подразделяются на *диэлектрики*, *полупроводники* и *проводники*. У *диэлектриков* валентная зона заполнена электронами, которые не могут перемещаться, так как их положения зафиксированы в химических связях. Зона проводимости расположена

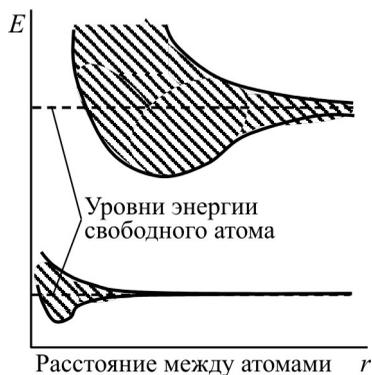


Рис. 5.8. Уровни энергии свободного атома

по энергиям намного выше валентной зоны, как показано на рис. 5.9, *а*, так что она недоступна тепловым флуктуациям и остается практически пустой. Другими словами, при комнатной температуре $T = 300$ К тепловой энергии недостаточно для переброски сколько-нибудь значительной части электронов из валентной зоны в зону проводимости, так что их количество в последней пренебрежимо мало. Еще один вариант изложения этого факта таков: ширина энергетической щели ΔE много больше величины тепловой энергии kT , где k – постоянная Больцмана.

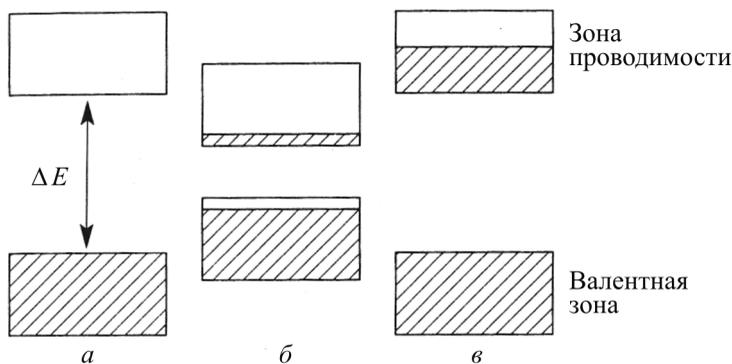


Рис. 5.9. Энергетические зоны в диэлектрике (*а*), собственном полупроводнике (*б*) и проводнике (*в*). Наличие электронов в зоне обозначено косой штриховкой

У *полупроводников* щель между валентной зоной и зоной проводимости много меньше, как показано на рис. 5.9, *б*, так что $\Delta E \sim 1$ эВ ближе к тепловой энергии $kT \sim 10^{-4}$ эВ. При комнатной температуре тепловое возбуждение может забросить часть электронов из валентной зоны в зону проводимости, где они принимают участие в переносе заряда, то есть в электрическом токе. Этот процесс называют *генерацией* тока. Процесс обратного перехода электрона из зоны проводимости в валентную зону называют *рекомбинацией*. Плотность электронов, достигающих зоны проводимости в процессе теплового возбуждения, относительно мала, но ей уже нельзя пренебречь, так что возникает

небольшая проводимость, отсюда и термин – полупроводник. Материалы такого типа называют *собственными полупроводниками*.

Полупроводник может быть легирован *донорами*, которые легко отдают электроны в зону проводимости, где те переносят ток. Материал может быть легирован и *акцепторами*, захватывающими электроны из валентной зоны и оставляющими в ней эффективные положительные заряды, называемые дырками, которые также являются носителями тока. Энергетические уровни таких доноров и акцепторов лежат внутри запрещенной зоны, как показано на рис. 5.10. В первом случае говорят о проводимости *n*-типа, так как основными носителями тока являются отрицательно (*negative*) заряженные электроны, а во втором – *p*-типа, так как основные носители – положительно (*positive*) заряженные вакансии (дырки). Материалы такого типа обладают *примесной* (несобственной) проводимостью. Акцепторные уровни расположены выше потолка валентной зоны на ΔE_A . Донорные уровни лежат ниже дна зоны проводимости на ΔE_D . Уровни глубоких ловушек электронов, которыми могут быть, например, дефекты кристаллического строения, лежат вблизи середины запрещенной зоны.

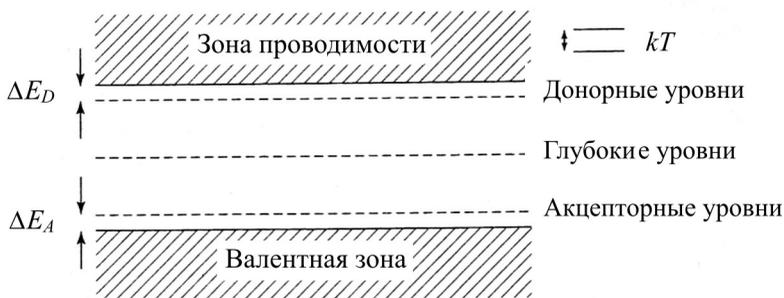


Рис. 5.10. Энергетическая схема локальных уровней в запрещенной зоне полупроводника

Проводники – это материалы с заполненной валентной зоной и зоной проводимости, частично заполненной делокализо-

ванными электронами, выступающими эффективными носителями электрического тока. Положительно заряженные ионы металлов, расположенные в узлах кристаллической решетки, отдают свои внешние электроны в зону проводимости и остаются положительно заряженной основой для обобществленных электронов. На рис. 5.9, в показана зонная структура для этого случая.

Образование зонного энергетического спектра является *квантово-механическим эффектом* и вытекает из соотношения неопределенностей. Валентные электроны переходят от атома к атому сквозь потенциальный барьер (туннельный эффект). Время жизни связано с неопределенностью энергии

$$\Delta E \sim h/\Delta\tau.$$

Среднее время жизни валентного электрона уменьшается от 10^{-8} с (изолированный атом) до 10^{-15} с, при этом ширина спектральных линий увеличивается с естественной 10^{-7} эВ до значений в кристалле $\Delta E \sim 1-10$ эВ.

5.4. Энергетический спектр кристалла

Энергетический спектр кристалла непосредственно связан с энергетическим спектром тех атомов, которые входят в его состав. В состав реального кристалла входят как собственные, так и примесные атомы. В производстве оптических волокон, интегральных микросхем используют достаточно чистые кристаллы. В них содержание примесных атомов обычно не превышает 10^{-6} % от концентрации собственных атомов ($\sim 5 \cdot 10^{22}$ см⁻³). Примесные атомы находятся в кристалле в среднем на таких расстояниях друг от друга, что практически не взаимодействуют между собой. Их энергетические спектры подобны спектрам свободных атомов, т.е. дискретны, хотя и изменены воздействием собственных атомов кристалла. Примесные уровни определяют тип и величину электропроводности кристалла и влияют на характеристики приборных микро-, опто- и наноэлектронных

структур. Взаимодействие собственных атомов друг с другом существенно изменяет их энергетический спектр. *Высоко расположенные дискретные энергетические уровни изолированных атомов превращаются (при объединении атомов в кристалл) в широкие энергетические полосы – зоны. Коренным образом изменяется и зависимость потенциальной энергии электрона от координат $U(x)$; она становится периодической.*

Одномерная энергетическая модель кристалла схематически представлена на рис. 5.11. Соседние атомы кристалла так изменяют потенциальное поле друг друга, что оно превращается в периодическую совокупность потенциальных барьеров и потенциальных ям. Функция $U(x)$ становится периодической. Решение уравнения Шрёдингера для такого случая (независимо от конкретных особенностей периодического потенциала) всегда дает *зонный энергетический спектр*. Внешние электроны атомов принадлежат всему кристаллу, так как легко туннелируют сквозь потенциальные барьеры (см. рис. 5.11).

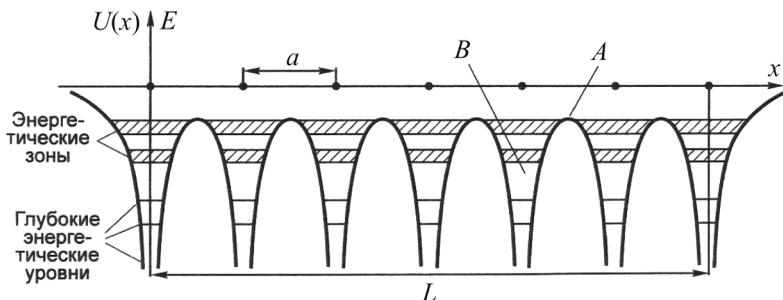


Рис. 5.11. Одномерная энергетическая модель кристалла:
 a – межатомное расстояние; L – общий размер кристалла;
 A – потенциальный барьер, ограничивающий переход электронов от одного атома к соседнему;
 B – потенциальная яма

Связь зонного энергетического спектра кристалла с дискретным энергетическим спектром атомов, из которых состоит кристалл, можно пояснить с помощью мысленной процедуры

образования кристалла из первоначально разнесенных на большое расстояние друг от друга атомов.

Основные электрические и оптические свойства кристаллов определяются особенностями разрешенных зон: валентной зоны и зоны проводимости, а также энергетическим промежутком ΔE между этими зонами – запрещенной зоной. Если запрещенная зона отсутствует ($\Delta E = 0$), то соответствующее вещество относится к металлам (проводникам). К металлам относятся и такие вещества, для которых $\Delta E > 0$, но не все квантовые состояния валентной зоны заняты электронами. У непроводников (диэлектриков и полупроводников) $\Delta E > 0$, квантовые состояния валентной зоны при абсолютном нуле температур ($T = 0$) полностью заняты электронами, а зона проводимости электронов их не содержит.

Если $0 < \Delta E < 3 \text{ эВ}$, то вещество относится к полупроводникам. Для диэлектриков $\Delta E > 3 \text{ эВ}$. Такое разделение полупроводников и диэлектриков условно. При анализе электрических и оптических процессов в кристаллах обычно используют лишь две разрешенные и одну запрещенную зоны (если $\Delta E > 0$), как это показано на рис. 5.9, 5.10.

Валентная зона в проводниках не полностью занята электронами. Самый верхний энергетический уровень (E_F), который в металлах (при $T = 0$) занят электроном, называется уровнем Ферми. Валентные электроны не локализованы вблизи отдельных атомов, а свободно перемещаются по всему кристаллу, подобно молекулам газа в некотором сосуде. Систему электронов в проводящих кристаллах называют электронным газом или электронной жидкостью. Уровень Ферми выполняет для электронной жидкости в кристалле ту же роль, что и уровень жидкости в сообщающихся сосудах. Если привести в соприкосновение два кристалла с различными уровнями Ферми, то электроны будут «перетекать» из одного кристалла в другой до тех пор, пока не выровняются уровни Ферми. Строгое определение смысла уровня Ферми дается в термодинамике.

5.5. Понятие эффективной массы электрона

Под действием силы внешнего электрического поля, $F = eE$, собственные и примесные электроны проводимости приобретают ускорение (a) и скорость направленного движения, образуя электрический ток в полупроводнике (E – напряженность поля). Это не означает, однако, что для определения ускорения (a) можно непосредственно пользоваться вторым законом Ньютона ($a = F/m$). Дело в том, что характер движения электрона в кристалле определяется не только внешней силой, но и воздействием на электрон атомов кристаллической решетки. В квантовой теории показано, что воздействие атомов кристалла на движение электрона можно формально учесть, заменив массу электрона на другую физическую величину, которая называется эффективной массой и обозначается m^* . При такой замене

$$a = F/m^* . \quad (5.1)$$

Эффективная масса электрона может изменяться от значения, близкого к массе электрона (у границ зоны), до бесконечности (в середине зоны), и быть как положительной (в нижней части зоны), так и отрицательной (в верхней части зоны).

Электроны с отрицательной эффективной массой перемещаются в сторону, противоположную направлению действия внешней силы. Именно такое движение электронов формально учитывается введением фиктивных положительных носителей тока (*дырок*). Если в верхней части валентной зоны все квантовые состояния заняты электронами, то их поток против внешнего поля ($m^* < 0$) в точности компенсируется потоком (по полю) электронов нижней половины валентной зоны ($m^* > 0$). Этим объясняется отсутствие электропроводности ($\sigma = 0$) у полупроводников, если $T = 0$. При появлении в верхней части валентной зоны свободных квантовых состояний ($T > 0$) обратный поток электронов становится меньше прямого и в кристалле возникает электрический ток. Величина этого тока такова, как

если бы он создавался положительными зарядами в количестве, равном числу свободных квантовых состояний в верхней части валентной зоны. Эти формально введенные положительные электрические заряды и называются дырками. Очевидно, что их эффективная масса положительна; они ускоряются в направлении действия внешнего поля.

Введение понятия эффективной массы электронов и дырок облегчает объяснение особенностей функционирования многих микро-, опто- и наноструктур.

Описанная ситуация в квантовой теории имеет и классические аналоги. Предположим, что мы рассматриваем движение тела массой m в поле силы тяжести F_T . В свободном состоянии его ускорение $a(=g) = F_T/m$ (где g – ускорение силы тяжести). В жидкости движущая сила уменьшится на величину выталкивающей силы Архимеда F_A , и ускорение определится соотношением $a = (F_T - F_A)/m$. Учитывая, что $F_T = mg = \rho V g m F_A = \rho_j V g$ (ρ и ρ_j – плотности тела и жидкости соответственно, V – объем тела), можно окончательно записать

$$a = F_T / m^*,$$

где $m^* = m / (1 - \rho_j / \rho)$.

Видно, что влияние жидкости на движение тела под действием силы тяжести F_T удобно учитывать введением эффективной массы m^* , которая может принимать как положительные значения (тело тонет), так и отрицательные (тело всплывает, т.е. ускоряется против силы тяжести). Если $|m^*| = \infty$, то $\rho_j = \rho$, тело будет оставаться неподвижным при любом конечном значении F_T .

5.6. Экситонные эффекты

Фотоны с энергией, равной или превышающей ширину запрещенной зоны полупроводника ΔE , могут создавать электронно-дырочные пары. Обычно возникшие электрон и дырка дальше двигаются независимо друг от друга, однако в некото-

рых случаях благодаря кулоновскому взаимодействию между ними электрон и дырка могут оставаться «вместе», формируя новую электрически нейтральную квазичастицу, похожую на атом водорода, которую называют *экситоном* (от лат. *excite* – возбуждаю). Такие частицы не влияют на электропроводность вещества, поскольку не обладают электрическим зарядом. Размер экситона составляет несколько параметров решетки и сравним с типичными размерами наноструктур. Представление об экситоне введено в 1931 году Я.И. Френкелем для объяснения отсутствия фотопроводимости некоторых кристаллов: при поглощении света энергия расходуется не на создание носителей заряда, а на образование экситонов.

Простейшая модель описывает экситон как электрон и дырку, которые вращаются внутри решетки относительно общего центра масс под воздействием кулоновского притяжения, как это показано на рис. 5.12. Экситон имеет свойства частицы, он способен перемещаться по кристаллу и имеет характерный оптический спектр.

Существуют два основных типа экситонов:

1) экситоны с незначительным перекрытием волновых функций электронов и дырок, радиусы которых охватывают множество атомов кристалла. Такие системы, характерные для полупроводников, называются экситонами Ванье–Мотта. Они распространяются по кристаллу в виде волны, вызывая «всплески» в спектре поглощения (экситонный резонанс);

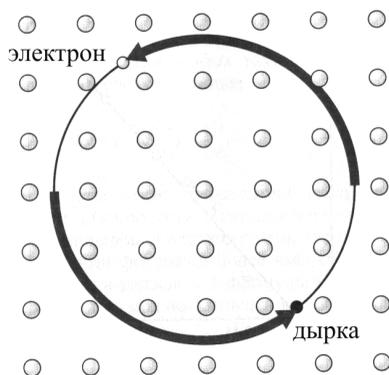


Рис. 5.12. Представление экситона в виде связанного состояния электронно-дырочной пары

2) экситоны, радиус которых имеет порядок постоянной решетки, характерны в основном для изоляторов (экситоны Френкеля).

Энергия фотона, необходимая для создания экситона, меньше энергии ΔE , необходимой просто для создания независимой электронно-дырочной пары в полупроводнике, поэтому полный процесс можно описать как создание экситона с его дальнейшим разделением на электрон и дырку. При этом второй процесс требует энергии, равной энергии связи экситона ΔE_i . Поэтому, как показано на рис. 5.13, связанные состояния экситона, расположены в запрещенной зоне полупроводника, чуть ниже границы зоны проводимости.

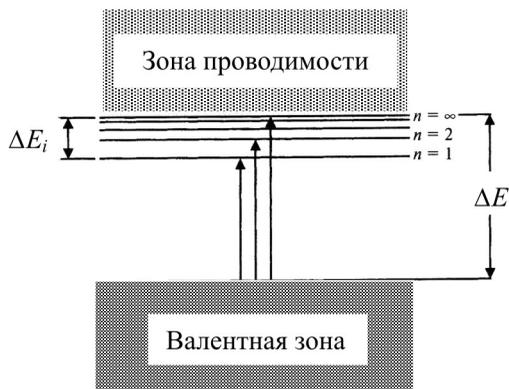


Рис. 5.13. Связанные состояния экситона
вблизи границы зоны проводимости

При малых концентрациях экситоны ведут себя в кристалле подобно газу. Если создающее экситоны излучение является достаточно мощным, то при больших концентрациях становится существенным их взаимодействие и возможно образование связанного состояния двух экситонов – экситонной молекулы – *би-экситона*. При критических концентрациях (зависящих от температуры) происходит «сжижение» экситонного газа, обладающего большой подвижностью в неоднородных полях.

Вопросы для самоконтроля

1. На какие группы по уровням осреднения свойств обычно подразделяют конструкционные материалы?
2. Какие модели силового взаимодействия атомов в кристаллической решетке вы знаете?
3. В чем состоит различие кристаллического и аморфного состояний в материалах?
4. Что называется энергией сублимации?
5. Какие типы кристаллических решеток вы знаете? Каковы их основные характеристики?
6. Что такое анизотропия?
7. Какие типы жидких кристаллов вы знаете?
8. Каковы основные дефекты кристаллического строения? Чем отличается линейная дислокация от винтовой дислокации? Что характеризует вектор Бюргера?
9. Каковы особенности зонного энергетического спектра изолированных атомов?
10. Охарактеризуйте классификацию твердых тел по зонной теории.
11. Какие процессы называют генерацией и рекомбинацией носителей тока в полупроводнике?
12. Механизмы собственной и примесной проводимости полупроводников.
13. Что характеризует уровень Ферми в энергетическом спектре кристалла?
14. Какая особенность движения носителей тока в кристалле под действием внешнего электрического поля учитывается введением эффективной массы электрона?
15. Что такое экситон, какие типы экситонов вам известны?

6. ФИЗИЧЕСКИЕ ОСНОВЫ ОПТИКИ

6.1. Электромагнитная природа света

Для описания свойств света привлекают упрощенные модели, чтобы в первом приближении понять, как именно испускают свои лучи источники света, будь то Солнце или лазер. Световые лучи согласно квантовой теории могут восприниматься как прямолинейные траектории частиц света, или фотонов, исходящих от источников видимого излучения. Но если попытаться, пользуясь точечной диафрагмой, получить резко ограниченный луч, то позади диафрагмы будут отмечены явления дифракции, вызывающие расходимость пучка относительно диаметра отверстия. Поэтому резкий световой луч создать не удастся. Это объясняется волновой моделью света. В принципе, корпускулярная и волновая модели могут быть объединены на уровне унифицированного теоретического описания, которое, однако, требует достаточно глубоких математических знаний. Для объяснения многих интересующих нас явлений вполне достаточно только волновой либо только корпускулярной модели – опыт показывает, какая из них наиболее приемлема в том или ином случае. Так, для понимания поглощения и излучения света лучше прибегнуть к представлению частиц, а для описания распространения света и явлений интерференции можно привлечь волновое поле.

Волновая оптика описывает видимое излучение как поперечную электромагнитную волну, в которой периодически и с равной частотой ν возбуждается напряженность электрического поля E и связанная с ним напряженность магнитного поля H . Векторы E , H и направление распространения всегда взаимно ортогональны. На рис 6.1 представлены напряженности поля в определенный момент времени в зависимости от координаты места в направлении распространения.

Чтобы получить представление о пространственном распространении световых волн, рассматриваются их фазовые поверхности, или плоскости – например, места максимальных на-

пряженностей поля. Интервал между двумя соседними фазовыми поверхностями составляет длину волны λ . При плоской волне фазовые поверхности параллельны.

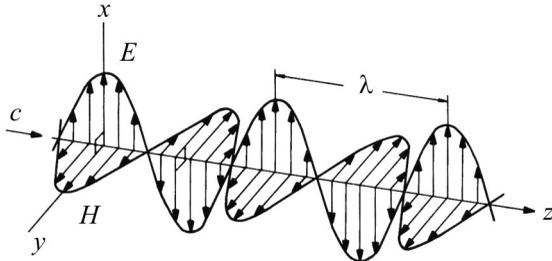


Рис. 6.1. Напряженность электрического (E) и магнитного (H) полей в световой волне в фиксированный момент времени. Волна распространяется в направлении z

Между частотой ν , длиной волны λ и скоростью распространения c существует следующая зависимость:

$$c = \lambda \cdot \nu. \quad (6.1)$$

В вакууме скорость света $c = 2,998 \cdot 10^8$ м/с. Для большинства оптических явлений достаточно рассмотреть напряженность электрического поля. Напряженность поля световой волны не поддается прямому измерению. Вместо этого можно определять интенсивность либо энергетическую освещенность I , выражаемую через временное среднее значение квадрата амплитуды напряженности поля E :

$$I = \overline{E^2} / \sqrt{\mu\mu_0 / \epsilon\epsilon_0}, \quad (6.2)$$

где $\epsilon_0 = 8,854 \cdot 10^{-12}$ Ф/м – абсолютная диэлектрическая проницаемость; ϵ – относительная диэлектрическая проницаемость; $\mu_0 = 4\pi \cdot 10^{-7}$ Гн/м – абсолютная магнитная проницаемость; μ – относительная магнитная проницаемость. Горизонтальная черта над E^2 символизирует значение, осредненное по времени. Единицей измерения напряженности электрического поля является В/м,

а единицей измерения плотности мощности служит Вт/м². Коэффициент пропорциональности $Z = \sqrt{\mu\mu_0/\epsilon\epsilon_0}$ обладает размерностью сопротивления и обозначается поэтому как «волновое сопротивление». Для вакуума и воздуха ($\epsilon = 1, \mu = 1$) $Z = 377$ В/А.

В прозрачной среде свет распространяется медленнее, чем в вакууме. Скорость света v в такой среде определяется как $v = c/n$. Постоянная вещества n именуется показателем преломления и выражается через относительную диэлектрическую и магнитную проницаемости среды:

$$n = \sqrt{\epsilon \cdot \mu} . \quad (6.3)$$

При попадании света на границу раздела между двумя оптическими средами с показателями преломления n_1 и n_2 углы падения α_1 и преломления α_2 связаны известным *законом преломления* (закон Снеллиуса):

$$n_1 \cdot \sin \alpha_1 = n_2 \cdot \sin \alpha_2 , \quad (6.4)$$

из которого следует, что при возрастании угла падения увеличивается и угол преломления. В случае падения луча из оптически более плотной среды в менее плотную ($n_1 > n_2$) угол преломления достигает максимального значения $\alpha_2 = 90^\circ$, т.е. луч после падения под предельным углом ($\alpha_1 = \alpha_{\text{пр}}$) на поверхность раздела сред будет распространяться по касательной к поверхности в точке падения. Из уравнения (6.4) следует, что

$$\sin \alpha_{\text{пр}} = n_2/n_1 . \quad (6.5)$$

Последующее увеличение угла падения приводит к отражению луча. Описанное явление называют *полным внутренним отражением*, а угол, определяемый равенством (6.5), – *предельным углом полного внутреннего отражения*. Например, на границе стекло ($n_1 = 1,5$), воздух ($n_2 = 1$) $\alpha_{\text{пр}} = \arcsin 1/1,5 \approx 42^\circ$. Примеры полного внутреннего отражения представлены на рис. 6.2.

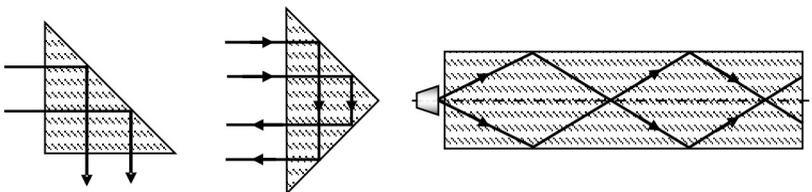


Рис. 6.2. Схемы хода лучей в поворотных призмах и световоде при полном внутреннем отражении

Простейшая волна, описываемая тригонометрическими функциями, такими как синус и косинус, называется гармонической волной (рис. 6.3):

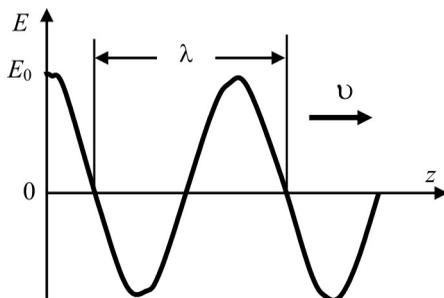


Рис. 6.3. Распространяющаяся волна

$$E = E_0 \cos \frac{2\pi}{\lambda} (vt - z), \quad (6.6)$$

где E_0 – амплитуда волны; z – направление ее распространения; λ – расстояние между ее минимумами или *длина волны*. Аргумент косинуса называется *фазой волны* φ . Продифференцировав постоянное значение фазы по времени,

$$\frac{d\varphi}{dt} = \frac{d}{dt} \left[\frac{2\pi}{\lambda} (vt - z) \right] = 0, \quad (6.7)$$

получим скорость $v = dz/dt$, с которой перемещается точка постоянной фазы. Она называется *фазовой скоростью волны*. За время

одного колебания в какой-то одной точке, например в точке $z = 0$ (его называют *периодом колебания* T), волна проходит расстояние λ , поэтому фазовая скорость

$$v = \frac{\lambda}{T} = \lambda \cdot \nu, \quad (6.8)$$

где ν – величина, обратная периоду, характеризующая число колебаний в единицу времени, и называется *частотой* волны. В оптике и спектроскопии применяют волновое число k и круговую (циклическую) частоту ω , определяемые выражениями

$$k = \frac{2\pi}{\lambda}, \quad \omega = 2\pi\nu. \quad (6.9)$$

Через эти параметры можно выразить и фазовую скорость, $v = \omega/k$, тогда уравнение, описывающее волну, принимает вид

$$E = E_0 \cos(\omega t - kz + \varphi_0), \quad (6.10)$$

где φ_0 – начальная фаза волны при $t = 0$ и $z = 0$.

Таким образом, световые волны можно описать *скалярным уравнением* (6.10) в направлении их распространения z . Для скорости света, $c \approx 3 \cdot 10^8$ м/с, и средней длины волны видимого света, $\lambda = 0,55$ мкм, частота видимого света

$$\nu = \frac{c}{\lambda} = 6 \cdot 10^{14} \text{ Гц}. \quad (6.11)$$

Если колебания напряженности электрического поля E происходят в одной плоскости, то световая волна обозначается как линейно-поляризованная, а плоскость Ez – как *плоскость поляризации* света. Свет большинства источников (Солнца, ламп накаливания) не поляризован и может трактоваться как статистическая смесь волн со всеми возможными направлениями поляризации.

Согласно квантовой теории свет демонстрирует как свойства волн, так и свойства частиц. На такую двойственность указывают и экспериментальные наблюдения. В корпускулярном

описании свет состоит из квантов, или фотонов с энергией W , движущихся со световой скоростью c .

$$W = h\nu = hc/\lambda. \quad (6.12)$$

Здесь $h = 6,626 \cdot 10^{-34}$ Дж·с есть постоянная Планка. В ядерной физике и физике лазера энергия фотона часто указывается не в джоулях, а в электрон-вольтах (эВ), причем 1 эВ показывает энергию ($W = eU$, $e = 1,602 \cdot 10^{-19}$ А·с), генерируемую электроном при ускорении его напряжением в 1 вольт (В).

Объемная плотность энергии ρ [Дж/м³] в световой волне выражается через объемную плотность фотонов Φ [м⁻³], а интенсивность – через плотность потока фотонов ϕ [м⁻²с⁻¹]:

$$\rho = h\nu \cdot \Phi, \quad I = h\nu \cdot \phi. \quad (6.13)$$

Видимый свет бывает разного цвета, причем эти цвета могут различаться по частоте или длине волн. К видимому диапазону в коротковолновой области примыкает ультрафиолетовый (УФ)-диапазон, а в длинноволновой – инфракрасный (ИК)-диапазон (табл. 6.1 и 6.2). Солнечный спектр имеет свой максимум в видимой области и примерно соответствует излучению черного тела с температурой 6000 К (рис. 6.4).

Т а б л и ц а 6.1

Длины волн (λ), частота (ν) и энергия фотонов ($h\nu$) электромагнитного излучения

	λ	ν , Гц	$h\nu$
Гамма-излучение	до 500 пм	до $6 \cdot 10^{18}$	24,8 кэВ
Рентгеновское излучение	до 50 нм	до $6 \cdot 10^{15}$	24,8 эВ
Ультрафиолетовое излучение	до 400 нм	до $7,5 \cdot 10^{14}$	3,1 эВ
Видимое излучение	до 700 нм	до $4,3 \cdot 10^{14}$	1,77 эВ
Инфракрасное излучение	до 100 мкм	до $3 \cdot 10^{12}$	12,4 мэВ
Микроволны (СВЧ)	до 1 см	до $3 \cdot 10^{10}$	124 мкэВ
Радиоволны	до 1 км	до $3 \cdot 10^5$	1,24 нэВ

Примечание. Указаны ориентировочные численные значения области спектра.

Таблица 6.2

Длины волн, частота и энергия фотонов
в диапазоне лазерного излучения

	λ , нм	ν , 10^{14} Гц	$h\nu$, эВ
Вакуумный УФ-диапазон	100–200	30–15	12,4–6,2
Дальний УФ-диапазон	200–280	15–10,7	6,2–4,4
Средний УФ-диапазон	280–315	10,7–9,5	4,4–3,9
Ближний УФ-диапазон	315–380	9,5–7,9	3,9–3,3
Свет (видимое излучение)	380–780	7,9–3,9	3,3–1,6
Ближний ИК-диапазон	780–3000	3,9–1,0	1,6–0,4
Средний ИК-диапазон	3000–50000	1,0–0,06	0,4–0,025
Дальний ИК-диапазон	50000–1 мм	0,0–0,003	0,025–0,001

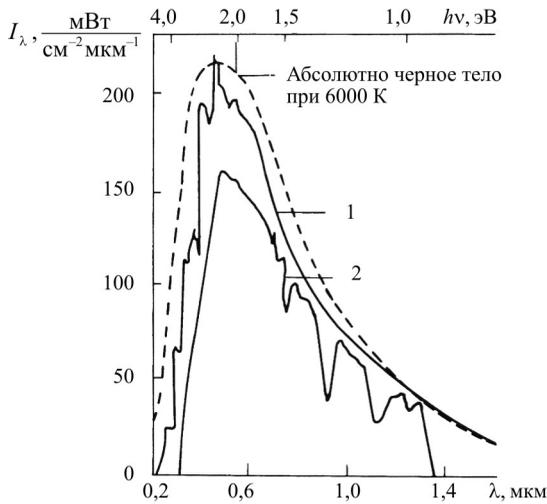


Рис. 6.4. Солнечный спектр в сравнении
с излучением абсолютно черного тела при 6000 К:
1 – спектр без атмосферы; 2 – спектр с атмосферой воздуха

6.2. Основные явления волновой оптики

Интерференция

При суперпозиции (сложении) нескольких световых волн их энергия перераспределяется в пространстве, при этом волновое поле будет содержать периодические светлые и темные участки интенсивности. Это явление называется *интерференцией* (от лат. *inter* – между и *ferentis* – несущий, переносящий).

Рассмотрим суперпозицию двух монохроматических плоских волн, исходящих из одного источника с разностью фаз φ_0 .

$$E_1 = E_0 \cos(\omega t - kz),$$

$$E_2 = E_0 \cos(\omega t - kz + \varphi_0).$$

При сложении двух волн результирующее колебание также происходит на частоте ω и имеет некоторую амплитуду E .

$$E = E_1 \cos(\omega t - kz_1) + E_2 \cos(\omega t - kz_2).$$

При этом интенсивность света, пропорциональная квадрату амплитуды, имеет вид

$$I = E_1^2 + E_2^2 + 2E_1E_2 \cos(k\Delta) = I_1 + I_2 + 2\sqrt{I_1I_2} \cos(k\Delta),$$

где $\Delta = z_2 - z_1$ – *оптическая разность хода лучей*. При $\Delta = m\lambda$, $m = 1, 2, \dots$ имеют место максимумы (усиливающая интерференция), а при $\Delta = (m+1/2)\lambda$, $m = 1, 2, \dots$ – минимумы (ослабляющая интерференция). Таким образом, если средний $\cos(k\Delta)$ постоянен со временем, то интерференция имеет место с тем или иным результатом суперпозиции, определяемым его значением, изменяющимся в пределах от -1 до $+1$. Интенсивность света будет изменяться в пределах от $I_{\max} = (E_1 + E_2)^2 > I_1 + I_2$ до $I_{\min} = (E_1 - E_2)^2 < I_1 + I_2$ (рис. 6.5).

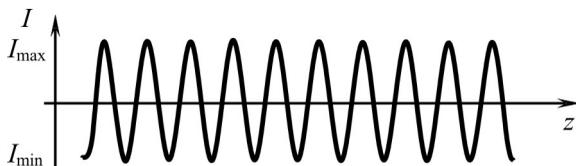


Рис. 6.5. Распределение интенсивности света при двухлучевой интерференции

Такие волны называются *когерентными*. Это весьма жесткое требование к когерентности не имеет места в случае реальных волн, испускаемых реальными источниками, так как излучатели света – атомы – немонахроматические источники. Их излучение в пределах макроисточников света (искусственных или естественных) не согласовано по фазам, а сами источники имеют конечные размеры. Немонахроматичность источников и их протяженность ведут соответственно к понятиям временной и пространственной когерентности, а в целом – к *пространственно-временной когерентности*, которая будет обсуждаться ниже.

Дифракция

Другим явлением, возможным даже с одной световой волной, является *дифракция* (лат. *diffractus* – преломленный) – огибание волной препятствий, встречающихся на ее пути, или, в более широком смысле, – любое отклонение распространения волн вблизи препятствий от законов геометрической оптики.

Дифракционные явления присущи всем волновым процессам, но особенно отчетливо проявляются лишь в тех случаях, когда длины волн излучений сопоставимы с размерами препятствий.

Явление дифракции можно объяснить с помощью *принципа Гюйгенса*, в соответствии с которым плоский фронт волны представляется совокупностью фиктивных источников (рис. 6.6), каждый из которых сам излучает сферическую волну. Распространение этих волн на малое расстояние дает новый волновой фронт. В результате плоский волновой фронт искривляется, последую-

щие фронты приобретают все большую кривизну и, наконец, волновой фронт становится сферическим, а волна – расходящейся. Принцип Гюйгенса не объясняет появление луча света.

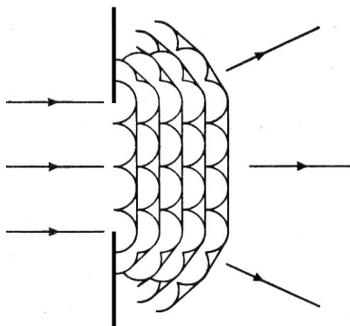


Рис. 6.6. Построение Гюйгенса



Рис. 6.7. Зонная пластинка Френеля

Чтобы объяснить прямолинейное распространение света, принцип Гюйгенса был развит Френелем, который предположил, что световая волна, возбуждаемая каким-либо источником света, может быть представлена как результат интерференции когерентных вторичных волн, «излучаемых» фиктивными источниками. При делении сферического фронта волны на зоны, вторичные волны которых находятся в противофазах, результатом интерференции является фокусировка света в виде луча. Пластинкой с зонами Френеля (рис. 6.7) можно усиливать или ослаблять источник света, блокируя светлые либо темные зоны.

Принцип Ферма

Главной чертой *геометрической* оптики, которая исследует световые лучи в виде прямых линий, является то, что они движутся по путям, соответствующим кратчайшему времени пробега между источником и местом назначения. Это утверждение является *принципом наименьшего времени*, предложенным в 1657 году французским математиком Пьером Ферма.

Возникает вопрос, откуда свет, как кажется, заранее знает путь, на прохождение которого будет затрачено наименьшее время? Если он начал двигаться по неверному пути, не будет ли более экономным по времени продолжить движение, чем возвращаться к источнику и начинать сначала?

Волновая теория света приходит на помощь особо элегантным способом. Предположим, что мы рассматриваем произвольный путь между двумя заданными точками и представляем себе волну, извивающуюся по этому пути (рис. 6.8). На верхней части рисунка мы видим искривленный путь между двумя фиксированными точками и другой искривленный путь, близкий к нему. На этих путях нарисованы волны с одинаковой длиной. Хотя они начинают путь с одной и той же амплитудой, но когда достигают конечной точки, их амплитуды сильно различаются. В пункте назначения гребни и впадины волн, прибывших этими разными путями, уничтожают друг друга: эта взаимная аннигиляция называется *деструктивной интерференцией*. Однако существует один путь (на нижней части рис. 6.8), для которого различие между гребнями соседних волн столь мало, что они не уничтожают, а усиливают друг друга: это взаимное усиление называется *конструктивной интерференцией*. Пути, на которых интер-

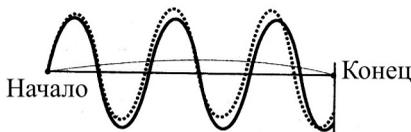
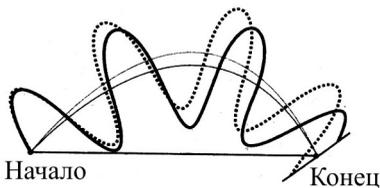


Рис. 6.8. Схемы движения волны из начальной в конечную точку

ференция конструктивна, очень близки к прямой линии, в общем случае это пути с наименьшим временем пробега между источником и пунктом назначения.

Свет не знает заранее, какой из путей окажется путем с наименьшим временем пробега: он испытывает все траектории, но только на путях, очень близких к пути с наименьшим временем пробега,

волны не гасят друг друга. Деструктивная и конструктивная интерференции становятся тем более точными, чем короче длина волны света, и только геометрическая прямая линия выживает при бесконечно малой длине волны, которая и является тем пределом, в котором физическая (волновая) оптика становится геометрической оптикой. Механизм интерференции дает ясное объяснение принципу наименьшего времени Ферма.

Поскольку в реальных волнах излучателями света являются атомы – немонахроматические источники, излучение которых не согласовано по фазам, результирующую интерференцию множества волн с различными длинами можно представить в виде *волнового пакета* (цуга). Этот цуг (рис. 6.9) является результатом сложения (*суперпозиции*) множества волн с различными длинами, каждая из которых соответствует определенному импульсу. Эти волны, складываясь там, где их гребни совпадают, образуют пик реальной волновой функции и гасят друг друга там, где их гребни совпадают со впадинами. Волновой пакет, образованный суперпозицией множества волн с различными длинами может быть достаточно четко зафиксирован в определенной области пространства, хотя мы ничего не можем сказать о том, какое из множества колебаний будет преобладать. Этот волновой пакет, который движется подобно классической частице, обладающей импульсом, называется *фотоном*.

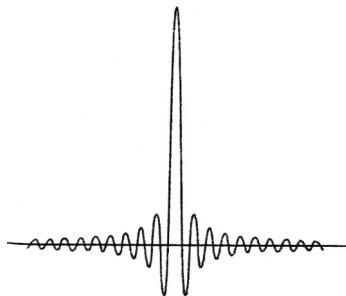


Рис. 6.9. Представление фотона в виде волнового пакета

Групповая скорость

Рассмотрим сложение двух волн, отличающихся только значениями волнового числа k и круговой частоты ω . Полное

электрическое поле, создаваемое этими волнами, распространяющимися в одном направлении, имеет вид

$$E = E \cos(\omega t - kz) + E \cos[(\omega + \Delta\omega)t - (k + \Delta k)z]. \quad (6.17)$$

В результате получаем волновой пакет, который модулируется (ограничивается по амплитуде) множителем $\cos[(\Delta\omega t - \Delta kz)/2]$, являющимся огибающей функцией движущейся волны. Эта огибающая функция тоже является волной и имеет скорость

$$v_g = \Delta\omega/\Delta k, \quad (6.18)$$

называемую *групповой скоростью волны*. В пределе малых приращений $\Delta\omega$ и Δk групповая скорость выражается производной

$$v_g = d\omega/dk. \quad (6.19)$$

Уравнение (6.19) справедливо и для более общего случая, когда результирующая волна составлена из множества волн с различными частотами, изменяющимися в узком диапазоне. В этом случае огибающая может оказаться импульсной функцией, а не косинусом. Тогда групповой скоростью будет скорость распространения этого импульса (см. рис. 6.9).

Отметим отличие групповой скорости от фазовой скорости (6.8). Если значение групповой скорости ограничено скоростью света, то фазовая скорость растет с увеличением частоты и может превосходить скорость света без нарушения принципа относительности. Такой реальный случай может произойти для фазовой скорости высокочастотной моды оптического волокна.

Групповой показатель преломления

Рассмотрим среду с *дисперсией*, в которой показатель преломления зависит от волнового числа: $n = n(k)$. Поскольку фазовая скорость $v = c/n$, а $v = \omega/k$, то

$$\omega = ck/n(k). \quad (6.20)$$

Групповую скорость в соответствии с формулой (6.19) найдем дифференцированием функции (6.20):

$$v_r = \frac{d\omega}{dk} = \frac{c}{n} \left(1 - \frac{k}{n} \frac{dn}{dk} \right) = v \left(1 - \frac{k}{n} \frac{dn}{dk} \right). \quad (6.21)$$

В среде с дисперсией $dn/dk \neq 0$ и фазовая скорость $v = c/n$ не равна групповой скорости.

Групповой показатель преломления n_r можно определить из соотношения $n_r = c/v_r$ или с учетом (6.21) из соотношения

$$n_r = n / \left(1 - \frac{k}{n} \frac{dn}{dk} \right). \quad (6.22)$$

В большинстве случаев групповой показатель преломления больше обычного, его значения зависят от частот волнового пакета.

Когерентность

Когерентностью называется согласованное протекание во времени и пространстве нескольких колебательных или волновых процессов.

Условию когерентности удовлетворяют *монохроматические волны* – неограниченные в пространстве волны одной определенной и строго постоянной частоты ($v = \text{const}$).

Реальные световые волны не являются строго монохроматическими, имеют разброс круговой частоты $\Delta\omega$. В силу фундаментальных физических причин излучение всегда имеет статистический характер. Каждый атом излучает относительно короткие вспышки или *волновые пакеты* длительностью $t \leq 10^{-8}$ с. Если атом возбудить несколько раз, то он может излучить несколько последовательных волновых пакетов. Временные интервалы между их появлениями могут быть велики по сравне-

нию с их длительностью. Фазы различных волновых пакетов не связаны друг с другом и не дают интерференционной картины.

Время когерентности t_k , при котором наблюдается интерференция, зависит от разброса круговой частоты.

$$t_k = 1/\Delta\omega. \quad (6.23)$$

Условие, при котором время когерентности не может превышать время излучения ($t_k < t$), является условием *временной когерентности*.

При переносе волны в пространстве фаза колебаний сохраняется только за время когерентности, за это время волна распространяется в вакууме на расстояние $l_k = ct_k$, называемое *длиной когерентности* (длиной волнового пакета) – расстоянием, при прохождении которого две или несколько волн утрачивают когерентность.

Условие, при котором оптическая разность хода не может превышать длину когерентности, называется условием *пространственной когерентности*.

$$\Delta < l_k. \quad (6.24)$$

Для источника света конечных размеров длина когерентности характеризуется максимальным поперечным направлением распространения волны – расстоянием, на котором возможно проявление интерференции.

$$l_k \approx \lambda/\varphi, \quad (6.25)$$

где λ – длина волны; φ – угловой размер источника.

Так, для излучения Солнца ($\lambda = 0,5$ мкм), которое с Земли характеризуется угловым размером $\varphi = 10^{-2}$ рад, $l_k \approx \lambda/\varphi = 0,5 \cdot 10^{-6} / 10^{-2} = 0,05$ мм. При такой малой длине когерентности интерференция солнечных лучей не наблюдается глазом, разрешающая способность которого $\sim 0,1$ мм.

6.3. Основные явления квантовой оптики

Атомы состоят из положительного ядра и электронной оболочки. Заряд атомного ядра (атомный номер) равен числу электронов, так что атом обычно является электрически нейтральным. Каждый электрон движется в электрическом (кулоновском) поле атомного ядра, которое частично экранируется другими электронами.

Молекулы состоят из нескольких атомных ядер и одной электронной оболочки, причем отдельные электроны могут быть отнесены к определенному ядру атома или равномерно распределяются по всей молекулярной зоне. Энергетические состояния молекул, как и в случае атомов, выражаются через электронные возбуждения, но дополнительно возникает еще энергия колебаний и вращений.

Во-первых, атомы в молекуле могут колебаться относительно своего равновесного положения, и, во-вторых, сама молекула может вращаться вокруг основных инерциальных осей. При этом общая энергия молекулы складывается из электронной энергии, колебательной энергии и вращательной энергии.

Многоатомные молекулы обнаруживают разные формы колебаний. Это можно объяснить на примере линейной, симметричной, трехатомной молекулы углекислого газа CO_2 (рис. 6.10). Для молекулы CO_2 возможны три основных формы колебаний. Каждое колебание является квантованным и способно существовать независимо от других форм колебаний. Относительно простая молекула CO_2 с небольшим числом атомов находит применение в газовых инфракрасных и ультрафиолетовых лазерах.

Рассмотрим теперь химическую связь молекул на основе системы двух атомов, например H и Cl. Эти атомы взаимно притягиваются, образуя молекулу хлороводорода HCl, причем между ними устанавливается определенное расстояние (межъядерный интервал) r_0 (рис. 6.11, *a*). При уменьшении этого расстояния атомы отталкиваются, при его увеличении – притягиваются.

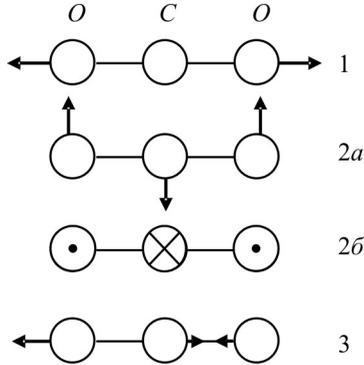


Рис. 6.10. Схемы колебаний молекулы CO_2 :
 1 – продольные симметричные; 2 – изгибные;
 3 – продольные асимметричные

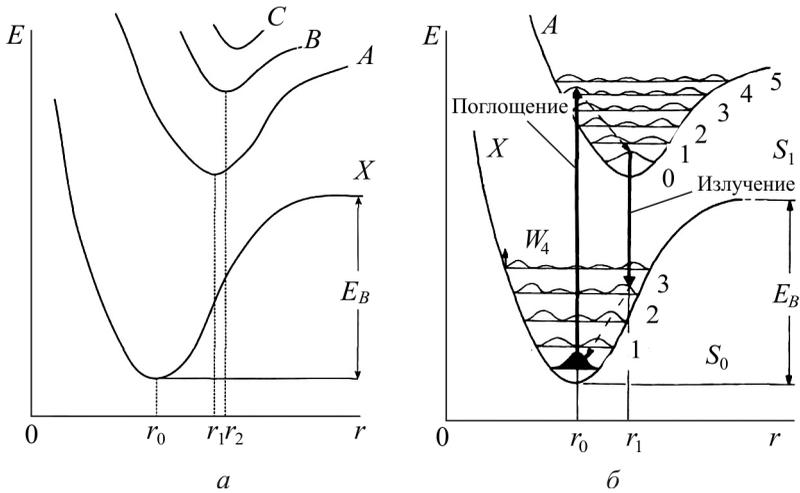


Рис. 6.11. Кривая потенциала X (а) дает схематическое представление о потенциальной энергии молекулы в основном состоянии в зависимости от межъядерного интервала. Энергия связи составляет E_B . Кривые потенциала электронно-возбужденных состояний обозначены как A , B , C и т.д.; колебательные уровни (б) имеют обозначения $0, 1, 2, \dots$, на них отображены пространственные распределения вероятности пребывания W . Переходы осуществляются преимущественно без изменения радиуса ядра между колебательными состояниями с максимальными значениями W

Кривая потенциала X на рис. 6.11, *a* отображает энергию взаимодействия (потенциальную энергию) двухатомной молекулы в зависимости от межъядерного интервала r . Эта характеристика демонстрирует свой минимум при r_0 . Для разрушения молекулярной связи необходимо соответствующим образом разделить атомы, а для этого требуется энергия связи E_B , показывающая глубину так называемой потенциальной ямы. Кривая X относится к основному состоянию электронов. Точно так же, как и в случае атомов, электроны могут возбуждаться с переходом на более высокие орбиты. Тогда при повышенных энергиях получаются другие кривые потенциала. Когда молекула переводится из основного состояния в возбужденные состояния, изменяется межъядерный интервал (r_1, r_2, r_3 и т.д.) и, соответственно, энергия связи.

В качестве условных обозначений используются: X для электронного основного состояния, A, B, \dots для возбужденных состояний. При переходах электронов с вышележащего уровня на нижележащий электронный уровень возникает излучение.

Наряду с электронной энергией молекулы в результате движения ядра появляются две энергетических составляющих. Во-первых, атомы в молекуле могут колебаться относительно своего равновесного положения и, во-вторых, сама молекула может вращаться вокруг основных инерциальных осей. При этом общая энергия молекулы складывается из электронной, колебательной и вращательной энергии.

Электронная энергия составляет от 1 до 20 эВ, колебательная энергия от 0,5 до 10^{-2} эВ, в то время как вращательная энергия будет менее 10^{-2} эВ. Схема уровней молекулы, таким образом, несколько сложнее схемы уровней атомов. Как видно из рис. 6.11, *b*, к каждому электронному уровню X, A, B, C, \dots относится несколько эквидистантных колебательных уровней, над которыми создаются затем вращательные уровни. Так же как электронная энергия, колебательная и вращательная энергия квантуется.

Молекулы могут возбуждаться в разные вращательно-колебательные уровни (см. рис. 6.11, б) с последующим переходом в нижележащие уровни, при этом происходит излучение. Здесь возможны многочисленные переходы между разными уровнями – электронными, вращательными или колебательными, а длина волны возникающего излучения находится в широком диапазоне спектра от ультрафиолетового до инфракрасного.

Вопросы для самоконтроля

1. Представление света в виде электромагнитной волны, понятие интенсивности света, физический смысл коэффициента преломления оптической среды.

2. Закон преломления света, явление полного внутреннего отражения.

3. Что характеризуют частота, длина волны света? В чем заключается явление поляризации света?

4. Свет как поток фотонов, плотность потока фотонов, энергия кванта света.

5. Приведите примеры длин волн, частот, энергий кванта для оптического диапазона спектра электромагнитного излучения.

6. В чем состоит явление интерференции, пространственно-временной когерентности? Объясните, почему интерференция света не наблюдается невооруженным глазом.

7. Принцип Ферма, явления конструктивной и деструктивной интерференции. Представление фотона в виде волнового пакета.

8. Объясните различие между групповой и фазовой скоростью света. От чего зависит групповой показатель преломления среды?

9. Перечислите формы колебаний возбужденной молекулы.

10. Объясните механизм квантовых явлений поглощения и излучения энергии сложной молекулы.

7. ЭЛЕМЕНТЫ НЕЛИНЕЙНОЙ ОПТИКИ

7.1. Механизмы оптической нелинейности

Характер взаимодействия интенсивного излучения с веществом становится нелинейным (по напряженности внешнего поля E) и многофотонным (по числу фотонов, поглощаемых в элементарном акте взаимодействия). Нелинейные явления в сильных электромагнитных полях возникают не в результате нарушения принципа суперпозиции для электромагнитного поля, а в результате влияния электромагнитного поля на поляризованность среды.

В современной оптике видное место занимают явления, связанные с нелинейностью отклика вещества на световое поле. Это такие явления, как *двухфотонное или многофотонное поглощение света, оптический пробой среды, вынужденное рассеяние света, самофокусировка световых пучков, самомодуляция импульсов, удвоение или утроение частоты света* и т.п. При всем многообразии нелинейно-оптических явлений можно выделить некоторые общие черты, присущие каждому из них.

Во-первых, это сильная зависимость от интенсивности света. Как правило, нелинейно-оптический эффект становится заметным лишь при достаточно большой интенсивности света. Неслучайно нелинейная оптика появилась лишь после создания лазера. Нелинейная оптика – это оптика сильных световых полей, оптика мощного лазерного излучения.

Во-вторых, в нелинейно-оптических процессах возникают новые спектральные компоненты поля, различные световые волны сильно взаимодействуют между собой, между ними происходит энергообмен, вплоть до полного преобразования одной волны в другую. Поэтому *для нелинейных эффектов невозможно применение принципа суперпозиции*, состоящего в том, что различные световые волны, отличающиеся частотой, направлением распространения, поляризацией, распространяются и взаимодействуют со средой независимо друг от друга.

Основными уравнениями, описывающими оптическое излучение в линейной либо нелинейной среде, являются уравнения Максвелла. В этих уравнениях вектор электрической индукции

$$D = E + 4\pi P, \quad (7.1)$$

где P – вектор поляризации среды, равный полному дипольному моменту единицы объема среды; напомним, что дипольный момент элементарного диполя равен произведению величины заряда электрического диполя (пары разноименных электрических зарядов) на расстояние между ними. В изотропной среде векторы D и E однонаправлены (коллинеарны),

$$D = \varepsilon E, \quad (7.2)$$

где ε – скалярная диэлектрическая проницаемость среды. В анизотропных средах направления векторов D и E могут различаться.

В оптике вводится характеристика восприимчивости (поляризуемости) α оптической среды, характеризующая отклик (поляризацию) среды на внешнее электрическое поле. В изотропных линейных средах вектор поляризации прямо пропорционален напряженности внешнего электрического поля,

$$P = \alpha E. \quad (7.3)$$

Известны три типа поляризуемости среды:

– *электронная* (рис. 7.1), обусловленная деформацией (вытягиванием) орбиты электрона в атоме под действием кулоновской силы; при этом характерное время отклика среды на изменение внешнего электрического поля около 10^{-15} – 10^{-14} с;

– *ионная*, обусловленная смещением тяжелых ионов в молекуле; время отклика 10^{-13} – 10^{-11} с;

– *ориентационная* (рис. 7.2), вызванная изменением ориентации (поворотом) диполей молекул во внешнем электрическом поле; время отклика среды порядка 10^{-10} с.

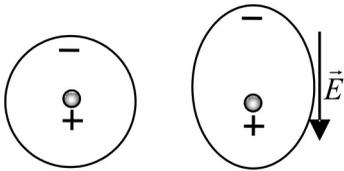


Рис. 7.1. Схема электронной поляризации молекулы

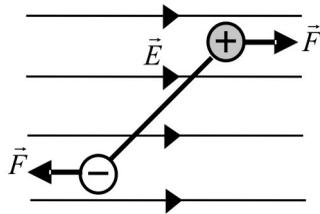


Рис. 7.2. Схема ориентационной поляризации молекулы

В линейной оптике при малых значениях мощности оптического излучения в среде скалярные коэффициенты диэлектрической проницаемости и поляризуемости среды постоянны и связаны соотношением

$$\varepsilon = 1 + 4\pi\alpha, \quad (7.4)$$

а в анизотропном диэлектрике направления векторов D и E могут различаться:

$$D_i = E_i + 4\pi \sum_{k=1}^3 \alpha_{ik} E_k, \quad (7.5)$$

при этом компоненты тензоров диэлектрической проницаемости и поляризуемости связаны соотношением

$$\varepsilon_{ik} = \delta_{ik} + 4\pi\alpha_{ik}, \quad (7.6)$$

где символ Кронекера $\delta_{ik} = 0$ при $i \neq k$ и 1 при $i = k$.

При больших мощностях излучения эти коэффициенты уже зависят от напряженности E электрического поля электромагнитной волны, при этом в общем случае анизотропной среды

$$P_i = \sum_{k=1}^3 \alpha_{ik}(E) E_k. \quad (7.7)$$

Функцию $\alpha_{ik}(E)$ удобно представить в виде ряда по степеням E :

$$\alpha_{ik}(E) = \alpha_{ik} + \sum_{j=1}^3 \chi_{ikj} E_j + \sum_{j=1}^3 \sum_{m=1}^3 \theta_{ijkm} E_j E_m + \dots \quad (7.8)$$

Первое слагаемое этого ряда (тензор 2-го ранга) описывает линейную восприимчивость среды, вторая сумма – квадратичную восприимчивость (тензор 3-го ранга) и третья, двойная сумма, – кубичную восприимчивость (тензор 4-го ранга). В типичных оптических средах, например в кварцевых стеклах и нелинейных кристаллах, порядок величин тензоров следующий:

$$\alpha_{ik} \approx 1, \chi_{ikj} \approx 10^{-12} - 10^{-11} \text{ м/В}, \theta_{ijkm} \approx 10^{-22} - 10^{-21} (\text{м/В})^2.$$

Выражение (7.7) для произвольной компоненты вектора электрической поляризации принимает вид

$$P_i = \sum_{k=1}^3 \alpha_{ik} E_k + \sum_{k=1}^3 \sum_{j=1}^3 \chi_{ikj} E_k E_j + \sum_{k=1}^3 \sum_{j=1}^3 \sum_{m=1}^3 \theta_{ijkm} E_k E_j E_m + \dots \quad (7.9)$$

Три группы слагаемых в приведенном выражении описывают, соответственно, *линейную, квадратичную и кубичную поляризации среды*. Нелинейные оптические среды подразделяются на квадратично-нелинейные и кубично-нелинейные. В кубично-нелинейных средах – в изотропных газах и жидкостях, а также в кристаллах, обладающих центром симметрии, тензор нелинейной восприимчивости χ_{ikj} третьего ранга равен нулю. При этом нелинейная поляризуемость среды пропорциональна третьей степени напряженности электрического поля электромагнитной волны в среде (рис. 7.3, а).

Для таких сред характерен квадратичный, по влиянию электрического поля на показатель преломления среды, электрооптический эффект Керра, объясняющий возникновение двойного лучепреломления под действием внешнего электрического поля. В квадратично-нелинейных средах компоненты тен-

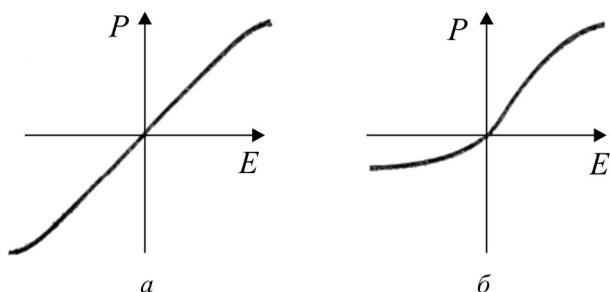


Рис. 7.3. Зависимость величины поляризации P от напряженности E электрического поля: a – в кубично-нелинейной среде; b – в квадратично-нелинейной среде

зора $\chi_{ikj} \neq 0$, и наибольший вклад в оптическую нелинейность вносит именно квадратичная поляризуемость (рис. 7.3, б). В таких кристаллах имеет место *пьезоэлектрический эффект* (связывающий механическую деформацию кристалла и его поляризацию). В квадратично-нелинейных средах наблюдается линейный, по влиянию электрического поля на показатель преломления среды, *электрооптический эффект Поккельса*, следствием которого является двойное лучепреломление в кристаллах, при котором входящий в кристалл луч света на выходе из кристалла раздваивается на обыкновенный и необыкновенный лучи.

7.2. Вынужденное рассеяние света

Случайные изменения плотности среды, обусловленные тепловыми движениями молекул (тепловые акустические волны), рассеивают световую волну и модулируют ее по частоте, при этом возникают сателлиты с частотами, равными сумме и разности частот световой волны и тепловых акустических колебаний.

При увеличении интенсивности падающего излучения выше порогового значения происходит следующее. Под действием электрического тока из-за явления электрострикции возникают импульсы избыточного давления, достигающие в поле ла-

зерного луча десятков тысяч атмосфер. Возникает акустическая волна давления (гиперзвук, 10^{10} Гц), изменяющая показатель преломления по закону бегущей волны. Эти изменения показателя преломления образуют в среде периодическую решетку, на которой и происходит рассеяние световой волны. При этом интенсивность сателлитов становится сравнимой с интенсивностью падающей волны, а количество их возрастает. Описанный эффект называется вынужденным рассеянием *Мандельштама–Бриллюэна*. Спонтанное рассеяние света на тепловых акустических волнах было рассмотрено Бриллюэном еще в 1922 году. (Одновременно с Бриллюэном и независимо от него рассеяние света в твердых телах теоретически исследовал Л.И. Мандельштам.) Вынужденное рассеяние, когда акустическая волна, рассеивающая свет, сама возбуждается этим светом, было открыто в 1964 г.

При достаточно больших интенсивностях падающего излучения нелинейная среда может стать генератором звука со световой накачкой. С помощью лазеров удается возбуждать мощные (до 10 кВт) гиперзвуковые колебания во многих жидкостях и твердых телах.

При рассеянии интенсивного лазерного излучения в жидкостях и кристаллах, помимо описанных выше боковых спектральных компонент, обнаруживаются компоненты с частотами, в точности кратными частоте падающего излучения (двукратными, трехкратными и т.д.), называемые *оптическими гармониками*. В некоторых кристаллах эти гармоники могут составлять до 50 % рассеянного излучения. Например, если направить красное излучение рубинового лазера (0,69 мкм) на кристалл дигидрофосфата калия, то на выходе можно получить невидимое ультрафиолетовое излучение (0,345 мкм).

7.3. Самофокусировка

Известно, что первоначально параллельный пучок света по мере распространения в среде расплывается за счет дифракционных явлений. Это справедливо при малых интенсивностях света,

пока еще среда остается линейной. С увеличением мощности светового пучка его расходимость начинает уменьшаться. При некоторой критической мощности пучок может распространяться, вообще не испытывая расходимости (режим *самоканализации*). При мощности, превышающей критическую мощность, пучок излучения скачком сжимается к оси и сходится в точку на некотором расстоянии от места входа. Происходит процесс *самофокусировки*.

Физические причины этого эффекта заключаются в изменении показателя преломления среды в сильном световом поле. Среда в зоне пучка становится оптически неоднородной; показатель преломления среды определяется теперь распределением интенсивности световой волны. Это приводит к явлению *нелинейной рефракции*, т.е. периферийные лучи пучка отклоняются к его оси в зону с большей оптической плотностью. Таким образом, нелинейная рефракция начинает конкурировать с дифракционной расходимостью. При взаимной компенсации этих процессов и наступает самоканализация, переходящая в самофокусировку при превышении критической мощности пучка. Процесс самофокусировки выделяется среди прочих нелинейных эффектов тем, что он обладает «лавиным» характером. Действительно, даже малое увеличение интенсивности в некотором участке светового пучка приводит к концентрации лучей в этой области, а следовательно, и к дополнительному возрастанию интенсивности.

Отметим, что критические мощности самофокусировки относительно невелики (для нитробензола – 25 кВт, для некоторых сортов оптического стекла – 1 Вт), что создает реальные предпосылки использования описанного эффекта для передачи энергии на значительные расстояния.

Самофокусировка – это явление сжатия апертурно-ограниченного пучка света в кубичной нелинейной среде, которое сопровождается увеличением плотности мощности излучения в поперечном сечении распространяющегося излучения и образованием нитевидных волноводных каналов в такой среде.

Для кубичной ($\chi = 0, \theta \neq 0$) нелинейной среды характерна следующая зависимость мгновенного показателя преломления $n(t)$ от напряженности $E(t)$ электрического поля распространяющейся волны:

$$n(t) = n_0 + n_2 |E(t)|^2. \quad (7.10)$$

Распределение мощности излучения в поперечном сечении светового пучка приблизительно соответствует функции с максимумом ее значения в центре пучка. При превышении порогового значения мощности показатель преломления нелинейной среды в центре пучка возрастает, а к периферии пучка плавно уменьшается.

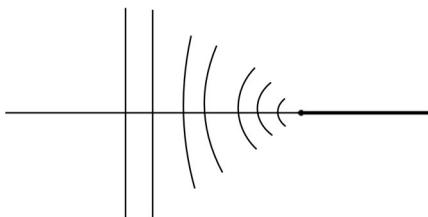


Рис. 7.4. Преобразование световой волны в сходящуюся при самофокусировке

В результате среда пространства становится подобной положительной градиентной линзе и преобразует изначально плоский волновой фронт световой волны в сходящийся (рис. 7.4).

Явление самофокусировки вследствие чрезвычайно высокой плотности мощности излучения в канале распространения нередко сопровождается необратимыми явлениями – оптическим пробоем среды либо интенсивным поглощением или рассеянием света на дефектах структуры и инородных примесях.

7.4. Нелинейные эффекты в волоконных световодах

Возникновение нелинейных эффектов в оптических волоконных световодах связано с индуцированной поляризацией P электрических диполей, которая уже не является линейной, а удовлетворяет более общему соотношению

$$P = \varepsilon_0 (\chi^1 E + \chi^2 EE + \chi^2 EEE + \dots), \quad (7.11)$$

где ε_0 – диэлектрическая проницаемость вакуума, χ^j ($j = 1, 2, \dots$) – восприимчивость j -го порядка. Главный вклад в P вносит линейная восприимчивость χ^1 , она определяет показатель преломления n . С восприимчивостью второго порядка χ^2 связаны такие эффекты, как генерация второй гармоники и генерация суммарной частоты. Однако эта восприимчивость ненулевая только для сред, в которых на молекулярном уровне отсутствует симметрия инверсии. Поскольку в кварцевых стеклах молекула SiO_2 обладает центром симметрии, $\chi^2 = 0$. Поэтому в оптических световодах не могут иметь место эффекты второго порядка. Тем не менее, слабые нелинейные эффекты второго порядка могут возникать из-за примесей внутри сердцевины.

Если на ранней стадии развития волоконно-оптических линий связи (ВОЛС) единственными проблемами являлись погонные оптические потери и волоконно-оптическая дисперсия, то сейчас на первое место стали выходить проблемы, связанные с нелинейными эффектами, особенно остро проявляющимися при передаче высокоскоростной цифровой информации.

В современной технике ВОЛС известно несколько механизмов нелинейности, возникающей в оптоволокне. Наибольший вклад в искажение передаваемой информации вносит изменение показателя преломления материала световода в зависимости от оптической мощности пропускаемого сигнала (рис. 7.5).

Зависимость показателя преломления от интенсивности приводит к множеству нелинейных эффектов. Два наиболее широко изученных нелинейных эффекта: *фазовая самомодуляция* и *фазовая кросс-модуляция*.

Фазовая самомодуляция обусловлена самонаведенным набегом фазы, который оптическое поле приобретает при распространении в волоконном световоде.

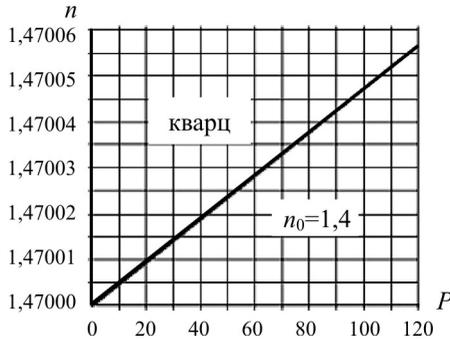


Рис. 7.5. Зависимость показателя преломления кварца от оптической мощности излучения

Фазовая кросс-модуляция обусловлена нелинейным набегом фазы оптического поля, который наведен другим полем на другой длине волны, распространяющимся совместно.

Другой класс нелинейных эффектов вызван вынужденным неупругим рассеянием, при котором оптическое поле передает часть своей энергии нелинейной среде. В эту категорию попадают два важных нелинейных эффекта; оба они связаны с возникновением колебательных мод кварца. Это эффекты *вынужденного комбинационного рассеяния* (ВКР) и *вынужденного рассеяния Мандельштама–Бриллюэна* (ВРМБ).

Явление ВКР света было открыто и объяснено в 1962–1963 годах. Оно послужило толчком к изучению вынужденного рассеяния других видов. При пропускании через кристалл (сапфира или кварца) мощного лазерного излучения в кристалле возбуждалась акустическая волна и одновременно генерировалось оптическое излучение. И акустический, и рассеянный оптический лучи испускались в строго определенных направлениях и возникали только при условии, когда мощность лазера превышала определенное пороговое значение.

Таким образом, оптический сигнал рассеивается и смещается в область более длинных волн. Если при ВРМБ спектр стимулированного излучения узкий (десятки МГц) и смещен в длин-

новолновую сторону на ~ 10 ТГц, то при ВКР спектр стимулированного излучения широкий (~ 7 ТГц) смещен в длинноволновую сторону на величину порядка ~ 10 ТГц. При схожести ВРМБ и ВКР основное различие между этими эффектами состоит в том, что в ВКР принимают участие оптические фононы (кванты колебаний атомов кристаллической решетки), тогда как в ВРМБ – акустические.

В ВОЛС при достижении входной мощности, равной порогу ВРМБ, может начаться интенсивное рассеяние света в обратном направлении, приводящее к деградации качества связи. Поэтому уровень передаваемой мощности всегда должен быть меньше этого порога.

Из-за нелинейных оптических эффектов возникают фундаментальные ограничения по скорости передачи информации. Практическим пределом следует считать скорость в 10 Тбит/с. Для увеличения скоростей информационных потоков более рациональным является увеличение числа оптических жил в волоконном кабеле.

7.5. Оптические солитоны

Для большинства оптических материалов в видимом диапазоне показатель преломления n растет с частотой. Это явление называется *нормальной дисперсией* (от лат. *dispergo* – рассеиваю) показателя преломления. Вблизи полос поглощения света наблюдается уменьшение n с частотой – *аномальная дисперсия*. Явление дисперсии в видимом диапазоне объясняется реакцией внешних валентных электронов. На частоте поглощения света они начинают осциллировать в резонансном режиме, как показано на рис. 7.6, и по мере дальнейшего увеличения частоты перестают давать отклик. После достижения резонанса n снова начинает возрастать и одновременно возникают потери интенсивности света, вызванные его поглощением.

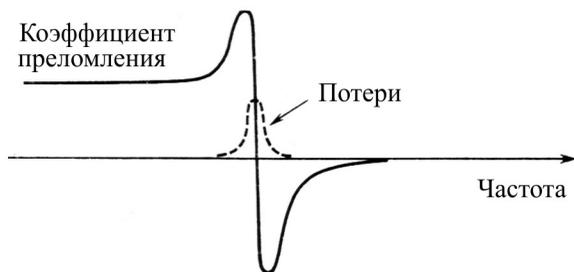


Рис. 7.6. Явление резонанса

В области аномальной дисперсии особенно ярко проявляются нелинейные свойства оптических световодов. Здесь могут существовать *солитоны*, образования, обусловленные совместным действием дисперсионных и нелинейных эффектов. Термин «солитон» (от англ. *solitary wave* – уединённая волна) относится к специальному типу волновых пакетов, которые могут распространяться на значительные расстояния без искажения своей формы и сохраняются при столкновениях друг с другом; во многом они ведут себя подобно частице.

Теория солитонов тесно связана с явлением *модуляционной неустойчивости*, которая в волоконной оптике наблюдается в области аномальной дисперсии (вблизи полосы поглощения вещества, где наблюдается локальное уменьшение показателя преломления с частотой). Совместное действие дисперсии групповых скоростей и нелинейных эффектов в световоде в области отрицательных дисперсий является одной из основных причин, которая объясняет возникновение оптических солитонов.

Для волокна с *положительной дисперсией* ее влияние сводится к расширению спектра и расплыванию импульса со временем. *Отрицательная же дисперсия* среды оказывает нестандартное влияние: световой импульс сначала несколько расширяется, затем стабилизируется, а сам спектр импульса сужается. Особый интерес к солитонам обусловлен тем, что это единственный случай стационарной волны, основная энергия которой заключена

в конечной области пространства. При взаимодействии друг с другом они не разрушаются и не рассеиваются.

Оптические солитоны нашли применение по крайней мере в трех областях:

- при создании солитонных лазеров;
- при создании солитонных линий связи.
- в оптических устройствах цифровой обработки информации.

Основная идея *солитонных лазеров* – использование волоконного световода для осуществления синхронной подачи части энергии обратно в резонатор лазера, работающего в режиме синхронизации мод. Световод изменяет форму импульса, формируя солитон. После нескольких циклов формируется стационарное состояние, в котором импульсы являются солитонами световода. Длительностью импульса можно управлять, изменяя длину световода. При этом длительность может быть гораздо меньше, чем в случае одного лазера без световода (~50 фс). В последнее время с использованием многосолитонных импульсов в световоде получены очень короткие импульсы, до 6 фс.

Самая простая схема *солитонной линии связи* предложена в 1983 году Хасегавой (рис. 7.7). В ней реализована архитектура линейной последовательной цепи, состоящей из линейных сегментов световодов длиной L с усилителями. Усиление организовано так, что на концах каждого сегмента установлены направленные ответвители, через которые в линию связи (световод) в обоих направлениях вводится непрерывное излучение накачки от лазера, работающего на длине волны 1460 нм. В схеме может быть использовано одномодовое волокно – как обычное, так и со сдвигом дисперсии, с эффективной площадью сердцевины 25 мкм^2 , работающее на длине волны 1550 нм. Реализованные значения L составили 40–50 км, а общая длина линии – 600 км. С помощью таких усилителей была достигнута скорость 160 Гбит/с. Преимущество солитонных систем перед обычными методами оптической передачи – возможность повы-

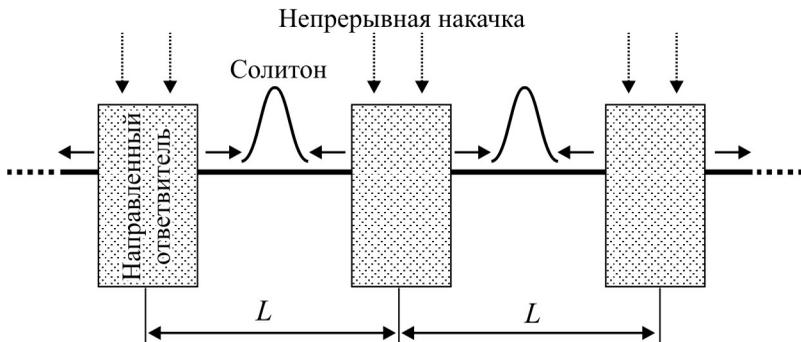


Рис. 7.7. Схема солитонной линии связи

шения скорости передачи и увеличения длины регенерационного участка.

Последние исследования показывают возможность эффективного использования солитонов в оптических устройствах *обработки цифровой информации*. Волоконные интерферометры позволяют переключить солитонный сигнал. С помощью таких переключателей реализуются логические функции «И», «ИЛИ», «НЕ».

Нелинейные эффекты широко используются в самых различных областях современной оптической техники: в высокоточных оптических датчиках, биосенсорах, кремниевой фотонике, интерферометрии, ВОЛС и др. Это направление науки и техники быстро развивается и предъявляет серьезные требования к уровню подготовки специалистов, работающих в данных областях. В связи с разработкой мощных оптических излучателей и высококачественных оптических сред области практического применения нелинейной оптики непрерывно расширяются, при этом величина пороговой мощности, при которой наблюдаются эффекты нелинейности, имеет тенденцию к снижению. Нелинейная оптика, в том числе нелинейная волоконная оптика, лежит в основе действия как современных, так и перспективных устройств оптических систем передачи и обработки информации.

Вопросы для самоконтроля

1. Что такое нелинейная оптика? Перечислите оптические эффекты, вызванные оптической нелинейностью.
2. Перечислите типы поляризуемости среды. Какими механизмами они вызваны?
3. Электрооптический эффект Керра. Каким видом поляризации он вызван?
4. Физическая сущность пьезоэлектрического эффекта и электрооптического эффекта Погкельса.
5. Физическое содержание эффекта вынужденного рассеяния Мандельштама–Бриллюэна.
6. Объясните эффект самофокусировки света.
7. Объясните нелинейные эффекты фазовой самомодуляции, фазовой кросс-модуляции, вынужденного комбинационного рассеяния и вынужденного рассеяния Мандельштама Бриллюэна в волоконных световодах.
8. Объясните принцип возникновения оптических потерь при аномальной дисперсии в оптическом волокне.
9. Что такое солитон, в чем преимущества солитонных линий передачи информации?

ЧАСТЬ 3. ФИЗИЧЕСКИЕ И НАНОТЕХНОЛОГИЧЕСКИЕ ОСНОВЫ ФОТОНИКИ

8. ПОЛУПРОВОДНИКОВЫЕ КВАНТОВЫЕ СТРУКТУРЫ

8.1. Роль полупроводниковых структур в оптоэлектронике

В основе действия элементов оптоэлектроники лежат универсальные схемные свойства полупроводниковых структур. Значимость универсальных схемных свойств полупроводниковых структур для массовой технологии интегральных микросхем была осознана специалистами далеко не сразу. За изобретение транзистора Нобелевская премия была присуждена У. Шокли в 1956 году, т.е. через 7 лет после разработки, а за открытие, исследование и использование в микро- и оптоэлектронике основополагающих свойств полупроводниковых структур – через десятки лет (Ж.И. Алферов, Дж. Килби, Г. Кремер, Нобелевская премия 2000 г.).

Изобретения универсальных конструкционных элементов появляются довольно редко, но всегда оказываются весьма ценными, так как определяют длительный технический прогресс в соответствующей области. В механических устройствах универсальными конструкционными элементами оказались, например, стержень и колесо, до сих пор позволяющие создавать все более и более совершенные механические системы и изделия. Природа также преуспела в создании необозримого разнообразия живых систем, «используя» универсальный конструкционный элемент – клетку. Нет никаких сомнений в том, что переход к нанооптоэлектронике не ограничится разработкой лишь лабораторных образцов, если будут найдены универсальные конструкционные элементы, комбинацией которых удастся создавать наноприборы различного назначения, используя либо группо-

вую технологию, либо процессы самосборки. Для нанотехнологии по принципу «снизу вверх» такими универсальными конструктивными элементами являются атомы.

Полупроводниковая структура представляет собой некую границу раздела, в которой присутствует полупроводниковый материал. Сюда относятся граница раздела между областями с дырочной и электронной проводимостью внутри полупроводникового кристалла (p – n -переход), граница раздела между слоями полупроводника с различной шириной запрещенной зоны (гетеропереход) и др.

Электронно-дырочные переходы – основа элементной базы интегральных микросхем, а гетеропереходы – основа оптоэлектронных приборов различного назначения. Многие наноструктуры созданы при разработке все более миниатюрных структур рассмотренного выше типа. С уменьшением размера структур были выявлены новые полезные для оптоэлектроники физические (в основном квантовомеханические) эффекты. Обычно их называют размерными.

8.2. Твердотельные гетероструктуры. Полупроводниковый гетеропереход

Интерес к твердотельным гетероструктурам обусловлен тем, что они являются основой практически всех современных приборов и устройств электроники, фотоники и оптоинформатики. Гетероструктуры широко применяются в повседневной жизни, достаточно вспомнить бытовую технику, аудио- и видеосистемы, мобильные телефоны и компьютеры, а также в большинстве областей человеческой деятельности, начиная от производства и кончая медициной и образованием. Уровнем проникновения гетероструктур в жизнь человека определяется ее качество. В связи с этим первостепенное значение приобретают технологии создания гетероструктур различного типа, их развитие и совершенствование.

В переводе с греческого *heteros* – другой или иной. В русском языке *гетеро* соответствует слову *разный*.

Гетеропереход представляет собой контакт между двумя различными веществами либо между веществом и вакуумом. Гетероструктура – объект, обладающий по крайней мере одним гетеропереходом.

В физике полупроводников термин *гетероструктура* обозначает выращенный на подложке слоистый пирог из различных полупроводников, в общем случае отличающихся шириной запрещённой зоны. Это определение сужает область применимости термина *гетероструктура*, так как относится только к слоистым полупроводниковым материалам.

Полупроводниковым гетеропереходом называют контакт двух полупроводников различного вида и разного типа проводимости, например *p-Ge – n-GaAs*. Отличие гетеропереходов от обычного *p-n*-перехода заключается в том, что в обычных *p-n*-переходах используется один и тот же вид полупроводника, например *p-Si – n-Si*. Поскольку в гетеропереходах используются разные материалы, необходимо, чтобы у этих материалов с высокой точностью совпадали два параметра: температурный коэффициент расширения и постоянная решетки.

Электронно-дырочным переходом или *p-n-переходом* называется граница раздела между полупроводниками с электронной и дырочной проводимостью.

Положения энергетических уровней Ферми, которые в полупроводниках заняты электронами при $T = 0$, в электронном (E_F^n) и дырочном (E_F^p) кристаллах не одинаковы ($E_F^n \neq E_F^p$). Термодинамическое равновесие между приведенными в контакт кристаллами наступает в результате выравнивания энергий Ферми ($E_F^n = E_F^p$). Это выравнивание сопровождается перетеканием «электронной жидкости» из кристалла, где уровень Ферми выше, в кристалл, где он ниже. В результате указанных процессов контакт кристаллов *p*- и *n*-типа приводит в области образо-

вавшегося p - n -перехода к взаимному смещению энергетических зон по оси E , как это показано на рис. 8.1. На рисунке виден энергетический скачок (потенциальный барьер), проявляющийся в изгибе всех энергетических уровней в валентной зоне и зоне проводимости.



Рис. 8.1. Энергетическая схема кристалла в области p - n -перехода (без учета влияния внешнего потенциала, приложенного к p - n -переходу)

Наличие и особенности потенциального барьера в области p - n -перехода определяют свойства приборов, выполненных на основе p - n -перехода.

При протекании прямого тока через p - n -переход основные носители тока (например, электроны из n -области) «принудительно» проникают (инжектируются) в область с противоположным типом проводимости (в p -область). В этой области указанные носители являются неосновными и неравновесными.

Введение неравновесных избыточных носителей заряда в полупроводниковый или диэлектрический кристалл под действием электрического поля называется инжекцией. Инжекция

характерна для контактов металл–полупроводник и для p – n -переходов. Будучи неравновесными и неосновными в области с противоположным типом проводимости, инжектированные носители претерпевают *рекомбинацию* (присоединение) типа «зона-зона» с основными носителями. При этом концентрация инжектированных носителей убывает и через некоторое время τ составляет лишь $1/e \approx 1/2,7$ часть от исходной величины. Это время называется *временем жизни неравновесных носителей заряда*. За время τ инжектированные носители успевают углубиться в область с противоположным типом проводимости на некоторое расстояние L от p – n -перехода. Это расстояние называется *диффузионной длиной неравновесных носителей* (электронов L_n и дырок L_p). Диффузионная длина L связана со временем жизни τ соотношением

$$L_n \approx \sqrt{D_n \tau_n}; \quad L_p \approx \sqrt{D_p \tau_p}, \quad (8.1)$$

где D_n и D_p – постоянные, называемые коэффициентами диффузии электронов и дырок соответственно.

Величины τ и L определяют конструкцию и характеристики многих полупроводниковых приборов. Они зависят от типа полупроводника, концентрации и вида примесей и дефектов в нем.

Отметим также, что рекомбинация каждого инжектированного через p – n -переход неравновесного носителя заряда может сопровождаться излучением кванта света $h\nu$, что символизирует зигзагообразная стрелка на рис. 8.1. Этот эффект лежит в основе действия полупроводниковых светодиодов и квантовых генераторов. При малых токах через p – n -переход рекомбинационное излучение является спонтанным, p – n -переход работает как светодиод. С увеличением тока, начиная с его порогового значения ($i = i_{\text{пор}}$), спонтанное излучение p – n -перехода преобразуется в лазерное, и p – n -переход начинает работать как квантовый генератор (*инжекционный квантовый генератор*).

Отметим особенности и свойства p - n -перехода. При контакте донорного полупроводника с акцепторным в окрестности p - n -перехода (x_{p-n}) концентрации доноров и акцепторов изменяются. Электроны из n -полупроводника диффундируют в p -полупроводник. В результате у границы контакта нескомпенсированные заряды образуют *двойной электрический слой*, поле которого препятствует дальнейшему перетеканию зарядов.

С выравниванием уровней Ферми энергетические зоны искривляются, возникают потенциальные барьеры для электронов и дырок. Стационарная высота барьера ($e\phi_k$) устанавливается при наступлении динамического равновесия между описанными встречными потоками электронов и дырок в области p - n -перехода. Потенциальный барьер в области p - n -перехода несимметричен. Его высота убывает, если на p -область подан положительный потенциал, и возрастает при обратной полярности включения p - n -перехода. Этот факт лежит в основе эффекта выпрямления p - n -переходом переменного электрического тока (рис. 8.2).

Пунктирная прямая линия на рис. 8.2 соответствует вольт-амперной характеристике линейного сопротивления, подчиняющегося закону

Ома. Если внешняя разность потенциалов U на p - n -переходе увеличивается, то потенциальный барьер $e\phi_k$ (e – заряд электрона, ϕ_k – разность потенциалов на границе контакта) становится выше и равняется при обратном смещении $e(\phi_k + U)$, при прямом смещении потенциальный барьер $e(\phi_k - U)$ становится ниже. Поэтому ток через p - n -переход увеличивается с ростом напря-

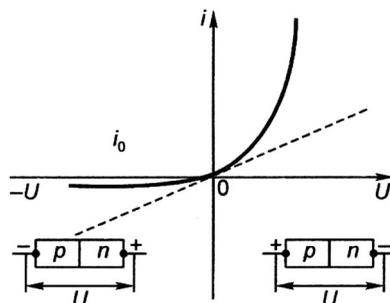


Рис. 8.2. Вольт-амперная характеристика p - n -перехода: пунктир – линейная (омическая) характеристика; вставки поясняют включение p - n -перехода

жения менее резко, чем по закону Ома, при обратном смещении и более резко – при прямом. В результате вольт-амперная характеристика становится нелинейной (сплошная линия на рис. 8.2). Прямое направление внешнего поля, *уменьшающее запирающий слой, при котором электроны и дырки рекомбинируют, называется пропускным*. Обратное направление внешнего поля, *расширяющее запирающий слой, называется запирающим*. В этом направлении ток не проходит.

Таким образом, *p–n*-переход обладает односторонней (вентильной) проводимостью, дифференциальным сопротивлением и приобретает свойство выпрямлять переменный электрический сигнал.

8.3. Квантоворазмерные структуры, их самоорганизация

Важнейшим свойством наноструктур является зависимость их свойств от характерного размера неоднородностей. Наиболее широко известное проявление этого свойства называется *эффектом квантового ограничения (quantum confinement)*.

При плавном уменьшении размеров образца от больших (макроскопических) значений, например метра или сантиметра, до очень маленьких свойства сначала остаются неизменными, затем начинают медленно меняться, а при размерах менее 100 нм могут измениться радикально. Если размеры образца в одном измерении лежат в нанометровом диапазоне, а в двух других остаются большими, то получившаяся структура называется *квантовой ямой*. Если образец мал в двух измерениях и имеет большие размеры в третьем, то такой объект называют *квантовой проволокой (шнуром)*. Предельный случай этого процесса уменьшения размеров, при котором размеры во всех трех измерениях лежат в нижней части нанометрового диапазона, называется *квантовой точкой*. Эпитет «квантовый» в названиях этих трех типов наноструктур используют потому, что в области ультра-

малых масштабов возникает изменение свойств квантовомеханической природы. Таким образом, среди низкоразмерных структур можно выделить три элементарные структуры. Это квантовые ямы, квантовые нити и квантовые точки (рис. 8.3).

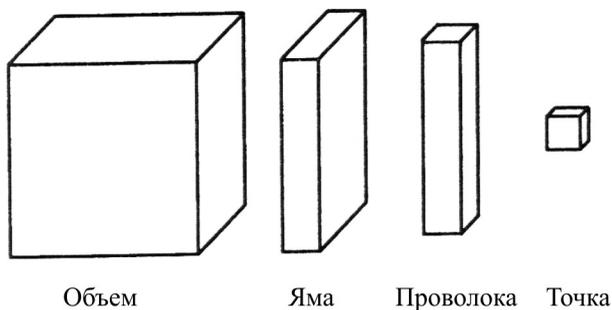


Рис. 8.3. Последовательность прямоугольных наноструктур

Эти элементарные структуры представляют собой кристаллический материал, пространственно ограниченный в одном, двух и трех направлениях. Для изготовления наноструктур используют всевозможные полупроводниковые соединения, а также полупроводники четвертой группы Si и Ge. В качестве примера на рис. 8.4 представлены изображения реальных элементарных наноструктур, полученные с помощью электронного микроскопа.

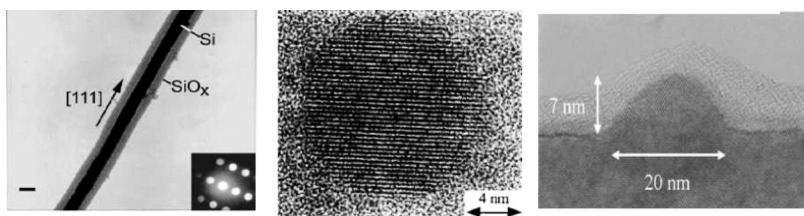


Рис. 8.4. Изображения (слева направо) квантовой нити, квантовой точки CdS в SiO₂, квантовой точки InAs в GaAs, полученные с помощью просвечивающего электронного микроскопа

Квантовое ограничение приводит, как видно из рис. 8.5, к ненулевому минимальному значению энергии и дискретности энергий разрешенных состояний.

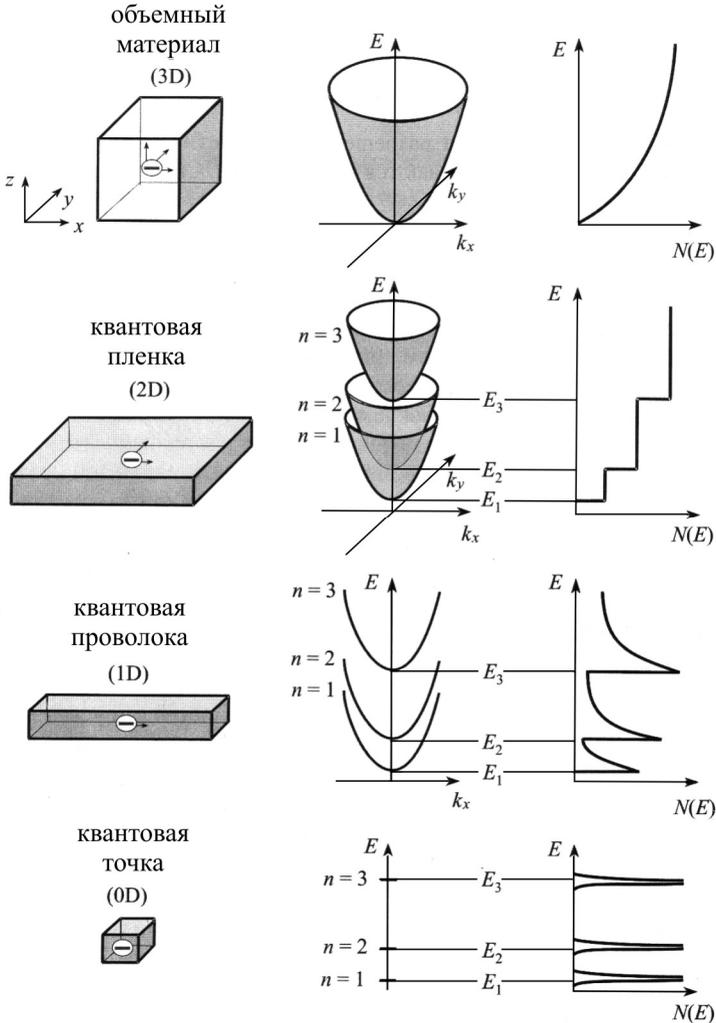


Рис. 8.5. Элементарные низкоразмерные структуры, их энергетические диаграммы и плотности электронных состояний в сравнении с трехмерной структурой

Квантовые пленки (*quantum films*) представляют собой двумерные (2D) структуры, в которых квантовое ограничение действует только в одном направлении – перпендикулярно пленке (направление z на рис. 8.5). Носители заряда в таких структурах могут свободно двигаться в плоскости xy . Их энергия складывается из квантованных значений, определяемых эффектом квантового ограничения в направлении z (в соответствии с толщиной пленки), и непрерывных составляющих в направлениях x и y . Энергетическая диаграмма квантовой пленки представляет собой семейство параболических зон, которые, перекрываясь, образуют подзоны. Минимальная энергия электрона в n -й подзоне мала, поэтому электрон с такой энергией неподвижен в плоскости пленки.

Зависимость плотности электронных состояний от энергии в квантовой пленке имеет ступенчатый вид (вместо параболической зависимости в трехмерных структурах); электроны в квантовых пленках обычно называют двумерным электронным газом (*two-dimensional electron gas*).

Квантовые проволоки (шнуры) (*quantum wires*) – это одномерные (1D) структуры. В отличие от квантовых пленок они имеют не один, а два нанометровых размера, в направлении которых и действует эффект квантового ограничения. Носители заряда могут свободно двигаться только в одном направлении – вдоль оси шнура. Таким образом, вклад в энергию носителя заряда дают кинетическая составляющая вдоль одного направления и квантованные значения в двух других направлениях.

Квантовые точки (*quantum dots*) – это нульмерные (0D) структуры, в которых движение носителей заряда ограничено во всех трех направлениях. В каждом из этих направлений энергия электрона оказывается квантованной, а плотность состояний представляет собой набор острых пиков. Из-за сходства энергетических характеристик атомов и квантовых точек последние иногда называют *искусственными атомами*. Квантовые точки состоят из сравнительно небольшого количества атомов. В этом

отношении к ним близки атомные кластеры и нанокристаллиты (кристаллиты нанометровых размеров), где также имеет место эффект квантового ограничения.

Рассмотренные элементарные низкоразмерные структуры в определенном смысле являются идеализированными объектами. Низкоразмерные структуры, представляющие практический интерес, должны располагаться на какой-либо подложке и иметь контакт с другими структурами и функциональными элементами. Более того, приборные применения требуют комбинации нескольких элементарных структур. Но, несмотря на появление в сложных комбинированных структурах новых квантово-механических эффектов, определяющую роль в них продолжает играть квантовое ограничение.

Для изготовления низкоразмерных структур используют два принципиальных подхода, которые можно охарактеризовать как «геометрический» и «электронный». Геометрический подход предполагает использование технологий, обеспечивающих формирование объектов с нанометровыми размерами. Для этого используются специальные нанотехнологические приемы. Электронный подход основан на возможности управления размерами областей с определенным типом и концентрацией носителей заряда в полупроводниках посредством электрического поля.

В качестве примера нульмерной квантовой структуры на рис. 8.6 показан массив квантовых точек.

Примером одномерной полупроводниковой структуры является периодическая структура кремний–воздух (рис. 8.7), в которой отношение показателей преломления кремния и воздуха составляет в ближней инфракрасной области 3,4 – беспрецедентно большое значение. Пористый кремний сегодня рассматривается как перспективный оптический материал, который позволит создавать оптоэлектронные системы высокой степени интеграции. Сочетание высоких кремниевых технологий с квантовыми размерными эффектами и принципами формирования фотонных запрещенных зон привело к развитию нового направления – кремниевой фотоники.

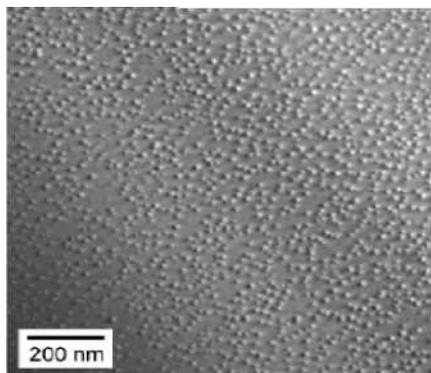


Рис. 8.6. Изображение массива квантовых точек, полученное с помощью просвечивающего электронного микроскопа

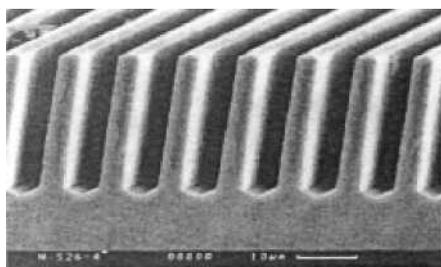


Рис. 8.7. Периодическая структура кремний–воздух, полученная методом анизотропного травления с использованием фотолитографической маски. Период структуры 8 мкм

Весьма интересными представляются периодические структуры на основе оксида алюминия. Они получают электрохимическим травлением металлического алюминия. С использованием электронно-литографических шаблонов получены совершенные двумерные периодические структуры, напоминающие пчелиные соты с диаметром пор менее 100 нм (рис. 8.8). Селективное травление алюминия при определенном сочетании условий травления позволяет получать регулярные структуры даже без использова-

ния каких-либо масок или шаблонов (рис. 8.9). Диаметр пор при этом может составлять всего несколько нанометров, что недостижимо для современных литографических методов. Периодичность пор связана с саморегуляцией процесса окисления алюминия при электрохимической реакции. Исходный проводящий материал (алюминий) в ходе реакции окисляется до Al_2O_3 .

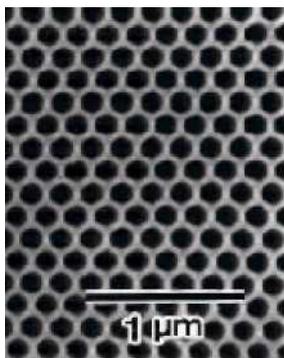


Рис. 8.8. Структура с двумерной периодичностью из оксида алюминия, полученная с использованием литографического шаблона

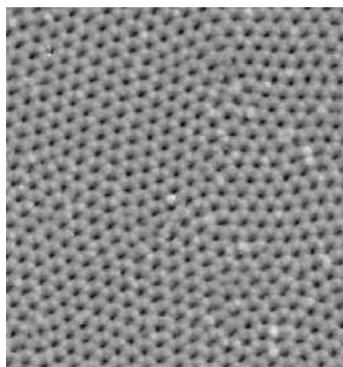


Рис. 8.9. Пористый оксид алюминия с регулярными порами, полученный без использования масок или шаблонов при травлении. Диаметр пор менее 50 нм. Нерегулярность пор обусловлена зернистой структурой исходной поликристаллической пленки алюминия

Трехмерные периодические структуры представляют наибольшие технологические трудности для экспериментальной реализации. Предложено два подхода к созданию диэлектрических структур с субмикронным периодом изменения показателя преломления. Первый основан на формировании плотноупакованных сферических глобул одинакового размера (коллоидные кристаллы), второй подход основан на построении многослойных структур с периодическим изменением показателя преломления в каждом слое.

Прототипами трехмерных фотонных кристаллов для оптической области электромагнитного спектра являются искусственные опалы, получаемые седиментацией (осаждением) частиц оксида кремния, их термообработкой до спекания в твердую структуру. Такие структуры получили название «инвертированных опалов» и в настоящее время активно синтезируются и исследуются в различных лабораториях. Эти структуры реализованы не только с использованием оксида кремния (рис. 8.10), но и оксидов титана, фуллеренов.

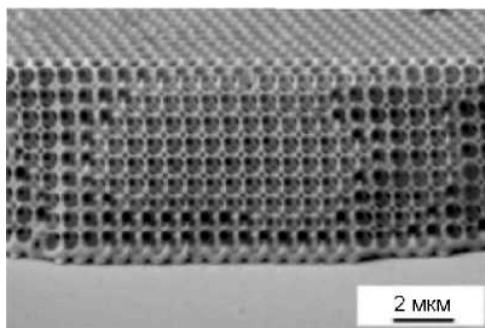


Рис. 8.10. Трехмерная структура из поликристаллического кремния, полученная с помощью опаловой матрицы

Из низкоразмерных наноструктур можно построить методами саморегулирования более сложные наноструктуры.

Саморегулирование является одной из наиболее общих закономерностей в природе. Оно осуществляется различными путями, но всегда с одной общей целью – обеспечить наибольшую устойчивость системы. В нанотехнологии практическое применение нашли *самосборка (self-assembly)* и *самоорганизация (self-organization)*.

Самосборка (самоупорядочение) – это процесс преимущественной концентрации молекул растворенного в жидкости вещества и формирования специфического расположения этих молекул на твердой поверхности (адсорбции). Ее движущей силой

является хемосорбция, которая проявляется в высокоэнергетических реакциях между адсорбатом и адсорбирующей поверхностью. В отличие от сильного взаимодействия между адсорбируемой молекулой и поверхностью, взаимодействие между самими молекулами остается слабым.

В органическом и неорганическом мире существует большое количество примеров самосборки.

Пленки мономолекулярной толщины, сформировавшиеся по механизму самосборки, имеют очень низкую плотность дефектов, достаточно стабильны и отличаются механической прочностью. Их используют в качестве трафарета для литографических процессов. При этом нанометровое разрешение достигается путем использования зондов сканирующего туннельного или атомного силового микроскопа.

Молекулярные блоки для самосборки должны содержать три основные функциональные группы: группу, прикрепляющую их к поверхности, промежуточную группу и поверхностную функциональную группу. Эти группы не являются взаимозаменяемыми. Комбинация различных по составу групп постоянно приводит к появлению новых форм самосборки.

В качестве групп, прикрепляющих весь молекулярный блок к поверхности подложки, чаще всего используют силаны RSiX_3 ($\text{R} = \text{CH}_3, \text{C}_2\text{H}_5, \dots$) Это нужно для образования связей с гидроксильными (ОН) группами, которые обычно покрывают поверхность кремния и другие технологически важные поверхности. В качестве X-компонента, замещающего в силане водород, используются метакси-группы, хлор или их комбинация. Состав прикрепляющей группы существенно влияет на упорядоченное расположение адсорбированных молекул и на плотность их упаковки. Например, для поверхности арсенида галлия и золота хорошие результаты дает тиол (RSH).

Промежуточная группа определяет взаимодействие всего хемосорбированного молекулярного блока с обрабатывающим его зондом. Отдаление поверхностной функциональной группы

от подложки при увеличении размеров промежуточной группы (например, путем повторения СНг-группы в ней) позволяет располагать зонд ближе к пленке и тем самым понижать дозу экспонирования и пороговое напряжение. Фенильные группы, обладая определенной проводимостью, хорошо подходят в качестве промежуточных групп при электронном экспонировании зондом сканирующего туннельного микроскопа.

Поверхностные функциональные группы определяют свойства «новой» поверхности. Например, аминовые группы (NH_2) могут быть использованы для прикрепления к ним определенных молекул. Галогены (хлор, йод и др.) имеют большие сечения электронного захвата, что облегчает десорбцию галогенсодержащих фрагментов. Их последующая обработка может осуществляться с целью замены галогенных групп более активными. Поверхности, покрытые алкильными группами, инертны и гидрофобны. По своей химической активности они идентичны парафину, вследствие чего хорошо подходят для масок, используемых при жидкостном травлении и ограниченно – при сухом травлении.

Одна из концепций создания наноструктур, получившая название «снизу – вверх», состоит в том, чтобы набрать, соединить и выстроить отдельные атомы и молекулы в упорядоченную структуру. Этот подход можно осуществить с помощью самосборки или некоторой последовательности каталитических химических реакций. Такие процессы широко распространены в биологических системах, где, например, катализаторы, называемые ферментами, собирая аминокислоты, формируют живые ткани, образующие и поддерживающие органы тела.

Противоположная концепция формирования наноструктур обозначается «сверху – вниз». При таком подходе процесс начинается с обработки макромасштабного объекта или структуры и состоит в постепенном уменьшении их размеров. Один из широко распространенных процессов этого класса называют литографией. Он состоит в облучении образца, покрытого слоем,

чувствительным к такому воздействию, через некоторый шаблон. Затем этот шаблон удаляется, а на поверхности с помощью химической обработки формируют наноструктуру.

Процесс, иллюстрирующий использование самосборки для создания нанометровых элементов на кремниевой подложке, показан на рис. 8.11.

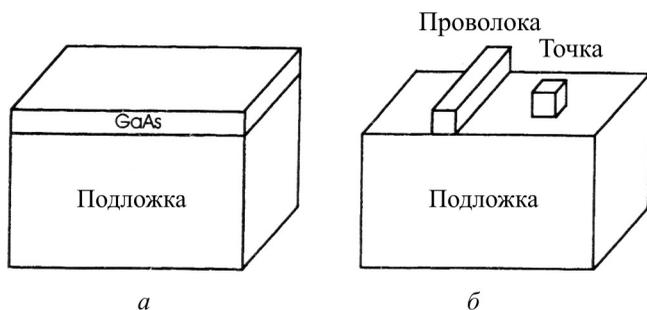


Рис. 8.11. Квантовая яма из арсенида галлия на подложке (а); квантовая проволока и квантовая точка, полученные методом литографии (б)

Перед нанесением пленки подложку очищают и пассивируют водородом в растворе HF. Затем ее окунают в раствор органосилового мономера и высушивают, чтобы сформировать на ее поверхности мономолекулярную пленку из молекул, один конец которых закреплен на подложке, а другой образует новую поверхность. Приготовленную таким образом мономолекулярную пленку, типичная толщина которой составляет около 1 нм, обрабатывают по требуемому рисунку низкоэнергетическими электронами, инжектируемыми с зонда сканирующего туннельного или атомного силового микроскопа. Напряжение при этом выбирают в диапазоне 2–10 В. После этого образец окунают в раствор с коллоидными частицами палладия, которые прикрепляются к необлученным областям пленки. Затем образец снова высушивают и помещают в ванну для электролитического осаждения никеля. Островки палладия на поверхности служат

каталитическими центрами для осаждения никеля. За счет бокового роста никелевых островков на палладии промежутки между палладиевыми островками заполняются, и образующаяся толстая пленка никеля имеет сплошную бездефектную структуру. Приготовленную таким образом профилированную металлическую пленку используют в качестве маски при последующем травлении. На практике разрешение при этом составляет 15–20 нм, хотя с теоретической точки зрения минимальный размер элементов интегральной микросхемы ограничен лишь размерами используемых для самосборки молекул.

Монокристаллическую пленку из одного материала, воспроизводящую постоянную решетки монокристаллической подложки из другого материала, называют *сверхрешеткой* (*superlattice*). Когда оба материала имеют идентичные или очень близкие постоянные решеток, они образуют так называемые *псевдоморфные* (*pseudomorphic*) сверхрешетки. Среди полупроводников таких материалов очень мало. Между тем равенство постоянных решеток не является строго необходимым условием для псевдоморфного роста одного материала на другом.

В пределах некоторой ограниченной толщины наносимой пленки возможно «заставить» осаждаемые атомы занимать позиции, соответствующие расположению атомов в подложке, даже если это расположение отличается от равновесного расположения атомов в объемном материале пленки. При этом образуется *напряженная* (*strained*) сверхрешетка, структура которой, однако, совершенна. Формирование напряженной сверхрешетки в случае, когда постоянная решетки у материала подложки меньше, чем у материала пленки, схематически проиллюстрировано на рис. 8.12. Напряжения в такой пленке возрастают по мере увеличения ее толщины. По достижении некоторой критической толщины они релаксируют посредством образования дислокаций несоответствия, высвобождая накопленную в напряженном состоянии энергию и понижая полную энергию системы. Кристаллическая решетка наносимого материала приоб-

ретает свой естественный вид, и при дальнейшем поступлении материала на подложку пленка растет с уже *релаксированной* (*relaxed*) решеткой. Критическая толщина пленки зависит от величины рассогласования постоянных решеток и модулей упругости материалов пленки и подложки при температуре осаждения. В принципе, не превышая критической толщины, можно сформировать напряженную сверхрешетку из любого полупроводника на подложке с тем же типом кристаллической решетки.

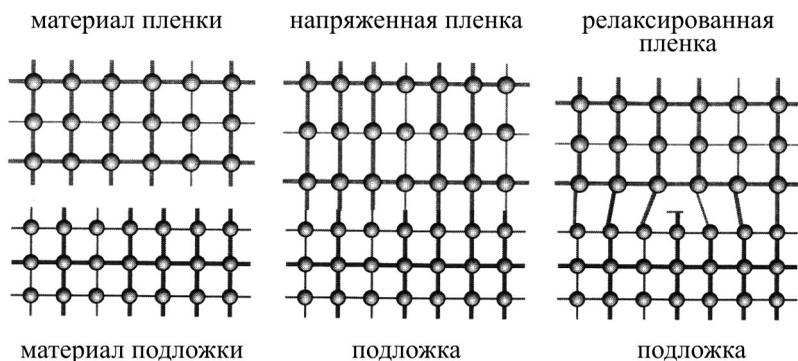


Рис. 8.12. Образование напряженной и релаксированной эпитаксиальной пленки

Полупроводниковые сверхрешетки состоят из чередующихся слоев двух полупроводников (рис. 8.13), различающихся составом или типом проводимости. Период повторения слоев составляет от нескольких нанометров до десятков нанометров.

Широкое применение находят два типа сверхрешеток: *композиционные* и *легированные*. Композиционные сверхрешетки – это *гетероструктуры* из чередующихся слоев различного состава и ширины запрещенной зоны, но с близкими значениями постоянной решетки. Легированные сверхрешетки – это периодическая последовательность слоев *n*- и *p*-типа одного и того же полупроводника.

Дополнительный периодический потенциал сверхрешетки изменяет зонную структуру исходных полупроводников. Поэтому

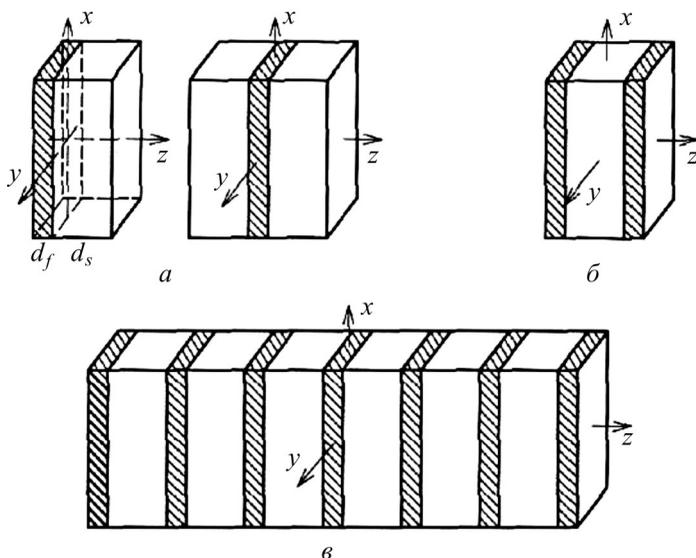


Рис. 8.13. Многослойные гетероструктуры:
 а – бислои, б – трислои, в – сверхрешётки

сверхрешетку можно рассматривать как новый, синтезированный полупроводник, обладающий необычными свойствами. Подбором материала и состава чередующихся слоев можно варьировать зонную структуру сверхрешетки в широких пределах. Это достигается совокупностью методов *зонной инженерии*.

Сверхрешетки используются в полупроводниковых приборах: лазерах, светодиодах, фотоприемниках, оптоволоконных волноводах и др. Принципы действия подобных приборов основаны на характерных для сверхрешеток явлениях: квантовом ограничении носителей зарядов в потенциальных ямах, пространственном разделении электронов и дырок, резонансном туннелировании, возможности перестройки энергетического спектра.

В гетеропереходе претерпевают скачки параметры полупроводников: ширина запрещенной зоны, подвижность носителей заряда, их эффективные массы и т.д. Скачкообразное изме-

нение свойств полупроводника на гетеропереходе дает возможность целенаправленно управлять этими свойствами путем подбора сопрягаемых полупроводниковых материалов. Гетеропереходы используются для совершенствования существующих полупроводниковых приборов и создания принципиально новых приборов различного назначения.

8.4. Применение квантовых структур в приборах оптоэлектроники

Оптоэлектроника – научно-техническое направление, в котором исследуются и используются эффекты взаимного преобразования электрических и оптических сигналов в веществе и на этой основе создаются устройства для генерации, передачи, хранения, обработки и отображения информации.

К основным элементам оптоэлектроники относятся транзисторы, полупроводниковые источники некогерентного (светодиоды) и когерентного (лазеры) излучения, а также полупроводниковые фотоприемники. В основе действия многих вариантов перечисленных приборов лежат p – n -переходы. Более эффективными для оптоэлектроники являются гетеропереходы на границах контакта двух различных по химическому составу полупроводников. Особенно широко используются в оптоэлектронике гетеропереходы на основе соединений элементов III и V групп таблицы Менделеева: Si–Ge, GaAs–Ge и т.д.

Транзисторы

*Транзисторами (триодами) называются устройства, предназначенные для усиления и генерирования электромагнитных колебаний. Термин «транзистор» (от англ. *transfer* – переносить и *resistor* – сопротивление) означает трехэлектродный полупроводниковый электронный прибор, в котором ток в цепи двух электродов управляется третьим. Действие транзистора можно сравнить с действием плотины, которая, перегораживая реку (по-*

стоянный источник), создает перепад уровней воды. Затрачивая очень небольшую энергию на вертикальное перемещение затвора, мы можем управлять потоком воды огромной мощности, т.е. энергией мощного постоянного источника.

В основу первого поколения транзисторов был положен эффект эмиссии электронов из нагретого катода и управления потоком этих электронов в вакууме с помощью электрического поля. Термоэлектронная эмиссия была открыта американским изобретателем Т.А. Эдисоном (1889) и исследована английским физиком О.В. Ричардсоном (Нобелевская премия, 1928). На основе описанного эффекта был создан активный схемный элемент – радиолампа, позволяющая выпрямлять и усиливать электрический сигнал, а также генерировать электромагнитные колебания.

Радиолампа (рис. 8.14) – это стеклянная вакуумная колба (лампа) с двумя (катод, анод) электродами и третьим управляющим электродом (сеткой). Катод при нагреве создает вблизи своей поверхности некоторую концентрацию электронов. «Тянущее» поле положительно заряженного анода формирует из этих электронов анодный ток. С помощью электрического потенциала, поданного на сетку, можно управлять анодным током, например усиливать его. Это позволяет

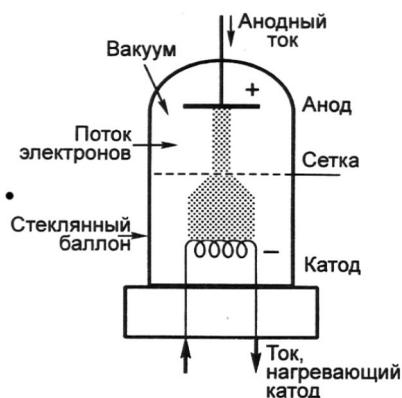


Рис. 8.14. Схематическое представление триода

использовать триод в электронных схемах в качестве усилителя и генератора сигналов. Поэтому триод относят к активными схемным элементам в отличие от пассивных – резисторов (электрические сопротивления), конденсаторов (электроемкости), катушки индуктивности.

К пятидесятым годам прошлого столетия ламповая электроника полностью исчерпала возможности удовлетворять запросы потребителей электронной техники. Возникла принципиально иная электроника, основанная на особых свойствах полупроводниковых структур.

В основе полупроводниковых транзисторов лежат $p-n$ -переходы. Большая разновидность транзисторов разделяется на две группы – *биполярные и полевые (униполярные)*. Протекание тока в полевом (униполярном) транзисторе обусловлено носителями заряда только одного знака – электронами или дырками (основные носители). В биполярном транзисторе ток обусловлен движением зарядов обоих знаков.

В униполярном транзисторе ток протекает в узком канале и регулируется внешним полем, перпендикулярным току. Поэтому подобные транзисторы называются еще и каналными или полевыми. Регулирующее поле создается специальным электродом – затвором, который может работать на основе $p-n$ -перехода.

В структуре любого транзистора есть три вывода – это *затвор (база), исток (эмиттер) и сток (коллектор)* (рис. 8.15). Управление током в выходной цепи осуществляется либо за счет изменения входного тока, либо входного напряжения. При этом даже небольшое варьирование входных величин может приводить к существенному изменению выходного напряжения и тока. Если пропустить через участок «затвор – исток» слабый ток, он будет усилен транзистором в десятки и даже сотни раз, а усиленный ток потечет через участок «исток – сток». Таким образом, при помощи изменения напряжения на затворе можно регулировать ток между истоком и стоком. В этом смысле затвор является аналогом сетки лампового триода (см. рис. 8.14); исток и сток – аналоги катода и анода. Усиление тока связано с тем, что внешние электрические поля и токи могут изменять плотность носителей заряда в полупроводнике и оказывать существенное влияние на его электропроводность.

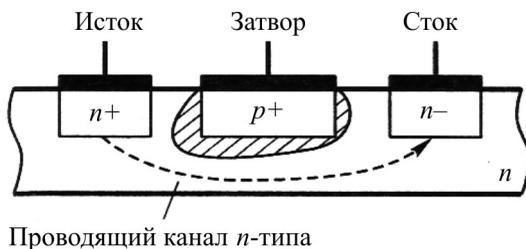


Рис. 8.15. Структурная схема униполярного транзистора с затвором на основе p - n -перехода

Важнейшей сферой применения транзисторов является цифровая техника (память, процессоры, компьютеры, цифровая связь и т.п.), где они исполняют роль переключателей. В настоящий момент вся современная цифровая техника основана на так называемых МОП-транзисторах, изготовленных на основе трехслойной структуры (металл – оксид – полупроводник). Транзистор может работать в составе интегральной схемы, собираемой на одном кремниевом кристалле – *чипе*. Чип составляет элементарный «кирпичик» для построения памяти, процессора и т.п. На одном чипе, обычно размером 1–2 см, размещаются десятки миллионов МОП-транзисторов, размеры каждого из которых не превышают 45–60 нанометров (рис. 8.16).

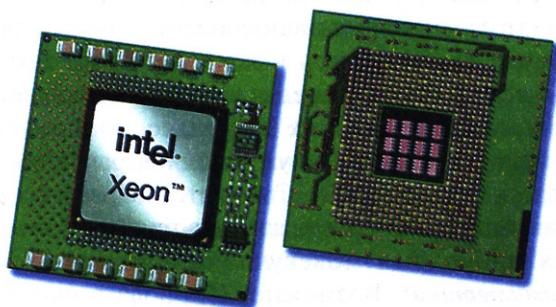


Рис. 8.16. Современный процессор производства компании Intel. Размер элементов транзистора в таких процессорах менее 50 нм

На протяжении последних десятков лет происходит стремительная миниатюризация (т.е. уменьшение размеров) МОП и увеличение степени их интеграции (т.е. количества на одном чипе), причем в ближайшие годы ожидается увеличение степени интеграции вплоть до миллиарда транзисторов на одном чипе. Однако полупроводниковая кремниевая электроника фактически подошла к пределу своих возможностей, связанному с фундаментальными физическими ограничениями, не позволяющими в дальнейшем на ее основе создавать все более производительные и миниатюрные устройства. Традиционный затвор с диэлектриком из двуокиси кремния (SiO_2) имеет толщину всего в несколько атомных слоев (~1,2 нм). Дальнейшее уменьшение его толщины приводит к значительным утечкам за счет туннельного тока (проявление квантовых эффектов) и, как следствие, к увеличению потребления энергии и тепловыделения транзистора.

Следующим шагом миниатюризации полевых транзисторов стало применение полупроводниковых углеродных нанотрубок, соединяющих два золотых электрода. Схематически такое устройство показано на рис. 8.17. При приложении небольшого напряжения к затвору, которым является кремниевая подложка, по нанотрубке между истоком и стоком течет ток. Если ток течет, элемент находится в состоянии «включено», и в состоянии «выключено» – в противном случае. Обнаружено, что небольшое напряжение на затворе может изменить проводимость нанотрубки более чем в 10^6 раз, что сравнимо со значениями для кремниевых полевых транзисторов. Время переключения такого устройства будет очень маленьким, а возможная тактовая частота составляет Терагерцы, что в 1000 раз быстрее тактовой частоты существующих процессоров. Золотые исток и сток формируются методами нанолитографии, а диаметр соединяющей их нанотрубки составляет около одного нанометра. Такие малые размеры позволяют поместить на чип еще большее количество переключателей.

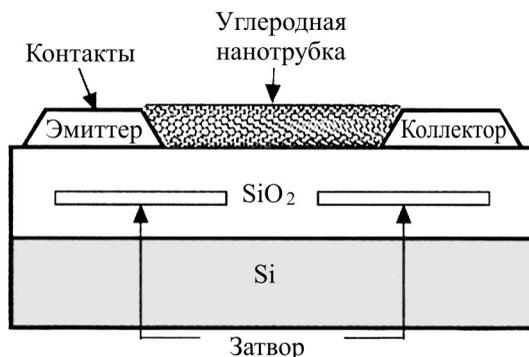


Рис. 8.17. Схема полевого транзистора на основе углеродной нанотрубки

Следует отметить, что помимо разновидностей полупроводниковых транзисторов ведутся разработки объектов совершенно иной категории – одноэлектронных транзисторов, работающих на одной единственной молекуле. На 1 см^2 поверхности возможно размещение 10^{13} молекулярных транзисторов, что в 10^4 раза больше плотности сборки в современных чипах. Время отклика молекулярного транзистора на внешнее воздействие равно $\sim 10^{-15}$ с, тогда как в современных устройствах оно составляет $\sim 10^{-9}$ с. В итоге эффективность молекулярного транзистора по сравнению с современными транзисторами должна повыситься в $\sim 10^{10}$ раз.

Однако ключевой проблемой молекулярной электроники остается интеграция молекул в схему. Принцип решения проблемы ясен – это должен быть процесс *самосборки*, основанный на молекулярном распознавании взаимно дополняющих структур. Поэтому специалисты предсказывают появление молекулярных транзисторов (и компьютеров на их основе) примерно к 2015 году.

Разрабатываются и *оптические транзисторы* как основные элементы для *фотоники*, в которых в качестве передающего звена выступают не электроны, а фотоны.

Светодиоды

Светоизлучающий диод (СИД) – это полупроводниковый диод, излучающий свет при пропускании тока через $p-n$ -переход в прямом направлении. Физическую основу работы СИД составляют процессы *инжекции* неосновных носителей заряда в активную область $p-n$ -структуры и излучательная рекомбинация (выделение энергии при переходе электрона из зоны проводимости в валентную зону) инжектированных (от лат. *injectio* – вбрасывание) носителей.

СИД включает активный элемент из полупроводникового монокристалла в виде кубика (чипа) с типичным размером $0,3 \times 0,3 \times 0,25$ мм, содержащего $p-n$ - или гетеропереход и омические контакты для управления питающим напряжением. СИД может иметь усложненную конструкцию, повышающую эффективность ввода излучения в волокно.

По сравнению с полупроводниковыми лазерами СИД обладают более широким спектром излучения, полоса модуляции у них уже, а эффективность связи с оптическим волокном ниже. Однако, не будучи «пороговым прибором», они имеют также ряд преимуществ: высокую надежность (большой срок службы), простоту изготовления, низкую себестоимость. Поэтому они находят широкое применение в качестве источников света в системах передачи на небольшие расстояния с шириной полосы до нескольких десятков МГц, например для внутренней связи в пределах здания, самолета, корабля, автомобиля и т.д.

По способу вывода излучения СИД подразделяются на диоды с *поверхностными излучателями* и на диоды с *торцевыми излучателями*.

В СИД с поверхностным излучателем свет излучается в направлении, перпендикулярном поверхности перехода (рис. 8.18, *a*). Для улучшения отвода тепла от активного слоя одна сторона выращенного слоя прижимается к тепловоду, а вывод излучения осуществляется со стороны подложки. Для вывода излучения через

поглощающую подложку на арсениде галлия в AlGaAs СИД в подложке вырезается круглое отверстие и вводится оптическое волокно. Такую конструкцию диода называют *диодом Барраса*.

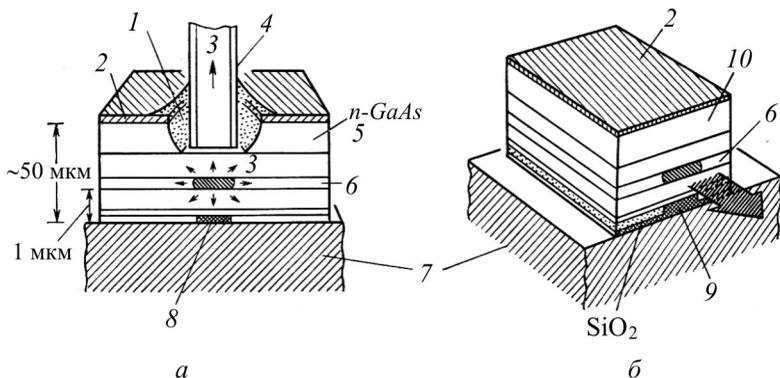


Рис 8.18. Структура СИД с поверхностным излучателем (а) и торцевым излучателем (б): 1 – эпоксидная смола; 2 – омический контакт; 3 – излучение; 4 – оптическое волокно; 5 – подложка на $n\text{-GaAs}$; 6 – активный слой; 7 – теплоотвод; 8 – нижний омический контакт (диаметром 50 мкм), 9 – нижний омический контакт (ширина 65 мкм); 10 – подложка

В СИД с торцевыми излучателями (рис. 8.18, б) вывод излучения, выходящего из активного слоя, осуществляется с торца, как в полупроводниковых лазерах. Для уменьшения самопоглощения применяется следующий способ: по соседству с активным слоем формируется световодный слой с малыми внутренними потерями, активный слой делается тонким, порядка 0,03–0,1 мкм. В результате по сравнению с диодами с поверхностным излучателем яркость оказывается в 5–10 раз большей.

Чип, содержащий $p\text{-}n$ -переход, может использоваться в *индикаторном СИД*, представляющем собой конструкцию (рис. 8.19), размещенную внутри линзообразной эпоксидной смолы (компаунда). Эффективный вывод излучения осуществляется с помощью отражательной пластины. Такие СИД в большинстве случаев применяются как миниатюрные элементы индикации в виде точек и цифр.

Простая структура кристалла с $p-n$ -переходом обеспечивает низкую стоимость индикаторных СИД, которая является одним из важнейших предъявляемых к ним параметров.

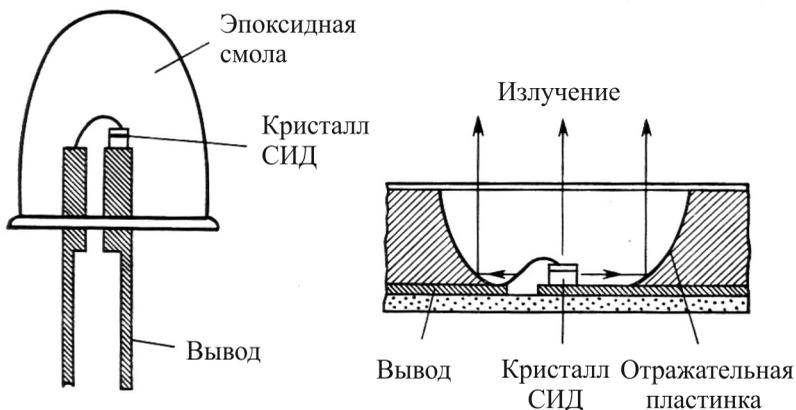


Рис. 8.19. Конструкция индикаторного СИД

Для индикации используется видимый свет с видимым излучением от зеленого до красного цвета. Для повышения КПД источника излучения часто применяется введение примеси азота, цинка, кислорода, которые становятся центрами люминесценции (излучения).

Физический принцип функционирования *газоразрядных индикаторов* (ГРИ) основан на преобразовании электрической энергии в световую при возбуждении «электронным ударом» атомов газа, который в рассматриваемом случае является активной средой, с последующим преобразованием энергии возбуждения в видимое излучение. В качестве газа-наполнителя, как правило, используется неон с небольшим количеством аргона или ксенона. Оптимальные параметры газового наполнения выбираются с учетом обеспечения необходимой яркости свечения, минимизации напряжения разряда, потребляемой мощности, а также обеспечения заданных частотных характеристик и увеличения долговечности устройств. Поскольку собственное све-

чение газа-наполнителя имеет оранжевый цвет, варьирование цвета свечения достигается за счет использования различных фотолуминофоров.

Принципиальная схема ячейки ГРИ, на основе которой формируется конструкция ГРИ-панели, приведена на рис. 8.20. Несмотря на сложную систему управления, содержащую высоковольтные элементы, и невысокую яркость, ГРИ-панели площадью до 3 м^2 применяются в системах коллективного пользования.

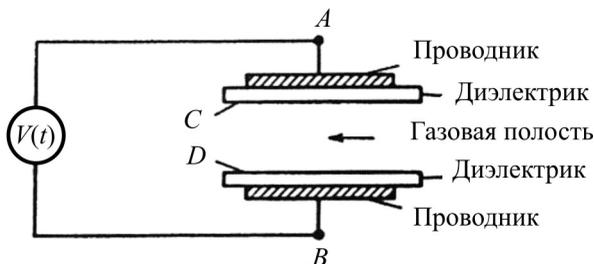


Рис. 8.20. Принципиальная схема ГРИ-ячейки

Электролюминесцентные индикаторы (ЭЛИ) являются еще одной разновидностью индикаторных систем плоской конструкции. Типичная конструкция тонкопленочного ЭЛИ представлена на рис. 8.21.

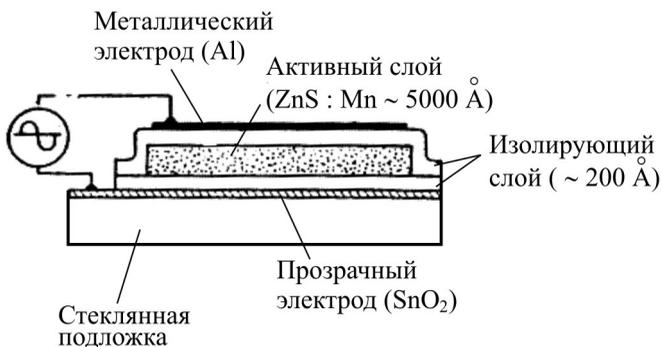


Рис. 8.21. Конструкция электролюминесцентного индикатора

Технология тонкопленочных ЭЛИ предусматривает последовательное нанесение на стеклянную подложку методом испарения: изолирующего слоя окиси иттрия, люминофора, в качестве которого наиболее часто используется сульфид цинка с примесью марганца, верхнего изолирующего слоя окиси иттрия, а также алюминиевых электродов. Генерация света происходит в ЭЛИ в области толщиной менее 2 мкм, перпендикулярной электродам, расстояние между которыми задается с помощью фотолитографических методов с точностью до 1 мкм.

ЭЛИ-панели характеризуются высоким контрастом даже в условиях интенсивной внешней засветки. Для получения полноцветного изображения используется широкая гамма материалов для люминофорных сред.

В последнее время достаточно большую популярность получили индикаторы на основе *микроэлектромеханических систем* (MEMS-индикаторы). Индикаторные системы MEMS представляют собой массивы модуляторов света, работающих как на просвет, так и на отражение, при этом их конфигурация может быть как матричной (двумерной), так и одномерной (в виде линейной шкалы).

В основе MEMS-технологии лежит формирование подвижных или деформируемых отражателей, создаваемых в кремниевой подложке, при этом схема управления индикатором может быть сформирована на этой же кремниевой подложке.

Некоторые примеры MEMS-технологий представлены на рис. 8.22. После светофильтра световой поток проходит через фокусирующую оптику и попадает на поверхность микрозеркального модулятора, после чего промодулированный пучок фокусируется в плоскости выходного объектива и направляется на экран.

В этой схеме микрозеркала имеют два рабочих состояния (открытое состояние, при котором луч попадает на экран, а также закрытое состояние, когда он попадает в оптический поглотитель).

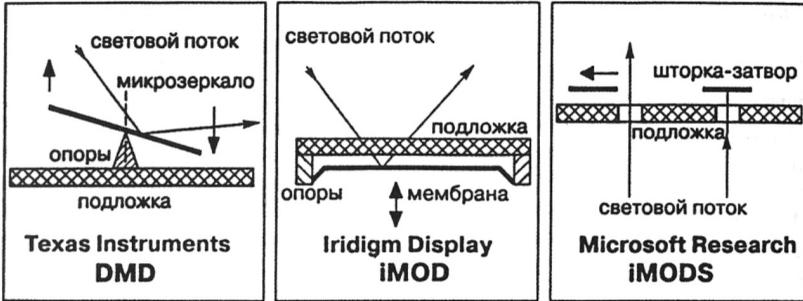


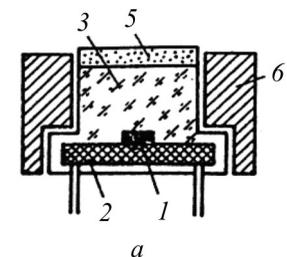
Рис. 8.22. Схемы различных технологических подходов MEMS-технологий

Каждый из этих технологических подходов имеет свои отличительные признаки, обусловленные нацеленностью на определенные сектора рынка индикаторных систем: в некоторых случаях требуется сложная оптика (DMD), в других же случаях (iMOD, iMODS) она не требуется.

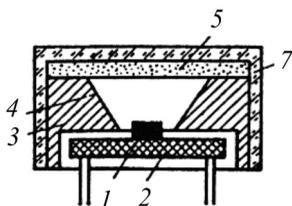
Полупроводниковые знаковосинтезирующие индикаторы (ПЗСИ) являются одним из наиболее универсальных классов приборов для отображения информации разнообразного характера, которые обладают оптимальной совокупностью свойств, обеспечивающих их широкое использование в устройствах и системах индивидуального, группового и даже коллективного пользования.

Необходимость увеличения размера знака при одновременном рациональном использовании полупроводниковых структур привела к разработке гибридных ПЗСИ, основанных на принципе рассеяния света (рис. 8.23).

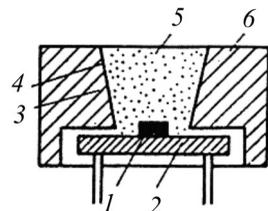
В такой конструкции формирование изображения осуществляется при использовании светоизлучающих кристаллов, размещенных в светорассеивающей полости светопровода. В этом случае размеры и форма светящейся области определяются габаритами и конфигурацией светопровода. Можно выделить три варианта конструкции светопровода:



а



б



в

Рис. 8.23. Различные варианты конструкции полупроводниковых знаковсинтезирующих индикаторов на принципе рассеяния света:

- 1 – кристалл-излучатель; 2 – держатель; 3 – светопроводящая полость; 4 – отражающие стенки; 5 – частицы рассеивателя; 6 – корпус; 7 – прозрачная крышка-корпус

электрического управления двулучепреломлением, «твист-эффект») и полевые эффекты (фазовый переход, эффект «гость-хозяин»).

– пластмассовый светопровод с верхним рассеивающим слоем (рис. 8.23, а);

– полый светопровод с рассеивающей пленкой (рис. 8.23, б);

– светопровод с отражающими зеркальными стенками (рис. 8.23, в).

Жидкокристаллические индикаторы (ЖК-индикаторы) широко применяются для буквенно-цифровой индикации в часах, настольных микрокалькуляторах, в плоских телевизионных экранах больших размеров. Они обладают рядом достоинств: возможностью исполнения в виде плоской конструкции, низким управляющим напряжением, простотой исполнения управляющей схемы, малой потребляемой мощностью и т.д. Однако они недостаточно эффективны с точки зрения контрастности изображения.

Рассмотрим электрооптические эффекты в ЖК-индикаторах, которые подразделяются на токовые эффекты (динамическое рассеяние, эффект

Динамическое рассеяние света. Если через слой нематического ЖК с отрицательной диэлектрической анизотропией пропустить постоянный или переменный ток низкой частоты, то прозрачный слой ЖК мутнеет: происходит рассеяние света. Поскольку ЖК-ячейка довольно «толстая» (больше 6 мкм), разрушение ранее упорядоченной структуры и перевод слоя жидкости в состояние турбулентности, в котором осуществляется рассеяние света, требует приложения сильного электрического поля. При подмешивании холестерического ЖК нематический ЖК будет испытывать влияние со стороны «закрученной» структуры его молекул. Возникает динамическое рассеяние света, которое сохраняется даже после снятия электрического поля. Это состояние устраняется применением высокочастотного электрического поля, в котором исчезает динамическое рассеяние.

Эффект электрического управления двулучепреломлением. На ориентацию молекул ЖК оказывает влияние состояние поверхностей пластин ЖК-ячейки. Если пластины обработать поверхностно-активным веществом, то, как показано на рис. 8.24, можно получить *гомеотропную* упаковку молекул, перпендикулярную пластинам, и *гомогенную* упаковку молекул, параллельную пластинам. Вне ЖК-ячейки скрещенно размещают поляризатор и анализатор. Для индикации используется свойство двойного лучепреломления. Свет, ставший линейно-поляризованным после прохождения поляризатора, становится эллиптически поляризованным после прохождения ЖК-ячейки, обладающей свойством двойного лучепреломления. Тем самым интенсивностью света, прошедшего через ЖК-ячейку, можно управлять с помощью приложенного электрического поля. Так как интенсивность проходящего света зависит от длины волны, то, изменяя напряжение, можно менять цветовой тон.

«*Твист-эффект*» реализуется при помощи однонаправленного натирания поверхностей пластин во взаимно перпендикулярных направлениях и введения нематического жидкого кристалла с положительной диэлектрической анизотропией. Поскольку

молекулы жидкого кристалла между двумя пластинами оказываются скрученными на 90° , то происходит поворот плоскости поляризации линейно-поляризованного света, прошедшего через ЖК-ячейку. Приложенное к электродам низкое рабочее напряжение (менее 1 В) позволяет управлять прохождением света.

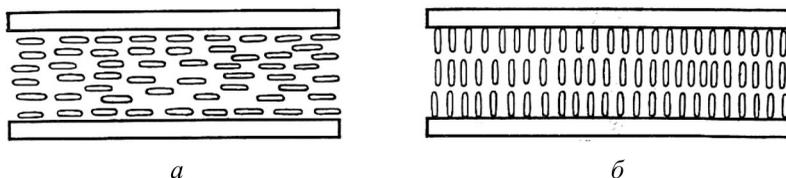


Рис. 8.24. Упаковка молекул в ЖК:
a – гомогенная; *б* – гомеотропная

Фазовый переход связан с изменением состояния нематического жидкого кристалла при изменении электрического поля с гомеотропной упаковкой молекул, при котором ЖК-ячейка прозрачна, в состояние, при котором ЖК-ячейка становится непрозрачной, рассеивая свет.

Оптический эффект «гость – хозяин». Краситель, обладающий свойством менять спектр проходящего света в зависимости от ориентации молекул, называется *плеохроическим* красителем. Если к жидкому кристаллу подмешать краситель, обладающий свойством менять спектр проходящего света в зависимости от ориентации молекул, тогда с помощью электрического поля оказывается возможным менять как ориентацию молекул ЖК, так и ориентацию молекул красителя, т.е. можно изменять цвет ЖК-ячейки. В такой ЖК-ячейке ЖК называют «хозяином», а краситель – «гостем», а само явление называют оптическим эффектом «гость – хозяин». На рис. 8.25 показано превращение ЖК *p*-типа с гомогенной упаковкой молекул в кристалл с гомеотропной упаковкой под действием электрического поля.

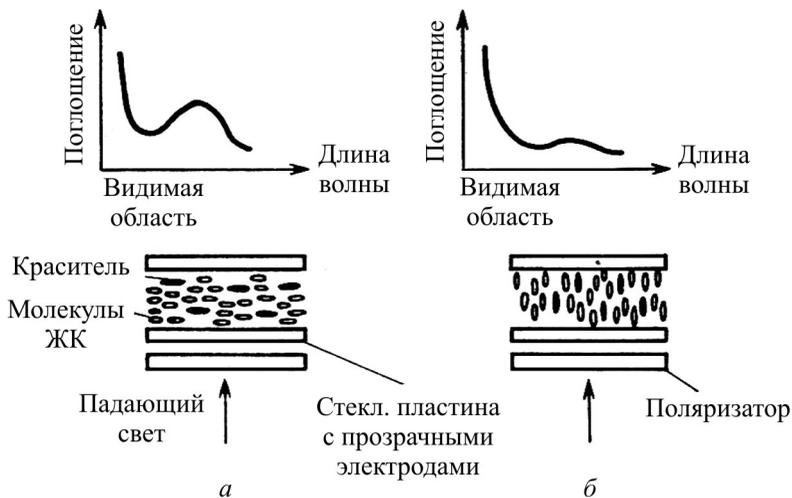


Рис. 8.25. Оптический эффект «гость – хозяин»:
 a – без электрического поля; b – в электрическом поле

Несмотря на то, что технология жидкокристаллических экранов достигла зрелого состояния, в настоящее время из-за неудовлетворительной цветопередачи и других недостатков идут интенсивные поиски альтернативных технологий. В качестве одной из возможных альтернатив в последнее время рассматриваются индикаторные системы на основе *органических электролюминесцентных структур* (OLED).

История разработки OLED, по-видимому, начинается с 1965 года, когда в национальной химической лаборатории (Оттава) было обнаружено голубое свечение антрацена. Следующий этап развития OLED начинается с исследований физика Р.Г. Фрэнда и химика А.Н. Холмса (Кембридж) в области молекулярной электроники с использованием полимера полипарафениленвинилена (PPV). Типичная конструкция OLED на основе PPV представлена на рис. 8.26.

В то время как в современных ЖКИ-экранах яркость проходящего через фильтры излучения подсветки составляет величину

порядка 100 Кд/м^2 , в первых приборах нового поколения этот параметр был в 50–60 раз выше (на уровне $5000\text{--}6000 \text{ Кд/м}^2$).

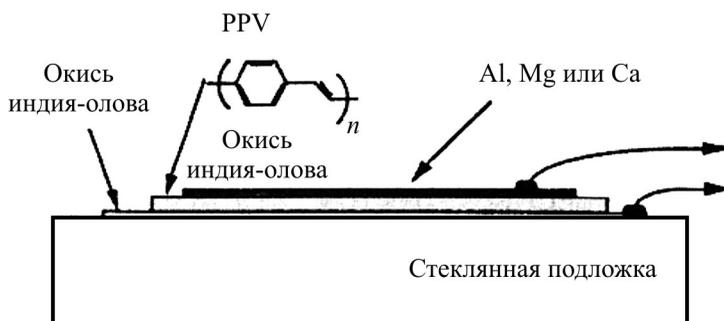


Рис. 8.26. Конструкция OLED на основе PPV

Интересным фактом, общим как для твердотельной, так и органической оптоэлектроники, является поистине стремительная динамика улучшения ключевых светотехнических параметров приборов начиная с 1990 года, при этом по уровню светоотдачи параметры OLED превосходят достижения СИД предыдущих поколений (рис. 8.27). Сверхяркие СИД, как видно на рис. 8.27, по светоотдаче на порядок превосходят лампы накаливания и значительно превосходят галогенные лампы.

Успешное решение физико-технологических проблем позволило в последнее время разработать и освоить в серийном производстве такой важный класс приборов, как *оптроны*, сочетающие в себе функции как генерации, так и регистрации излучения. Оптрон – это оптоэлектронный прибор, состоящий из оптического излучателя и фотоприемника (ФП). Физическую основу работы оптрона составляют процессы преобразования электрических сигналов в оптические (в излучателе), оптических сигналов – в электрические (в ФП).

ФП представляют собой приборы, в которых оптическая (световая) энергия преобразуется в электрическую энергию. Существуют следующие разновидности фотоприемников: *фото-*

диоды, солнечные батареи, лавинные фотодиоды, фототранзисторы, фоторезисторы. Принцип действия этих фотоприемников заключается в использовании эффектов: фотогальванического, фоторезистивного (внутреннего фотоэффекта), внутреннего усиления.



Рис. 8.27. Динамика улучшения светотехнических параметров твердотельных и органических СИД

Оптроны используются в датчиках, преобразующих электромагнитное излучение из одного диапазона в другой, например инфракрасное излучение в видимое и т.п. Наиболее распространенная конструкция этого прибора представлена на рис. 8.28.

С появлением технологии сверхъярких СИД, способных обеспечить передачу широкой гаммы цветов при одновременной высокой эффективности процесса излучательной рекомбинации, наблюдается подлинный бум в области нового поколения разнообразных систем визуального отображения информации коллективного и группового пользования. Одной из иллюстраций перспективности такого подхода является разработка миниатюрного дисплея, имеющего размеры 30×33×89 мм и дающего пользователю впечатление восприятия изображения на 30 см

мониторе с использованием СИД-технологии (рис. 8.29). Прибор имеет массу 70 г, потребляет менее 5 Вт и может работать от батарейки. Его можно прикрепить к легкому держателю наушника или оправе очков.

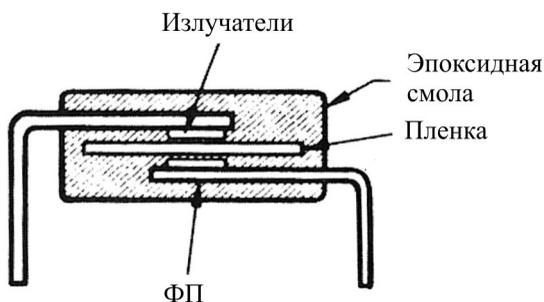


Рис. 8.28. Конструкция оптрона в керамическом корпусе

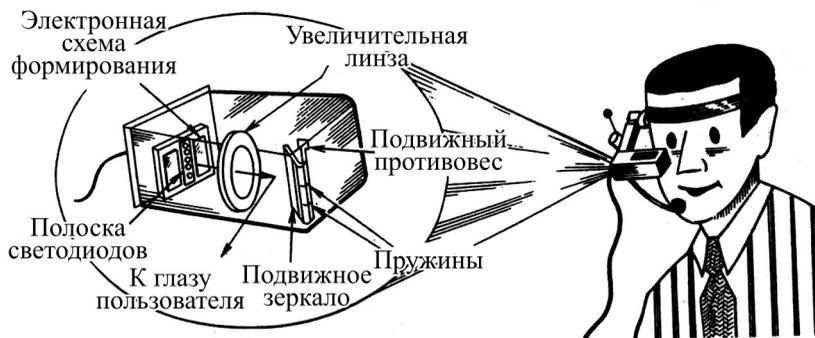


Рис. 8.29. Миниатюрный дисплей

Предельные возможности полупроводниковых излучателей демонстрируют лазеры с вертикальным резонатором (ЛВР), которые отличаются от обычных полосковых лазеров расположением зеркал резонатора Фабри-Перо параллельно плоскости полупроводниковой пластины. ЛВР характеризуются низкими значениями порогового тока, высокой частотой токовой модуляции (десятки гигагерц) и сверхминиатюрностью.

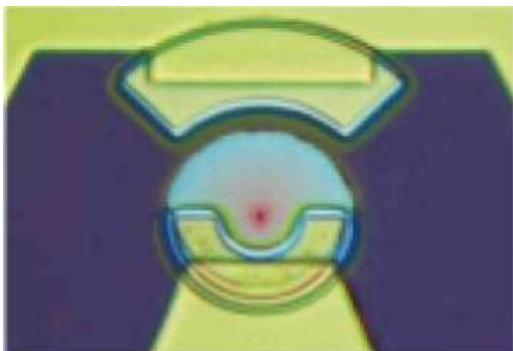


Рис. 8.30. Излучение диода с InAs с квантовой точкой в активном слое при токовой апертуре $A < 1$ мкм (яркая точка в центре)

Важный шаг на пути реализации однофотонных излучателей сделан при использовании в качестве активного слоя InAs квантовых точек малой плотности. При малой апертуре (от лат. *apertura* – отверстие) при токовом возбуждении наблюдалось излучение одиночной квантовой точки (рис. 8.30) Присутствие в спектре излучения лишь одной линии, соответствующей рекомбинации экситона одиночной InAs квантовой точки, подтвердило однофотонный характер излучения. Ожидается, что именно на базе ЛВР будут созданы полупроводниковые однофотонные СИД для использования в будущих системах квантовых вычислений и квантовой криптографии.

Вопросы для самоконтроля

1. Что называют гетеропереходом, гетероструктурой?
2. Объясните явления инжекции и рекомбинации электрических зарядов для p – n -переходов.
3. Объясните закономерности вольт-амперной характеристики для p – n -перехода.
4. Какие квантоворазмерные структуры вам известны? Каковы особенности энергетических диаграмм и плотностей электронных состояний в этих структурах?

5. В чем заключается процесс самосборки квантоворазмерных структур?

6. Что называют сверхрешеткой? Чем характеризуются псевдоморфные и напряженные сверхрешетки?

7. Каково различие между композиционными и легированными сверхрешетками?

8. Объясните физический принцип работы транзистора.

9. В чем заключается отличие биполярных транзисторов от полевых?

10. Назовите области применения транзисторов. Каковы пути повышения эффективности транзисторов в современных электронных и оптических схемах?

11. Объясните физический принцип работы светодиода.

12. Объясните физические принципы работы светодиодов: газоразрядных, электролюминесцентных, микроэлектромеханических, жидкокристаллических.

13. В чем заключается преимущество диодов на основе органических электролюминесцентных структур? Для чего нужны диоды с квантовой точкой? Дайте сравнительную характеристику эффективности различных типов светодиодов.

14. Что такое оптроны? Каковы области их применения?

9. ОСНОВЫ НАНОТЕХНОЛОГИЙ ПОЛУЧЕНИЯ ОПТИЧЕСКИХ МАТЕРИАЛОВ

9.1. Исторические аспекты

Так называемая приставка нано- вообще означает одну миллиардную (10^{-9}) чего-либо (от греч. *nános* – карлик). *Нанотехнология* имеет дело с разнообразными структурами вещества, характерный размер которых – порядка миллиардных долей метра. Хотя слово нанотехнология является относительно новым, устройства и структуры нанометровых размеров не новы. На самом деле они существуют на Земле столько же, сколько существует

сама жизнь. Моллюск морское ушко выращивает очень прочную, переливающуюся изнутри раковину, склеивая прочные наночастички мела особой смесью белков с углеводами. Трещины, появляющиеся снаружи, не могут распространяться в раковине из-за наноструктурированных кирпичиков. Раковины являются природной демонстрацией того, что структуры, сформированные из наночастиц, могут быть намного прочнее материала, однородного в объеме.

В точности неизвестно, когда человек впервые начал использовать преимущества наноразмерных материалов. Есть сведения, что в четвертом веке нашей эры римские стекловары делали стекло, содержащее наночастицы металлов. Огромное разнообразие прекрасных цветов витражей в средневековых храмах объясняется присутствием металлических наночастиц в стекле. Так, например, прекрасный рубиновый цвет стекла получали введением наночастиц золота в стеклянную матрицу. Декоративная глазурь с глянцем, характерная для средневековой гончарной посуды, включает сферические металлические наночастицы, обеспечивающие специфические оптические свойства. Оказалось, что красящие вещества, использовавшиеся аборигенами Австралии для нанесения стойких боевых раскрасок, а также краска для волос древнегреческих красавиц, также содержали наночастицы, обеспечивающие очень длительный и стойкий окрашивающий эффект.

Потенциальную важность маленьких частичек – *кластеров* – осознал ирландский (по рождению) химик Роберт Бойль, и это нашло отражение в его труде «Химик-скептик» в 1661 году. В нем Бойль критикует воззрения Аристотеля на материю, состоящую из четырех первооснов: земли, огня, воды и воздуха. Вместо этого он предполагает, что крошечные частички вещества соединяются разными способами и образуют, таким образом, то, что он называл *корпускулами*. Он описывает их как «крошечные массы, или кластеры, которым тяжело быстро разложиться на составляющие их частицы».

Первым из ученых, использовавшим измерения в нанометрах, принято считать Альберта Эйнштейна, который в 1905 году теоретически доказал, что размер молекулы сахара равен одному нанометру.

Идею же создания специальных приборов, способных проникнуть в глубину материи до границ наномира, выдвинул выдающийся американский инженер-электрик и изобретатель, физик, философ сербского происхождения Н. Тесла. Именно он предсказал создание в будущем электронного микроскопа.

В 1920-х годах американский физик-теоретик российского происхождения Г. А. Гамов впервые вывел решение уравнений Шрёдингера, описывающее возможность преодоления частицей потенциального барьера при условии, если ее энергия меньше его высоты. Открытое явление, называемое «туннельным эффектом» (туннелированием), позволило объяснить многие экспериментально наблюдавшиеся процессы. Найденное решение было применено для описания процессов, происходящих при вылете частицы из ядра, составляющих в настоящее время основу атомной науки и техники.

В 1932 году нидерландский профессор Ф. Цернике открыл метод фазового контраста и создал первый фазово-контрастный микроскоп (Нобелевская премия, 1953 году) – вариант оптического микроскопа, улучшавший качество показа деталей изображения, и исследовал с его помощью живые клетки (ранее для этого приходилось применять красители, убивавшие живые ткани).

В 1939 году немецкие физики Э.А. Руска, получивший Нобелевскую премию в 1986 году, и М. Кноль создали электронный микроскоп, обеспечивающий увеличение на тот период большее, чем существовавшие оптические микроскопы. Прибор стал прообразом нового поколения подобных устройств, позволивших в дальнейшем заглянуть в мир нанообъектов. В этом же году компания «Сименс», в которой работал Руска, выпустила первый коммерческий электронный микроскоп с разрешающей способностью 10 нм.

В 1956 году сотрудник картографической службы военного ведомства США Дж. О'Кифи предложил конструкцию микроскопа, в котором свет должен был выходить из крошечного отверстия в непрозрачном экране и освещать объект, расположенный очень близко от экрана. Свет, прошедший через образец или отраженный от него обратно в отверстие, должен был регистрироваться в процессе возвратно-поступательного движения (сканирования) образца. О'Кифи назвал свой метод *растровой микроскопией* ближнего поля и указал, что разрешение такого микроскопа ограничивается не длиной волны света, а только размером отверстия. Теоретически такое устройство могло бы давать изображение деталей размером меньше половины длины волны.

Однако идея о том, что в будущем человечество сможет создавать устройства и другие объекты, собирая их «молекула за молекулой», а то и «атом за атомом», восходит к знаменитой лекции «Там внизу – много места» одного из крупнейших физиков XX века, лауреата Нобелевской премии, профессора Калифорнийского технологического института Ричарда Фейнмана, прочитанной 29 декабря 1959 года. Опубликованные в феврале 1960 года материалы лекции были восприняты большинством современников как фантастика или шутка. Сам же Фейнман говорил, что в будущем, научившись манипулировать отдельными атомами, человечество сможет синтезировать всё что угодно: «Ни один физический или химический закон не мешает нам менять взаимное положение атомов...», т.е. использовать атомы как обыкновенный строительный материал, что-то вроде кирпичей или, в лучшем случае, узлов и деталей машин.



Ричард Филипп Фейнман

Наиболее актуальной оставалась задача разработки и создания инструментального (метрологического) оборудования для изучения атомного строения конструкционных материалов на наноуровне.

В 1964 году, спустя шесть лет после изобретения интегральной схемы, Г.Э. Мур, почетный президент и один из основателей американской корпорации *Intel* (сокр. от *Integrated Electronics Technologies Incorporated*), выдвинул предположение, что число транзисторов на кристалле будет удваиваться каждые два года. Это наблюдение получило название первого закона Мура. Показав зависимость роста производительности запоминающих микросхем от сроков их изготовления, он обнаружил закономерность: новые модели микросхем каждый раз появлялись через приблизительно равные промежутки времени (18–24 месяца). При этом их ёмкость возрастала каждый раз примерно вдвое.

Американский физик Р. Янг, работавший в Национальном бюро стандартов, в 1966 году предложил пьезоэлектрическое управляющее устройство (пьезодвигатель), применяемое сегодня в сканирующих туннельных микроскопах и для позиционирования наноинструментов (поиска нужных объектов на изучаемой поверхности).

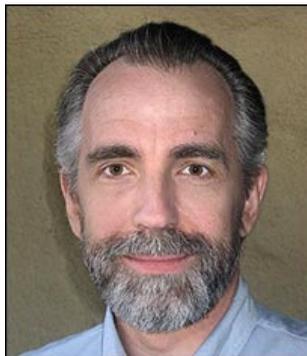
В это же время Д. Джонс теоретически конструировал замкнутые сфероидальные клетки из своеобразно свернутых нанографитовых слоев. Было показано, что объектом, внедренным в гексагональную решетку обычного графита и приводящим к образованию сложной искривленной поверхности, может быть пятиугольник, а физхимик-органик Е. Осава предположил существование полой, высокосимметричной молекулы C_{60} со структурой в виде усеченного икосаэдра, похожей на футбольный мяч (*фуллерена*).

Р. Янг в 1971 году предложил идею прибора «*Topografiner*», послужившего прообразом зондового микроскопа. В 1972 году Р. Янг сумел осуществить перемещение (и позиционирование) объектов в трех направлениях с точностью до 0,01 ангстрема \AA

(1 нм = 10 Å), применив перемещающие устройства на базе пьезоэлектриков. Со времени создания пьезодвигателя прошло более пяти лет.

В 1973 году российские ученые Д.А. Бочвар и Е.Г. Гальперин сделали первые теоретические квантово-химические расчеты наномолекулы фуллерена и доказали ее стабильность.

Современный вид идеи нанотехнологий начали приобретать в 1980-е годы в результате исследований К.Э. Дрекслера, работавшего в лаборатории искусственного интеллекта Массачусетского технологического института (США). Дрекслер выдвинул концепцию универсальных молекулярных роботов, работающих по заданной программе и собирающих любые объекты (в том числе и себе подобные) из подручных молекул. Всё это также сначала воспринималось как научная фантастика. Ученый уже тогда довольно точно предсказал немало грядущих достижений нанотехнологий, которые с 1989 года сбываются, причем часто со значительным опережением даже его прогнозов.



Основоположник молекулярной нанотехнологии Ким Эрик Дрекслер

Физик Н. Танигучи (*Taniguchi*) из Токийского университета в 1974 году впервые предложил термин *нанотехнология* – технология объектов, размеры которых порядка 10^{-9} м (атомы, молекулы), включающая процесс разделения, сборки и изменения материалов путем воздействия на них одним атомом или одной молекулой.

В 1975 году немецкие ученые-ботаники из Боннского университета (ФРГ) В. Бартхлотт и К. Найнюс (*Neinhuis*) обнаружили и запатентовали явление самоочистки поверхностей некото-

рых растений (*Lotus-effect*), а также тот факт, что этот феномен протекает в наноструктурированных поверхностных областях.

Весной 1981 года немецкие физики Г.К. Бинниг и Э. Руска, а также швейцарец Г. Рорер из Цюрихской лаборатории компании *IBM* (сокр. от англ. *International Business Machines Corporation*) испытали *растровый туннельный микроскоп* (Нобелевская премия, 1986 г.). Сканирующий туннельный микроскоп позволил построить трехмерную картину расположения атомов на поверхностях проводящих материалов. При движении острия иглы микроскопа над поверхностью кристалла, состоящего из кальция, иридия и олова, они смогли измерить неровности высотой в один атом. С помощью туннельного микроскопа стало возможным «захватить» атом с токопроводящей поверхности и поместить его в нужное место, т.е. манипулировать атомами, а следовательно, непосредственно собирать из них любой предмет, любое вещество.

Главной проблемой в исследовании были фоновые помехи: острие микроскопа, позиционировавшееся с точностью до долей атома, сбивалось от малейших шумов и вибраций даже вне лаборатории. Кроме того, сканирующий туннельный микроскоп позволяет исследовать нанообъекты только на электропроводной подложке.

Современные сканирующие микроскопы позволяют различать по вертикали над поверхностью размер около 0,01 нм (0,1 диаметра самого наименьшего атома – атома водорода), по горизонтали – около 0,2 нм. По сути, это уже не микроскопы, а наноскопы.

В 1985 году коллектив ученых в составе английского астрофизика, химика Г.В. Крото из Сассекского университета, американских химиков Р.Ф. Кёрла и С.К. О'Брайена и под руководством Р.Э. Смолли в Университете Райса (США) получили новый класс соединений – фуллерены и исследовали их свойства (Нобелевская премия, 1996 г.). В результате взрыва графитовой мишени лазерным пучком и исследования спектров паров гра-

фита была обнаружена молекула фуллерена C_{60} . Грани 60-атомного фуллерена – это 20 почти идеальных правильных шестиугольников и 12 пятиугольников. Позднее удалось получить фуллерены из 76, 78, 84, 90 и даже из нескольких сотен атомов углерода. Ученые также впервые сумели измерить объект размером 1 нм.

В 1986 году Г. Бинниг разработал сканирующий атомно-силовой микроскоп (АСМ), который позволял «увидеть» любые объекты, над которыми двигалась игла датчика. Такой микроскоп позволил, в отличие от туннельного микроскопа, осуществлять взаимодействие с любыми объектами, а не только с токопроводящими материалами.

К 1986 году в лабораториях мира работали уже не менее 40 сканирующих туннельных микроскопов (СТМ).

В настоящее время понятие нанотехнология включает в себя не только совокупность методов и способов синтеза, сборки, структурообразования и модифицирования материалов с наномасштабными свойствами, но и систему знаний, навыков, умений, аппаратурное, материаловедческое, информационное обеспечение процессов и технологических операций.

В научно-исследовательском институте «Дельта» под руководством П.Н. Лускиновича в 1987–1988 годах заработала первая российская нанотехнологическая установка, осуществлявшая направленный уход частиц с острия зонда микроскопа под влиянием нагрева.

Первый способ искусственного получения и выделения твердого кристаллического фуллерена был предложен в 1990 году В. Кречмером и Д. Хафманом с коллегами в Институте ядерной физики в г. Гейдельберге (Германия).

Углеродные нанотрубки открыл в 1991 году японский исследователь С. Ииджима (*Iijima*) из компании NEC (сокр. от англ. *Nippon Electric Corporation*). В поисках фуллеренов он изучал на полярном ионном микроскопе осадок (сажу), который образуется на катоде, когда распыляется графит при разряде

вольтовой дуги в атмосфере гелия. Его заинтересовал неприглядного вида серый «обрубок» диаметром 0,8 нм, вырастающий на катоде. Он оказался состоящим из странных графитовых наноцилиндров с угольно-черной сердцевиной (подобной карандашу). Электронная микроскопия осадка показала наличие протяженных полых объектов диаметром несколько десятков нанометров. Цилиндрические стенки цилиндров составляли сверхустойчивую структуру из шестигранных колец углерода, по краям закрытых полусферическими крышечками из семи- или восьмигранников. Так были открыты *нанотрубки и наноконусы*.

В 1992 году в природном углеродном минерале – шунгите (такое название минерал получил от названия поселка Шуньга в Карелии) были обнаружены природные фуллерены.

В дальнейшем различные наночастицы и наноструктуры обнаруживались во многих природных материалах, таких как лед, метеориты и даже на поверхностях обшивки орбитальных станций.

В 1994 году стали появляться первые коммерческие материалы на основе наночастиц – нанопорошки, нанопокрyтия, нанохимические препараты и т.д. С этого времени начинается бурное развитие *прикладной нанотехнологии*.

В 1997 году в Великобритании создан первый в Европе Институт наноструктурных материалов. Во многих институтах мира (США, Германии, Японии, Англии, Франции, Италии, Швейцарии, Израиля и др.) создаются лаборатории и отделы наноструктур, которые возглавляют известные ученые.

На 5-й Форсайтовской конференции Э. Дрекслер заявил, что, по его убеждению, к 2020 году станет возможной промышленная сборка наноустройств из отдельных атомов.

Профессор Высшей технической школы в г. Делфте (Нидерланды) С. Деккер создал транзистор на основе нанотрубок, используя их в качестве молекул. Для этого ему пришлось первым в мире измерить электрическую проводимость такой молекулы.

В Японии началась разработка программы *Astroboy* по развитию нанoeлектроники, способной работать в условиях

космического холода и при температуре в тысячи градусов выше нуля.

Еще через год (в 1999 году) американские ученые – физик М. Рид (Йельский университет) и химик Д. Тур (Райсский университет) разработали единые принципы манипуляции как одной молекулой, так и их цепочкой.

В 2000 году немецкий физик Ф. Гиссибл разглядел в кремнии субатомные частицы. Его коллега Роберт Магерле предложил технологию нанотомографии – создания трехмерной картины внутреннего строения вещества с разрешением 100 нм.

В 2003 году профессор Ф. Лью из университета Юты (США), используя разработки Гиссибла, с помощью атомного микроскопа построил образы орбит электронов путем анализа их возмущения при движении вокруг ядра.

В Японии действующая с 1999 года «Национальная программа работ по нанотехнологии» получила высший государственный приоритет «Огато». Проект спонсирует не только государство, но и дополнительно около 60 частных фирм. Кроме данного проекта, в Японии финансируется около десятка проектов, посвященных различным аспектам нанотехнологии – квантовым волнам, флуктуациям в квантовых системах и направленных на исследование и разработку квантовых функциональных схем. Крупнейшими являются проекты, связанные с атомной сборкой функциональных квантовых приборов и др. Основные разработки проводятся в центре перспективных технологий «Цукуба». По заявлениям руководителей этих проектов, они формируют технологию XXI века и планируют заложить основу для разработки терабитных (10^{12}) кристаллов.

Развитие исследований в области наноматериалов и нанотехнологий активно поддерживается правительством США. Так, еще администрацией Б. Клинтона была создана национальная программа исследований нанотехнологии (*National Nanotechnology Initiative*) с целью поддержки долгосрочных исследований и разработок, ведущих к значительным открытиям в облас-

ти новых наноматериалов, наноэлектроники, медицины и здравоохранения, энергетики, химической промышленности, биотехнологий, сельского хозяйства, информационных технологий и национальной безопасности.

Рассматривая наиболее весомые достижения российских ученых в области практической нанотехнологии, следует отметить награждение в 2000 году Нобелевской премией в области физики Ж.И. Алферова и его американских коллег Г. Кремера и Дж. Килби за создание полупроводниковых гетероструктур и интегральных схем.

В России фундаментальные научно-исследовательские работы по нанотехнологии проводятся по нескольким программам. К наиболее крупным из них относятся программа «Физика наноструктур» под руководством академика РАН Ж.И. Алферова и «Перспективные технологии и устройства в микро- и наноэлектронике» под руководством академика К.А. Валиева.

В Физико-техническом институте им. А.Ф. Иоффе под руководством Ж.И. Алферова осуществляются передовые разработки наногетероструктур, получившие международное признание. Ежегодно проводится международная конференция «Наноструктуры: физика и технологии». Значительные результаты нанотехнологических исследований достигнуты в Институте проблем технологии и макроэлектроники РАН под руководством члена-корреспондента РАН В.В. Аристова, а также в Физическом институте имени П.Н. Лебедева РАН под руководством члена-корреспондента РАН Ю.В. Копаева.

В области прикладных нанотехнологических исследований также можно отметить работы, проводимые корпорацией МДТ (*Molecular Device Tools for Nanotechnology*), созданной в 1991 году в г. Зеленограде выпускниками Московского физико-технического института (МФТИ).

В России инвестируются значительные средства для развития и интенсификации научно-исследовательских работ в области наноматериалов и нанотехнологий.

9.2. Наноматериалы

Повышенный интерес к методам получения сверхмелкозернистых объёмных и дисперсных материалов объясняется тем, что уменьшение размера структурных элементов (частиц, кристаллитов, зёрен) ниже некоторой пороговой величины приводит к заметному изменению свойств. Такие эффекты появляются, когда средний размер кристаллических зёрен не превышает 100 нм, и наиболее отчётливо наблюдаются, когда размер зёрен менее 10 нм. Изучение свойств сверхмелкозернистых материалов требует учёта не только их состава и структуры, но и дисперсности. Поликристаллические сверхмелкозернистые материалы со средним размером зёрен от 300 до 40 нм называют обычно *субмикрорекристаллическими*, а со средним размером зёрен менее 40 нм – *нанокристаллическими*. Условная классификация материалов по размеру D -частиц (зёрен) показана на рис. 9.1.

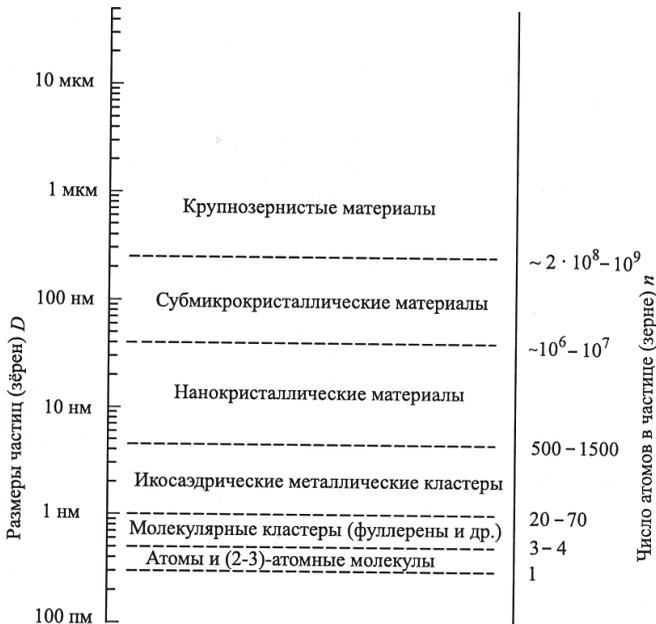


Рис. 9.1. Классификация веществ и материалов по размеру D частиц (зёрен)

Наноматериалы можно классифицировать также по геометрической форме и размерности структурных элементов, из которых они состоят. Основными типами нанокристаллических материалов по размерности являются кластерные материалы, волоконные материалы, плёнки и многослойные материалы, а также поликристаллические материалы, зёрна которых имеют сравнимые размеры во всех трёх направлениях (рис. 9.2).

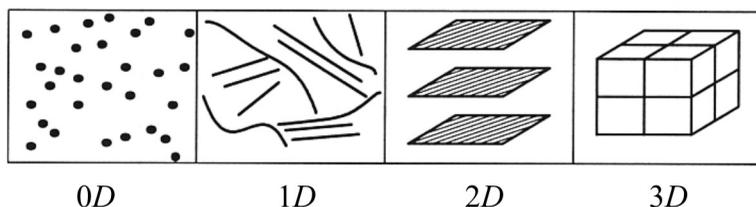


Рис. 9.2. Типы нанокристаллических материалов:
 0D – (нульмерные) кластеры; 1D – (одномерные) нанотрубки, волокна и прутки; 2D – (двумерные) плёнки и слои;
 3D – (трёхмерные) поликристаллы

Коллоидные кластеры

Коллоидные кластеры образуются в растворах в результате химических реакций и могут иметь размеры от 1 до 100 нм. Они могут долго существовать в жидкой фазе, не осаждаясь и не коагулируя, благодаря слабым межкластерным взаимодействиям, зарядовому отталкиванию и пассивации поверхности. По отношению к жидкой фазе коллоидные кластеры могут быть разделены на две группы: *гидрофильные* («любящие» воду) и *гидрофобные*.

Гидрофильные кластеры могут сорбировать на своей поверхности молекулы окружающей среды и образовывать с ними прочные комплексы. Кластеры этого типа окружены оболочкой, которая частично сохраняется и при коагуляции отдельных кластеров, и при переходе их в гелевую наносистему. Наиболее типичными представителями гидрофильных кластеров являются оксиды кремния, оксиды железа и других металлов.

Гидрофобные кластеры не адсорбируют на своей поверхности молекулы растворителя. Однако их поверхность можно модифицировать ионами из раствора, при этом она приобретает положительный или отрицательный заряд.

Для предотвращения слипания поверхность коллоидных кластеров пассивируют действием поверхностно активных веществ (ПАВ).

При растворении ПАВ в воде происходит структурирование молекул воды вокруг неполярных углеводородных радикалов. При этом молекулы ПАВ начинают самопроизвольно образовывать ассоциаты, которые называются *мицеллами* (предложение Мак-Бэна, 1913 г.).

Таким образом, мицеллы – это объединения ПАВ, обладающих полярной *гидрофильной* «головой» и неполярным *гидрофобным* углеводородным «хвостом». Если такое вещество растворить в любой жидкой среде, молекулы ПАВ будут заполнять ее поверхность до тех пор, пока их концентрация не достигнет некоторого предельного значения, называемого критической концентрацией мицеллообразования. При этом молекулы ПАВ будут обращаться к жидкой фазе тем концом, который ближе по химической природе к молекулам растворителя. Выше же критической концентрации мицеллообразования молекулы начнут «сжеживаться» (подобно поведению ежа в моменты опасности), образуя замкнутые ассоциаты, гидрофильная или гидрофобная часть которых полностью замкнута в объеме мицеллы (рис. 9.3). Если растворить ПАВ в полярной среде (например, в воде), то наружу будут обращены полярные части молекул, а мицеллы будут называться «прямыми». Если же молекулы ПАВ поместить не в воду, а, например, в неполярное масло, то они, словно ежик, свернутся в клубки, выставив, как иголки, наружу свои гидрофобные хвосты. Такие мицеллы носят название «обращенных». Движущей силой такого поведения молекул является уменьшение межфазного натяжения на границе мицелла-растворитель.

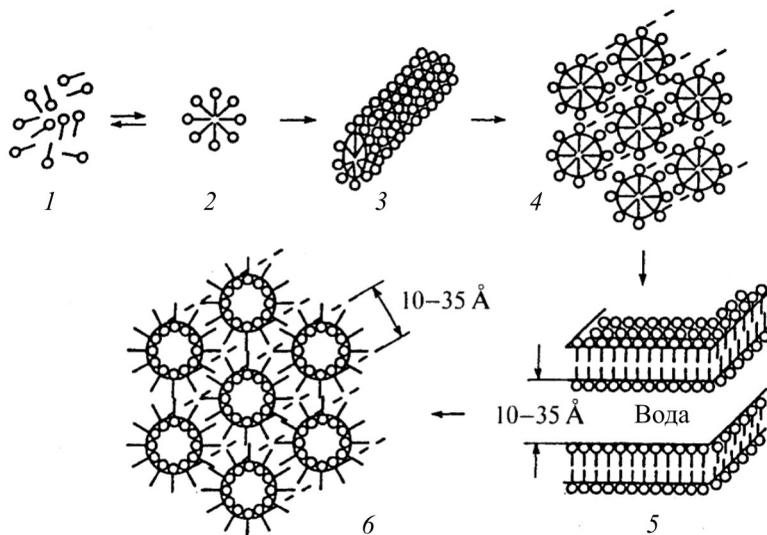


Рис. 9.3. Наноструктуры, возникающие в растворах с участием ПАВ:
 1 – мономеры; 2 – прямая мицелла; 3 – цилиндрическая мицелла;
 4 – гексагонально упакованные цилиндрические мицеллы;
 5 – ламелярная мицелла; 6 – гексагонально упакованные
 обратные мицеллы

При увеличении концентрации ПАВ мицеллы могут деформироваться и приобретать несферическую форму. Так, можно получить несферические типы мицелл: цилиндрические, гексагонально упакованные и другие. Если взять растворитель сложного состава, смешав компоненты согласно диаграмме состояния «полярный растворитель – неполярный растворитель – ПАВ» (рис. 9.4), то можно получить микрогетерогенные системы, полярная и неполярная фазы которых будут пространственно разделены мономолекулярной пленкой ПАВ в объеме кажущейся однородной жидкости.

Мицеллярные системы активно используют в процессах синтеза *наноструктур* и *наноматериалов*. Так, синтез в обращенных мицеллах является на сегодня самым распространенным

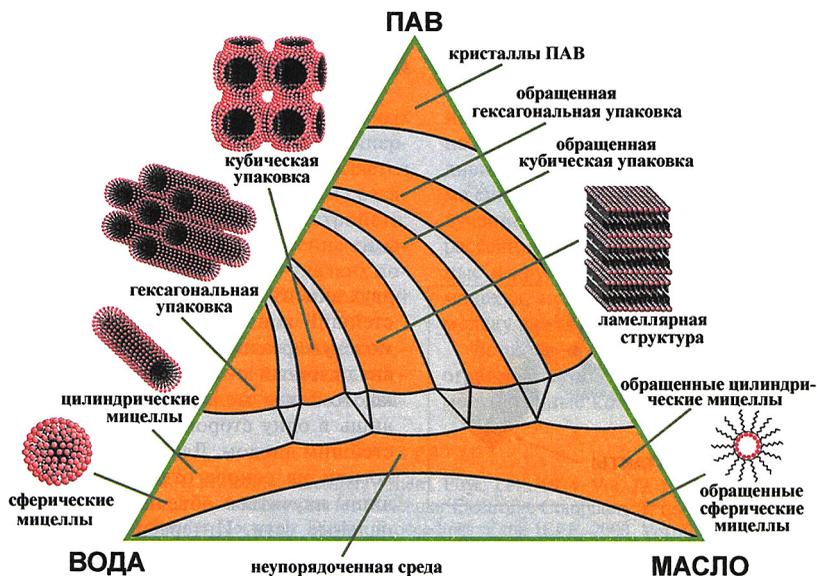


Рис. 9.4. Диаграмма состояния системы «вода – масло – ПАВ»

способом формирования однородных по размеру наночастиц, а прямые мицеллы применяют для синтеза *цеолитов*, *мезопористых молекулярных сит* и нанопористых соединений. В природе мицеллоподобные структуры образуются в крови, в межтканевой жидкости.

Фуллерены

Долгие годы считалось, что углерод может образовывать только две кристаллические модификации – *графит* и *алмаз*. Наиболее исследованным соединением углерода был графит. Графит состоит из плоских слоев углерода. Связи атомов в каждой плоскости чрезвычайно прочны, но отдельные плоскости связаны друг с другом не очень сильно и могут изгибаться и ломаться. Поэтому в карандашах с мягкими графитовыми стержнями слои графита при трении о бумагу отслаиваются и остаются на бумаге.

Графит часто используется в качестве смазки в некоторых механизмах с трущимися частями. Сильные связи между отдельными атомами графита объясняют высокую температуру его плавления. Графит не растворяется в воде и органических растворителях, но способен проводить электрический ток.

Значительно более прочным соединением атомов углерода является алмаз. Благодаря сильным ковалентным связям каждого атома углерода с тремя соседними атомами они образуют прочную объемную кристаллическую решетку. Алмаз настолько прочен и тверд, что используется для резки, сверления и полировки других материалов в промышленности.

Алмаз имеет очень высокую температуру плавления (около 4000 °С), ведь для разрушения кристаллической структуры нужно разорвать все сверхсильные ковалентные связи между атомами. Алмаз, как и графит, не растворяется в воде и других органических растворителях. Однако алмаз не проводит электрический ток. Дело в том, что все электроны в нем прочно удерживаются между атомами и не могут перемещаться по решетке.

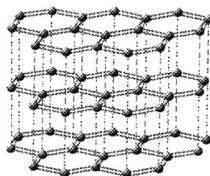
Люди издавна ценили алмазы за их необыкновенное сияние и блеск. Однако ученых они привлекали своей чрезвычайной твердостью и малой сжимаемостью. Алмазы прекрасно проводят тепло и очень слабо расширяются с увеличением температуры. Они не вступают в реакцию с большинством сильных кислот или оснований. Алмазы прозрачны в очень широком диапазоне: не только для видимого света, но и для ультрафиолетового и инфракрасного излучения.

Третью модификацию – *карбин*, *фуллерен* – открыли в 60-х годах XX в. Карбин представляет собой линейные кристаллы, в которых атомы углерода связаны двойными или чередующимися одинарными и тройными связями. В 70-х годах при попытках астрофизиков объяснить природу спектров излучения межзвездной пыли Г. Крото предположил, что их источником являются макромолекулы углерода C_{60} размером всего 1 нм, имеющие форму усеченного икосаэдра (рис. 9.5). Такую форму имеет футболь-

ный мяч. Термин «фуллерен» происходит от имени Ричарда Букминстера Фуллера, сконструировавшего оригинальный купол павильона США на выставке EXPO-67 в Монреале в форме сочлененных пентагонов и гексагонов. В 1985 г. группой английских химиков, возглавляемой Г. Крото, молекулы C_{60} были зарегистрированы с помощью масс-спектрографа в саже (Нобелевская премия по химии, 1996 г.).

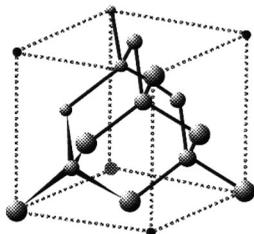


Уголь (объем)



Графит (плоскость)

Алмаз (кристалл)



Фуллерен (сфера)

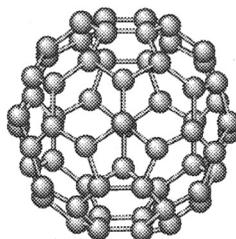


Рис. 9.5. В угле и графите атомы углерода расположены в плоскостях, в алмазе – в объемной кристаллической решетке, а в фуллерене – в сферическом каркасе

С химической точки зрения фуллерен представляет собой аллотропную модификацию углерода, уникальную своей молекулярной структурой – практически идеальные шарики удерживаются вместе только слабыми вандерваальсовыми силами. Очень необычна история открытия фуллеренов, которые сначала были предсказаны теоретически и лишь через 20 лет обнаружены экспериментально, вызвав настоящий «фуллереновый

бум» в научном мире. Уже сейчас опубликованы тысячи научных статей, книг, брошюр, посвященных данной теме, обсуждается возможность самых необычных применений этого соединения в наноэлектронике, медицине, технике (например, в качестве основы для синтеза искусственных алмазов, элементов квантовых компьютеров, для создания новых электронных приборов, а также в составе лекарств и смазочных веществ).

Следует отметить, что C_{60} – это далеко не единственный представитель фуллеренов. В настоящее время под фуллеренами понимаются углеродные молекулярные кластеры с четным, более 20, количеством атомов углерода, образующих три связи друг с другом.

Простейший его представитель C_{20} до сих пор экспериментально не получен – при синтезе преимущественно образуются молекулы C_{60} , а также высшие фуллерены с четным числом атомов, вероятно, из-за стабильности таких каркасов. Причина отсутствия малых фуллеренов связана, по-видимому, с тем, что с уменьшением размера структура становится более напряженной и склонной к полимеризации. Сам по себе фуллерен плохо растворим в воде и не проводит электрический ток, однако модификация поверхности фуллеренового «шара» или заполнение его внутреннего пространства атомами металлов приводит к заметному изменению физических свойств, например переходу в сверхпроводящее состояние или проявлению магнетизма. Такие молекулы фуллерена называются *фуллеритами*.

Нанотрубки

Из атомов углерода могут образовываться также различные структуры, составленные, как и в случае графита, из шестиугольных колец. Эти структуры являются полыми внутри и имеют замкнутую поверхность. Среди них выделяются *наночастицы* и *нанотрубки*. Наночастицы подобны фуллеренам, но значительно превосходят их по размерам. Они могут состоять

из нескольких слоев, образуя «луковичные структуры» или *онионы* (от англ. *onion* – луковица). Нанотрубки представляют собой графитовые сетки, свернутые в трубки, и могут быть как открытыми, так и закрытыми с концов (рис. 9.6), однослойными и многослойными.

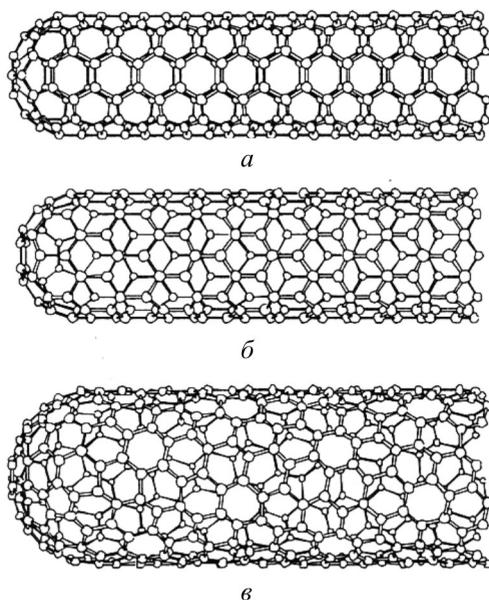


Рис. 9.6. Примеры некоторых структур углеродных нанотрубок: *а* – кресельная, *б* – зигзагообразная, *в* – хиральная

Однослойные углеродные нанотрубки (*carbon nanotubes*) открыли почти одновременно в 1991 году Сумио Ииджима и Тошинари Ичихаши (*Toshinari Ichi-hashii*) в Японии, а также Дональд С. Бетун (*Donald S. Bethune*) в США. Обе группы ученых описали образование фуллерена C_{60} в саже на стенках камеры дугового разряда и обнаружили, что сажа состоит из множества однослойных углеродных нанотрубок приблизительно с одинаковым диаметром.

Нанотрубки встречаются в природном материале – *шунгите*, – однако в настоящее время они получаются искусственно.

В зависимости от условий синтеза возможно получить как *одностенные углеродные нанотрубки*, похожие на очень тонкий и длинный цилиндр, «склеенный» из гексагональной углеродной сетки, так и многостенные нанотрубки. Последние представляют собой либо сверток, аналогичный бумажному свитку, либо набор одностенных труб, вставленных друг в друга наподобие матрешки (рис. 9.7). В зависимости от того, как именно из графитовой плоскости «вырезать полоску» и как ее свернуть, диаметр трубки может варьироваться в широких пределах от 0,4 до 100 нм, а длина – от 1 до 100 мкм.

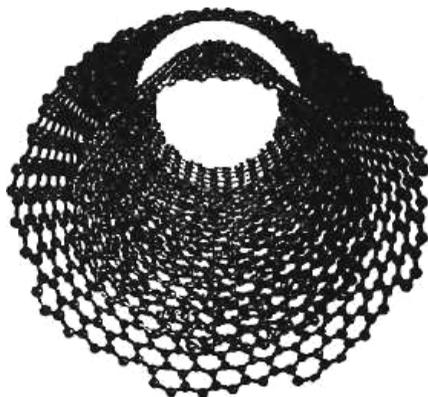


Рис. 9.7. Схема вложенных нанотрубок, когда одна трубка находится внутри другой

Углеродные нанотрубки обладают разнообразными свойствами, демонстрируя, несмотря на свою кажущуюся хрупкость и ажурность, высокую прочность на растяжение и изгиб, способность перестраиваться под действием механических напряжений, высокую проводимость, необычные магнитные и оптические свойства (табл. 9.1).

Поражает воображение разнообразие применений нанотрубок, одни из которых только придуманы «на бумаге», а другие –

Таблица 9.1

Свойства	Характеристики нанотрубок	Возможные применения
Механические	В 100 раз прочнее и в 6 раз легче стали	Сверхпрочные нити, создание композитных материалов, космический лифт
	Высокая гибкость	Наноактюаторы
	Высокая эластичность	Острия для атомно-силовой микроскопии
Электронные	Вариация электронных свойств в зависимости от диаметра трубок (переходы металл–полупроводник)	Компоненты нанoeлектроники (диоды, транзисторы и т.п.)
	Высокая проводимость металлических трубок	Прозрачные проводящие пластики, острия зондовых микроскопов
	Высокая полевая эмиссия	Электронная пушка, дисплеи, лампы, рентгеновские трубки
Физико-химические	Высокая удельная поверхность (100–1000 м ² /г)	Носители для катализаторов
	Высокая пористость и объем, форма каналов	Капсулы для активных молекул, защита инкапсулированных материалов, хранение водорода, металлов и газов, нанобатарей, нано pipетки

реализованы в действительности. Нанотрубки могут использоваться в качестве очень прочных микроскопических стержней и нитей: согласно результатам численного моделирования модуль Юнга однослойной нанотрубки достигает величин порядка 10^3 ГПа, что на порядок выше, чем у стали. Теоретики подсчитали, что из нанотрубок можно создать самые прочные волокна в мире, которые почти в 100 раз прочнее и в 6 раз легче стали. Недаром реализация одного из грандиозных проектов будущего, строительства *космического лифта*, так называется новая технология доставки грузов и пассажиров с поверхности Земли на космическую орбиту, основывается на уникальной механической прочности трубок. К настоящему моменту экспериментально подтверждено, что «микротрос» из нанотрубок

толщиной с человеческий волос способен удерживать грузы в сотни килограмм!

Идея космического лифта не нова. В 1895 году после посещения Эйфелевой башни русский ученый Константин Циолковский впервые высказал идею конструкции «космической башни», состоящей из катушки с кабелем, растянутым до высоты *геостационарной орбиты*. Эта идея пришла ему в голову после наблюдения за обычными лифтами, которые доставляли изумленных зрителей на вершину Эйфелевой башни. На рис. 9.8 показана схема такого лифта.



Рис. 9.8. Космический лифт для доставки грузов на геостационарную орбиту

Зачем нужен космический лифт? Для улучшения качества жизни и удешевления космических полетов. В настоящее время стоимость доставки 1 кг полезного груза на низкую околоземную орбиту равна около 20 тыс. долл., а на геостационарную орбиту – около 40 тыс. долл. При использовании космического лифта цена доставки 1 кг полезного груза на геостационарную орбиту будет не выше 400 долларов. Вследствие удешевления вывода спутников подешевеют и средства связи. Благодаря массовому выводу огромных солнечных батарей на околоземную орбиту подешевеет производство альтернативной энергии. В условиях невесомости упростится и станет массовым производство сложных фармацевтических препаратов и рост высокочистых кристаллов.

Но поистине удивительная особенность таких структур заключается в зависимости свойств от структуры самой нанотрубки. Так, например, вектор свертки графитовой плоскости определяет электронную структуру нанотрубки: будет ли она металлом или полупроводником.

Необычные электрические свойства нанотрубок делают их наиболее перспективным материалом нанoeлектроники: уже сейчас на их основе созданы опытные образцы полевых *транзисторов* и полупроводниковых *гетероструктур*, в которых переход металл/полупроводник реализуется в пределах единичной нанотрубки путем контролируемого создания дефектов в ее структуре.

Безусловно, компьютерная индустрия не обошла своим вниманием углеродные нанотрубки, создав и успешно протестировав прототипы первых плоских дисплеев, работающих на нанотрубчатой матрице. В таком дисплее под действием прикладываемого к одному из концов трубки напряжения с другого ее конца испускаются электроны, которые, попадая на фосфоресцирующий экран, вызывают свечение пикселя. В результате зерно изображения фактически определяется диаметром нанотрубки, а следовательно, обладает фантастически малыми размерами – не более одного микрона.

Однако нанотрубки могут выступать не только в роли исследуемого материала, но и как инструмент исследования. На их основе можно, к примеру, создать микроскопические *нановесы*, *нанотермометры* и другие *наноинструменты*, а если трубку «насадить» на острие *сканирующего туннельного* или *атомного силового микроскопа*, то получится идеальный микрозонд диаметром всего в нескольких атомов. Прикладывая определенное напряжение, с помощью такой иглы можно захватывать атомы или целые молекулы и как пинцетом перемещать их с места на место, а можно даже с высокой точностью производить запись и считывание информации с матрицы вплоть до плотности записи 250 Гбит/см².

Нанотрубки, в зависимости от длины, содержат от тысячи до миллионов атомов углерода. Открытые нанотрубки ведут себя как капилляры, втягивая в себя растворы или расплавы вещества, атомы которого меньше внутреннего диаметра трубки. Так могут создаваться *нанопроволоки*. При этом свойства металла внутри нанотрубки могут сильно отличаться от свойств этого металла в свободном состоянии, так как взаимодействие с углеродными стенками приводит к изменению его физических свойств. Благодаря этой способности есть возможность использовать фуллерены и нанотрубки в качестве нового наномасштабного материала в биологических системах и композитах. Углеродные нанотрубки и фуллерены являются наиболее удивительным открытием в области материаловедения за последние несколько десятилетий.

Наноленки

Тонкие наноструктурированные пленки представляют собой организованные наносистемы, в которых наноразмер проявляется только в одном измерении, а два других могут обладать макроразмерами. Известно несколько методов получения наноленок.

Метод эпитаксии (от греч. *epí* – на, над, при и *táxis* – расположение, порядок) заключается в наращивании монокристаллических слоев вещества на подложку (кристалл), при котором

кристаллографическая ориентация наращиваемого слоя повторяет кристаллографическую ориентацию подложки.

Метод химического парофазного осаждения веществ состоит в том, что исходное вещество испаряется в отдельной камере, а затем переносится через газовую фазу и осаждается в нужной пропорции на выбранную подложку.

Метод осаждения и выпаривания коллоидных растворов не позволяет получать достаточно прочные нанопленки.

Эффективным методом получения нанопленок служит *технология Ленгмюра–Блоджетт*. Этот метод был предложен лауреатом Нобелевской премии (1932) Ирвингом Ленгмюром еще в 1920 году, развит его коллегой Катариной Блоджетт в 1935 году. В соответствии с этим методом на поверхности воды формируется монослой ПАВ, в который могут входить ионы металлов и их комплексы. Затем с помощью ванн Ленгмюра–Блоджетт пленки с поверхности жидкой фазы переносят на твердую поверхность. В результате получают организованные нанопленки с регулируемым числом молекулярных слоев.

Познакомиться с технологией получения пленок Ленгмюра – Блоджетт можно при помощи обыкновенного стакана с водопроводной водой: капните в стакан капельку подсолнечного масла и, подождав пока капля растечется по поверхности, опустите в получившийся «блин» палец. На пальце образуется тонкая пленка подсолнечного масла, в чем можно легко убедиться, прикоснувшись к ней и испачкавшись окончательно. Такое же явление наблюдается и в более крупных масштабах при растекании маслянистых жидкостей по поверхности водоемов.

Термин «пленки Ленгмюра – Блоджетт» (*Langmuir – Blodgett films*) обозначает моно- или многослойные пленки, перенесенные с границы раздела жидкость/воздух на твердую подложку. Процесс формирования таких пленок подчиняется закономерностям самосборки. Молекулярная пленка на границе раздела жидкость/воздух называется *лэнгмюровской пленкой* (*Langmuir film*). Метод физического осаждения LB-пленок при погружении

в жидкость, на поверхности которой находится органическая пленка (или при подъеме из этой жидкости), называется LB-осаждением. В качестве жидкой среды чаще всего используется деионизированная вода, но могут применяться и другие жидкости (например, глицерин и ртуть). Все органические примеси должны быть удалены с поверхности воды путем фильтрации через фильтр из активированного угля.

Напомним, что вещества, мономолекулярные слои которых при переносе LB-методом взаимодействуют с водой (растворяются в ней), смачиваются или набухают, называются *гидрофильными (hydrophilic)*. Вещества, которые не взаимодействуют с водой (не растворяются в ней), не смачиваются и не набухают, называются *гидрофобными (hydrophobic)*. Существуют и так называемые *амфифильные (amphiphile)* вещества, которые растворяются и в воде, и в жирах. Один конец молекулы таких веществ – гидрофильный (и поэтому предпочитает быть погруженным в воду), тогда как другой ее конец – гидрофобный (и поэтому обычно находится в воздухе или в неполярном растворителе). Классический пример амфифильного вещества – стеариновая кислота, в которой длинный гидрокарбонатный «хвост» является гидрофобным, а основная (головная) карбоксильная группа – гидрофильной. Поскольку у амфифильных один конец – гидрофильный, а другой – гидрофобный, то они предпочитают располагаться на таких границах раздела, как воздух/вода или масло/вода. По этой причине их называют *поверхностно-активными (surfactants)*.

Уникальной особенностью LB-пленок является возможность формирования упорядоченной структуры на твердой поверхности из некристаллического материала. Это позволяет переносить мономолекулярные слои на различные подложки. В зависимости от способа обработки подложки ее поверхности можно придать гидрофильные или гидрофобные свойства.

Один из вариантов процесса получения пленок (вертикальный лифт) показан на рис. 9.9. Мономолекулярный слой амфифильного вещества может быть осажден с границы раздела во-

да/воздух путем вертикального смещения пластины. Когда подложка движется через молекулярный слой на границе раздела вода/воздух, этот слой может быть перенесен на нее в процессе извлечения данной подложки (подъема вверх) или ее погружения (опускания вниз). В зависимости от направления движения подложки сквозь монослой получается пленка с различной молекулярной ориентацией. При первом погружении подложки (рис. 9.9, *a*) на ее гидрофильной поверхности формируется монослой с ориентированными к подложке гидрофильными хвостами ПАВ. При втором погружении подложки на ней формируется мультислой с противоположной ориентацией. Последовательным переносом монослоев можно приготовить многослойную наноразмерную пленку из мономолекулярных (по толщине) слоев.

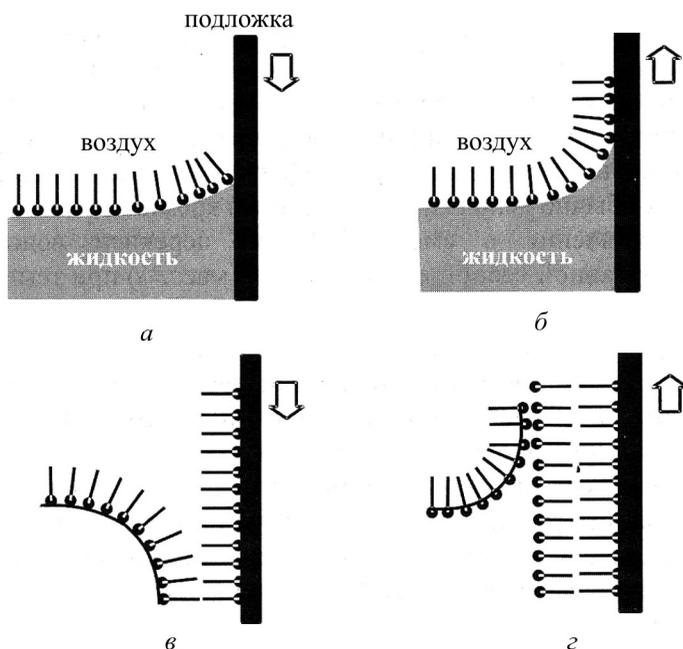


Рис. 9.9. Схема формирования многослойных пленок методом Лэнгмюра–Блоджетт: *a* – первое погружение; *б* – первый подъем; *в* – второе погружение; *г* – второй подъем

Процесс применения структурированной органической матрицы для синтеза и выращивания неорганических кристаллов получил название *темплатного синтеза*. Материалы – органонеорганические нанокомпозиты, полученные таким способом, именуют *гибридными наноматериалами*.

Техника Ленгмюра–Блоджетт внешне проста и дешева (не нужен сверхвысокий вакуум, высокие температуры и т.п.), однако первоначально требует значительных затрат для создания особо чистых помещений, так как любая пылинка, осевшая даже на одном из монослоев в гетероструктуре, это незалечиваемый дефект.

Тем не менее ленгмюровские пленки и нанокомпозиты на их основе уже нашли применение в качестве длинноволновых рентгеновских дифракционных решеток, резисторов, газовых сенсоров, наноразмерных диэлектрических полимерных покрытий и прослоек в различных устройствах и т.д.

Нанокристаллы

Периодические структуры с выраженной интерференционной окраской часто встречаются в живой природе. Структурами с одномерной периодичностью являются, например, покрытия на

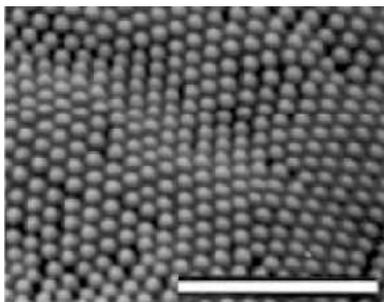


Рис. 9.10. Наружная поверхность глаза бабочки *Vanessa kershawi*. Длина метки равна 2 мкм

крыльях некоторых бабочек, хвостовых перьях павлина, панцирях некоторых жуков. Роль интерференции в окраске перьев павлинов отмечал еще И. Ньютон в 1730 году. Структуры с двумерной периодичностью присутствуют в строении глаз насекомых. Регулярная пористая структура глаз насекомых (рис. 9.10) является эффективным антиотражающим интерфейсом, обес-

печивающим прохождение света с одновременной возможностью физико-химического обмена с окружающей средой для внутренних тканей глаза.

В природе существуют трехмерные периодические структуры в виде коллоидных кристаллов. Они впервые были обнаружены при исследовании вирусов. Полудрагоценный минерал опал представляет собой коллоидный кристалл, состоящий из монодисперсных сферических глобул оксида кремния. Именно интерференцией света в трехмерной периодической структуре определяется их искрящийся цвет, зависящий от угла падения и наблюдения.

Упорядоченные структуры, период которых сопоставим с длинами волн света видимого или ближнего инфракрасного диапазонов, получили особое название – *фотонные кристаллы* (рис. 9.11). Это название указывает лишь на то, что поведение фотонов в таких средах весьма необычно.

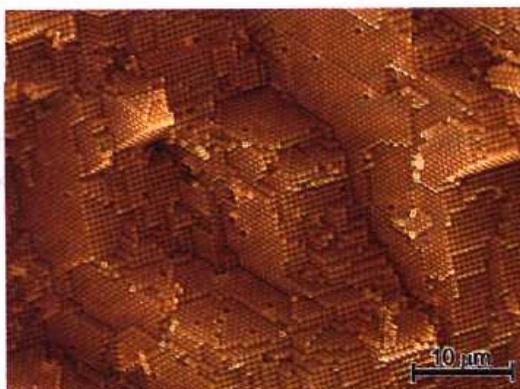


Рис. 9.11. Фотонный кристалл, состоящий из упорядоченных сферических частиц размером около 650 нм

Фотонный кристалл – это оптическая среда, в которой происходит периодическое изменение коэффициента преломления на масштабе, сопоставимом с длиной волны света видимого

или ближнего инфракрасного диапазонов. При этом различают одно-, двух- и трехмерные фотонные кристаллы в зависимости от того, наблюдается ли указанная периодичность в одном, двух или трех измерениях соответственно.

В качестве одномерного фотонного кристалла вполне можно рассматривать дифракционную решетку. Хорошо известно, что одномерная периодичность дифракционной решетки позволяет ей эффективно отражать свет некоторой длины волны, падающий на решетку под определенным углом. В фотонных кристаллах с трехмерной периодической структурой возможна ситуация, при которой свет некоторой длины волны отражается от такого объекта при любом угле падения. В эту оптическую среду фотон не может попасть извне.

Уникальное поведение фотонов в фотонных кристаллах позволит создать на основе последних самые разнообразные оптические устройства, которые произведут настоящую революцию в оптоэлектронике. Уже сейчас предложены пути использования фотонных кристаллов для создания высокоэффективных светодиодов и лазеров, новых оптоволоконных волноводов, оптических переключателей и фильтров, лежащих в основе сверхбыстрых фотонных компьютеров, в которых все логические операции производятся не электронами, а фотонами.

Известно, что передавать световые сигналы по воздуху, несмотря на его «прозрачность», неэффективно – в атмосфере содержится огромное количество пыли и капель воды, которые сильно рассеивают свет. Поэтому для передачи фотонов было придумано оптическое волокно, состоящее из сверхчистого стекла. Наноструктурированные волноводы по многим характеристикам значительно превосходят обычное стеклянное оптоволокно. Впервые оптическое волокно, созданное по принципу фотонного кристалла, было получено в 1995 году сотрудниками фирмы *Blaze Photonics* (Англия).

В технологии оптического волокна на принципе фотонного кристалла в кварцевую трубку диаметром 20 мм упаковыва-

ются капилляры диаметром порядка 1 мм с направляющим стержнем из того же кварцевого стекла, расположенным между капиллярами. Из полученной заготовки производится вытяжка волокна, при этом характеристические поперечные размеры изделия уменьшаются примерно в 10 000 раз.

Диаметр полых микроканалов, проходящих по всей длине волокна в 100 м, составлял 0,3 мкм при их периодическом зазоре в 2–3 мкм. Поперечное сечение такого волокна имело гексагональную форму, определяемую структурой укладки капилляров в трубке (рис. 9.12).

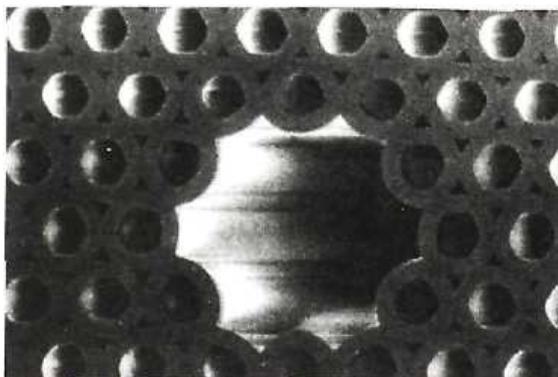


Рис. 9.12. Поперечное сечение увеличенного фрагмента оптического волокна (фотонный кристалл с полый сердцевиной)

Наличие таких полых микроканалов позволяет более чем на порядок увеличить относительную разность коэффициентов преломления сердцевины и оболочки по сравнению со стандартным волокном. Дополнительная «степень свободы», предоставляемая оптическими волокнами, созданными по принципу фотонного кристалла, делают их перспективными средами для систем оптической передачи информации.

Наноповерхности

В середине 1970-х годов профессора ботаники Боннского университета (ФРГ) В. Бартхлотт и К. Найньюс (*Neinhuis*) обнаружили, что листья и цветки некоторых растений почти не загрязняются и что этот феномен протекает в их наноструктурированных поверхностных областях. Впоследствии это явление ими было запатентовано и названо в честь наиболее яркого представителя таких растений «*лотос-эффектом*» (*Lotus-effect*).

Издrevле цветок лотоса считается символом незапятнанной чистоты. Как известно, листья и нежно-розовые цветки лотоса распускаются из грязной тины водоемов безупречно чистыми. Этот феномен *самоочистки* детально исследовался и позволил открыть удивительные возможности природы защищаться не только от грязи, но и от различных микроорганизмов. Данный эффект наблюдается не только у лотоса, но и у многих других растений, таких как капуста, камыш, водосбор, тюльпан, а также у животных (крылья стрекоз и бабочек). Они наделены природой свойством защиты от различных загрязнений по большей части неорганического происхождения (пыль, сажа), а также биологического происхождения (споры грибов, микробов, водоросли и т.д.).

Использование электронных микроскопов позволило ученым обнаружить, что поверхности листьев, цветков и побегов покрыты тонкой внеклеточной мембраной – поверхностным слоем. Эпидермис листьев и цветков некоторых растений выделяет воскоподобное вещество *кутин*, представляющее собой смесь высших жирных кислот и их эфиров. Жиры и жироподобные вещества, входящие в состав липидов – природных органических соединений, являются одним из основных компонентов биологических мембран. Липиды участвуют в обмене между растениями и окружающей средой (рис. 9.13).

Попавшая на поверхность листа капля воды удаляет с него частицу загрязнений. При этом частицы загрязнений не прони-

кают во внутреннюю часть капли, а равномерно распределяются по ее поверхности. Замечено, что гидрофобная субстанция удаляется каплей воды с гидрофобной поверхности. При рассмотрении условий протекания «эффекта лотоса» на наноскопическом уровне механизм этого явления становится более понятным.

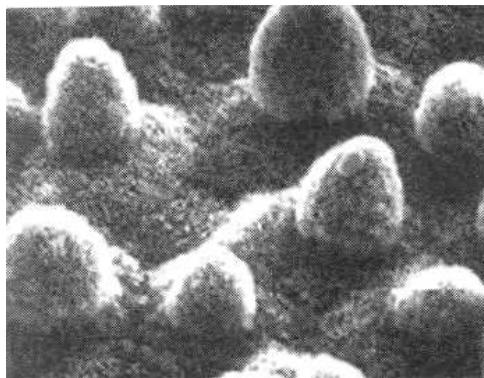


Рис. 9.13. Поверхность листа лотоса под электронным микроскопом

Можно представить себе массажную щетку, на зубьях которой лежит клочок бумаги, изображающий частицу загрязнений. Пятно «грязи» расположено только на самых вершинах зубьев, не соприкасаясь с поверхностью щетки (рис. 9.14, справа). Сила прилипания грязи обусловлена площадью поверхности взаимного контакта. Если бы поверхность была гладкой или имела макрорельеф, как на рис. 9.14 слева, то площадь контакта оказалась бы значительной и грязь удерживалась бы достаточно прочно. Однако из-за острых концов зубьев площадь контакта минимальна, и грязь как бы «висит на ножке». То же происходит и с каплей воды. Она не может «растечься» по остриям, и поэтому стремится свернуться в шарик.

Аналогичное явление происходит и с грязью на восковых кристалликах, покрывающих листья лотоса. Поверхность соприкосновения загрязнений с поверхностью листа также крайне

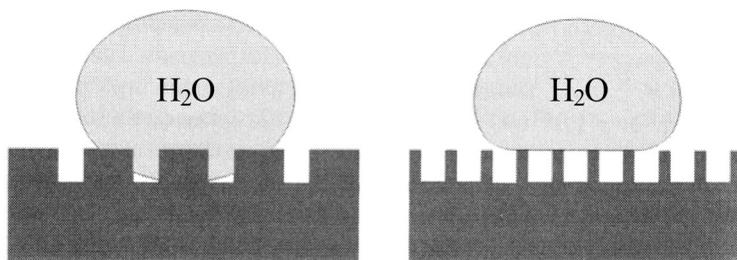


Рис. 9.14. Положение капли воды на макро (слева) и наноповерхности (справа)

незначительна. При этом силы сцепления между каплей воды и частицей загрязнения оказываются значительно более высокими, чем между этой же частицей и восковым слоем листа.

9.3. Оптические метаматериалы

Метаматериалы (от греч. *meta* – над, после) – это искусственные композитные среды, электрический и магнитный отклики которых существенно отличаются от соответствующих откликов в составляющих средах, благодаря чему возникают свойства, недостижимые в природных материалах.

Одним из наиболее ярких примеров таких материалов являются так называемые левые среды, в которых реализуется отрицательная рефракция (изменение преломления лучей при изменении показателя преломления среды) электромагнитных волн.

Интерес к отрицательной рефракции электромагнитных волн, при которой преломленный луч отклоняется по другую сторону от нормали к границе раздела сред, возник в начале XXI века после появления публикации группы ученых из университета Сан-Диего (США), возглавляемой Д.Р. Смитом. Они сообщили о создании композитных материалов, обладающих *отрицательным показателем преломления*, при этом они сослались на работы советского физика В. Веселаго 1967 года.

Веселаго показал, что в средах с одновременно отрицательными значениями диэлектрической и магнитной проницаемостей показатель преломления изменяет знак. Следствием этого является изменение на противоположное направления излучения, которое образует с векторами напряженностей электрического и магнитного полей левовинтовую тройку. Такие среды он назвал «*левыми*». Веселаго показал также, что в левых средах должны наблюдаться и другие аномальные явления: изменение знака групповой скорости, рассеяние света выпуклой линзой и, наоборот, его фокусировка вогнутой линзой.

Веселаго в действительности не был первым, кто теоретически исследовал преломление света в среде с отрицательной групповой скоростью. Детальный анализ отрицательной рефракции был дан еще в 1944 году Л.И. Мандельштамом. Кроме того, сам факт существования структур, в которых волна имеет отрицательную групповую скорость, также не являлся секретом. Механические модели одномерных сред подобного типа были исследованы еще в 1904 году Лэмбом, показавшим, что в так называемых обратных волнах фазовая скорость противоположна по направлению групповой скорости и потоку энергии.

Тем не менее именно статья Веселаго благодаря простоте своего изложения и широкой известности журнала, в котором она была опубликована, приобрела характер основополагающей работы в теории отрицательно преломляющих сред, называемых также левыми средами и средами Веселаго. В англоязычной литературе такие среды называют NIM (*negative index materials*) или LHM (*left-handed materials*). В последнее время все большее распространение получает термин «*метаматериалы*», указывающий на то, что свойства этих материалов зависят не от их химического состава, а от особенностей искусственно созданной конструкции из наноразмерных емкостных и индуктивных элементов. Резонансные свойства этих материалов таковы, что приводят к отрицательным значениям диэлектрической и магнитной проницаемостей в определенном диапазоне частот.

В 2000 году идею Веселаго о фокусирующих свойствах плоскопараллельной пластинки из материала с отрицательным показателем преломления подхватил английский физик Дж. Пендри, показавший, что в этом случае отсутствует дифракционный предел на размер фокального пятна, присущий обычным линзам. Подобное фокусирующее устройство Пендри назвал совершенной линзой (*perfect lens*). Это означает, что можно создавать оптические микроскопы с недоступным ранее разрешением.

В 2006 году Дж. Пендри выступил с новой идеей: если окружить объект материалом, показатель преломления которого плавно изменяется от 0 на внутренней поверхности до 1 на внешней границе, то свет будет огибать объект, который становится невидимым для наблюдателя. Идея была успешно реализована в микроволновом диапазоне. Возможность создания такого «плаща-невидимки», способного сделать объект невидимым, вызвала чрезвычайный интерес у военных.

Природа отрицательной рефракции

На рис. 9.15 показана схема хода лучей плоской волны под углом падения i к границе раздела двух сред. Фазовая скорость волны в верхней среде равна v_1 , а в нижней среде – v_2 .

Из построения следует закон преломления:

$$n_1 \cdot \sin i = n_2 \cdot \sin r ; \quad \frac{\sin i}{\sin r} = \frac{n_2}{n_1} = n = \frac{v_1}{v_2} . \quad (9.1)$$

Однако наряду с выражением (9.1) справедливо выражение

$$\frac{\sin i}{\sin(\pi - r)} = n. \quad (9.2)$$

Ему соответствует преломленная волна, в которой фазовая скорость и волновой вектор направлены не от границы, а к границе раздела (рис. 9.15, б).

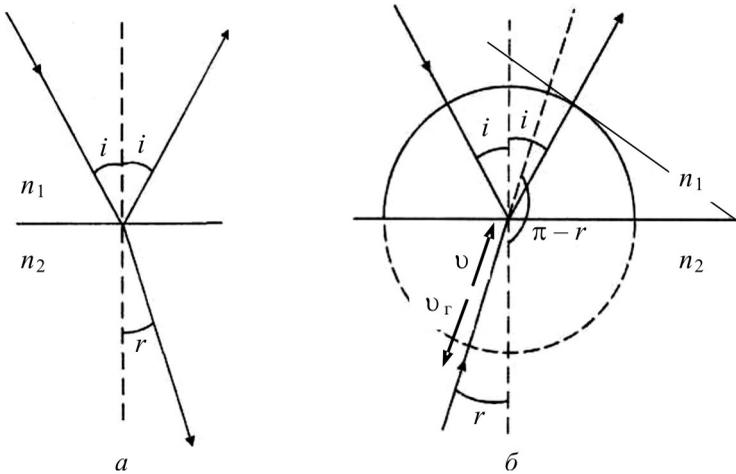


Рис. 9.15. Схема хода лучей при отражении и преломлении плоской волны в изотропной недиссипативной среде в случае положительной (а) и отрицательной (б) рефракции

С формальной математической точки зрения эта ситуация соответствует отрицательной фазовой скорости преломленной волны ($v = v_2 < 0$), отрицательному показателю преломления ($n < 0$) в (9.1) и отрицательному углу преломления ($r < 0$).

Таким образом, термин «отрицательная рефракция» обозначает ситуацию, когда в преломленной волне волновой вектор, направление которого совпадает с направлением фазовой скорости, имеет отрицательную проекцию на направление волнового вектора падающей волны. Термин «отрицательная фазовая скорость» является относительным: $v < 0$ лишь тогда, когда направление волнового вектора падающей волны определено как положительное. По этой причине термин «отрицательный показатель преломления» следует понимать в том смысле, что эффективный показатель преломления среды является отрицательным.

Мандельштам объяснил обсуждаемый эффект тем, что в данном случае преломленная волна является *обратной волной*, в которой фазовая и групповая скорости имеют противополож-

ные знаки (см. рис. 9.15, б). Групповая скорость v_r и поток энергии по-прежнему направлены от границы раздела вглубь второй среды, так что принцип причинности не нарушается.

Перспективы практического использования метаматериалов

Отрицательные значения эффективный показатель преломления может принимать лишь в магнитодиэлектриках. Первые попытки поиска таких материалов, у которых в определенном диапазоне частот одновременно становились бы отрицательными в соответствии с формулой (6.3) относительные диэлектрическая и магнитная проницаемости среды, были предприняты еще В. Весселаго. Он пытался создать материал с отрицательным преломлением на основе магнитного полупроводника CdCr_2Se_4 , однако эти усилия не увенчались успехом из-за существенных технологических трудностей, которые характеризуют синтез этого материала. Позднее положительные эффекты были достигнуты на композитных наноматериалах, полученных методами фото- и электронно-лучевой литографии.

Уникальные свойства метаматериалов, в частности возможность получения в них отрицательных значений показателя преломления, позволяют надеяться, что в недалеком будущем появятся созданные на их основе совершенно новые, неизвестные ранее технические устройства.

Одним из таких устройств являются совершенные линзы. Обычные линзы, по сути дела, являются дифракционными устройствами, разрешающая сила которых ограничена. Так, угловое расстояние между двумя точками, при котором они еще воспринимаются раздельно,

$$\delta\varphi = 1,22\lambda/D. \quad (9.3)$$

Для зрачка человеческого глаза при нормальном освещении ($D = 2$ мм, $\lambda = 500$ нм) получаем $\delta\varphi \approx 1'$.

Как было показано Дж. Пендри, плоскопараллельная пластинка из материала с показателем преломления $n = -1$, играющая роль фокусирующей линзы, не обладает этим недостатком. В этом случае фокусировка света имеет совершенно иную физическую природу: не дифракционную, а рефракционную. Поэтому световые лучи с помощью такой линзы могут быть сфокусированы практически в точку. Интенсивность света в фокусе теоретически может стремиться к бесконечности. Подобные совершенные линзы позволят резко увеличить разрешающую способность оптических микроскопов, сделать ее независимой от длины волны света.

Другая идея связана с созданием оболочек, способных сделать укрываемый ими объект невидимым для наблюдателя. Действительно, представим себе некий объект, окруженный оболочкой, показатель преломления которой изменяется от $n = 1$ на внешней поверхности до $n = 0$ на внутренней поверхности. Тогда внешнюю поверхность такого «плаща» световые лучи будут пересекать без отражения и преломления. По мере проникновения в глубь слоя угол преломления возрастает от значения, равного углу падения, до 90° , так как при $n = 0$ для любого падающего луча имеет место полное внутреннее отражение. Свет будет попросту обтекать предмет подобно тому, как обтекала бы его жидкость.

Поскольку отраженные лучи отсутствуют, наблюдатель предмета не видит. Правда, для этого надо позаботиться о том, чтобы отсутствовали или, по крайней мере, были незначительны потери света на поглощение в укрывающем слое, иначе наблюдатель увидит перед собой некоторое темное пятно непонятного происхождения. Для уменьшения потерь на поглощение света такие «плащи-невидимки» должны быть достаточно тонкими.

Метаматериалам на основе периодически распределенных элементов присущи следующие недостатки:

- сильная зависимость показателя преломления от частоты, угла падения и поляризации падающего излучения;

- невозможность плавного и динамического (посредством внешних полей) управления показателем преломления;
- сложность получения протяженных по площади, гибких пленок с отрицательным показателем преломления.

Эти недостатки делают проблематичным использование таких метаматериалов для тех основных целей, ради которых и затевалось их создание: использование в линзах субволнового разрешения и в качестве материалов для «плащей-невидимок», укрывающих объект от внешнего наблюдателя.

Преодоление этих недостатков связано с возможностями нанотехнологий по созданию отрицательно преломляющих материалов для оптического диапазона электромагнитных волн.

9.4. Методы формирования наноструктур

Развитию методов формирования наноструктур (нанотехнологии) во многом способствовало постоянное совершенствование традиционных методов изготовления полупроводниковых приборов, таких как химическое осаждение из газовой фазы, молекулярно-лучевая эпитаксия и электронно-лучевая литография. Кроме того, применение сканирующих атомарно-острых зондов и саморегулирующихся процессов значительно расширило возможности создания оптических наноструктур.

Различают два основных подхода, позволяющих формировать наноструктуры. Это технологии, реализующие принцип «сверху-вниз», и технологии, построенные на принципе «снизу-вверх». *Принцип «сверху-вниз» (top-down approach)* предполагает создание структур с требуемыми размерами и конфигурацией путем избирательного удаления материала, заранее нанесенного на подложку. При этом используются традиционные методы. Обыденным примером технологии «сверху-вниз» является создание скульптуры из монолитной каменной глыбы путем отсечения «лишнего» материала.

Альтернативный принцип *«снизу-вверх» (bottom-up approach)* предполагает формирование требуемых структур путем селектив-

ного осаждения атомов и молекул на заданные участки поверхности подложки. Так же поступает художник, нанося определенные краски на определенные участки поверхности холста.

Нанотехнологические приемы постоянно совершенствуются, появляются новые методы. Классификация способов получения наноматериалов представлена в табл. 9.2. Рассмотрим некоторые основные методы, уже прошедшие экспериментальную проверку и использующиеся достаточно широко.

Т а б л и ц а 9.2

Основные способы получения наноматериалов

Способ получения	Характеристика способа получения	Получаемые объекты
Фуллереновая дуга	Синтез в плазме дугового разряда между графитовыми электродами	Фуллереновая сажа, углеродные нанотрубки
Газофазный метод	Температура 4000 °С и выше для получения фуллерена C ₆₀	«Гостевые» наномолекулы
Каталитическое разложение углеводородов	Продувка по кварцевой трубке с металлическим порошком и температурой 700...1000 °С смеси газообразного углеводорода и буферного газа	Углеродные нити, многослойные нанотрубки, металлические частицы, покрытые графитовой оболочкой
Порошковая технология	Метод газофазного осаждения и компактирования; электроразрядное спекание; горячая обработка давлением; высокие статические и динамические давления при различных температурах	Металлы, сплавы
Интенсивная пластическая деформация	Равноканальное угловое прессование; деформация кручением; обработка давлением многослойных композитов	Металлы, сплавы
Кристаллизация из аморфного состояния	Обычные и высокие давления	Аморфные металлические покрытия
Пленочная технология	Химическое осаждение покрытий из газовой фазы (CVD); физическое осаждение из газовой фазы (PVD); электроосаждение, зольгель-технология	Металлы, сплавы, полимеры, химические соединения

9.4.1. Молекулярно-лучевая эпитаксия

Молекулярно-лучевая эпитаксия (*molecular beam epitaxy*) (МЛЭ) появилась как развитие метода химического осаждения пленок в сверхвысоком вакууме.

Химическое осаждение из газовой фазы (*chemical vapor deposition*) включает рост ориентированной монокристаллической пленки на монокристаллической подложке. Материал пленки при этом поступает из газовой фазы. Газовая среда может содержать как пары кристаллизующегося материала, так и газообразные реагенты, способные в процессе химических реакций на подложке образовывать необходимый для эпитаксиального роста материал. Осаждение пленок обычно проводится в проточных камерах, где газ-носитель, содержащий соответствующие реагенты, пропускается над нагретой монокристаллической подложкой. Химический состав, давление газа и температура подложки являются главными параметрами, контролирующими процесс осаждения пленок и их свойства.

Установка для химического осаждения из металлорганических соединений показана на рис. 9.16 на примере осаждения GaAs и гетероструктур GaAlAs. Главным преимуществом метода химического осаждения является возможность одновременной обработки большого количества подложек, что соответствует требованиям массового производства. Ограничением метода является плохо контролируемое загрязнение материала пленок, а также необходимость принимать серьезные меры безопасности при работе с взрывоопасными и токсичными гидридами.

При молекулярно-лучевой эпитаксии реагенты вводятся в рабочую камеру в виде молекулярных или атомных потоков. Эти потоки формируются путем испарения материала внутри замкнутой ячейки с очень малым выходным отверстием. Она называется *эффузионной (effusion)*, или *ячейкой Кнудсена*. Испаренные внутри нее молекулы и атомы, выходя из отверстия в сверхвысокий вакуум, движутся без соударений (баллистиче-

ски), создавая, таким образом, направленные, хорошо коллимированные потоки частиц.

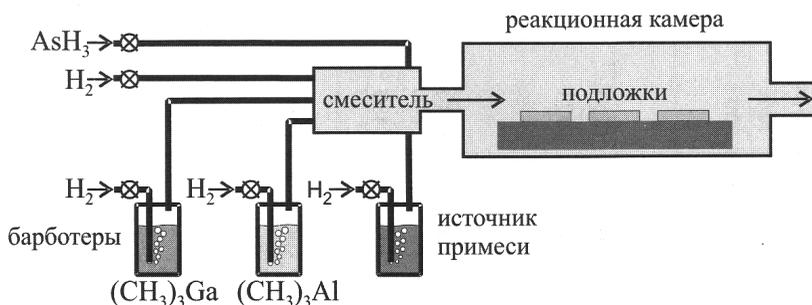


Рис. 9.16. Принципиальная компоновка установки для химического осаждения пленок из газовой фазы металлоорганических соединений

Для МЛЭ обычно используют несколько эффузионных ячеек – по одной на каждый испаряющийся материал. Кроме ячеек, для осаждения собственно материалов должны присутствовать также источники легирующих примесей. Наряду с испарением осаждаемого материала внутри эффузионной ячейки молекулярные потоки могут формироваться по такому же принципу и из паров или газообразных соединений. Для этого их вводят в сверхвысоковакуумную камеру через специальные подогреваемые сопла.

Конструкция типичной установки МЛЭ, размещаемой в сверхвысоковакуумной камере, схематически показана на рис. 9.17. Потоки атомов или молекул создаются в зоне генерации (I) за счет испарения жидких или сублимации твердых материалов, помещенных в эффузионные ячейки (источники). Потоки атомов (молекул) направляются на подложку, проходя зону смешивания (II), и осаждаются на ней в зоне роста (III), образуя пленку из вещества требуемого состава. Зону роста можно разделить на три области, первая из которых представляет собой подложку или очередной выросший моноатомный слой мате-

риала. Вторая область – газовая смесь компонентов гетероструктуры в приповерхностной области. Третья область – переходный слой, геометрия которого и протекающие в нем процессы сильно зависят от выбора условий роста.

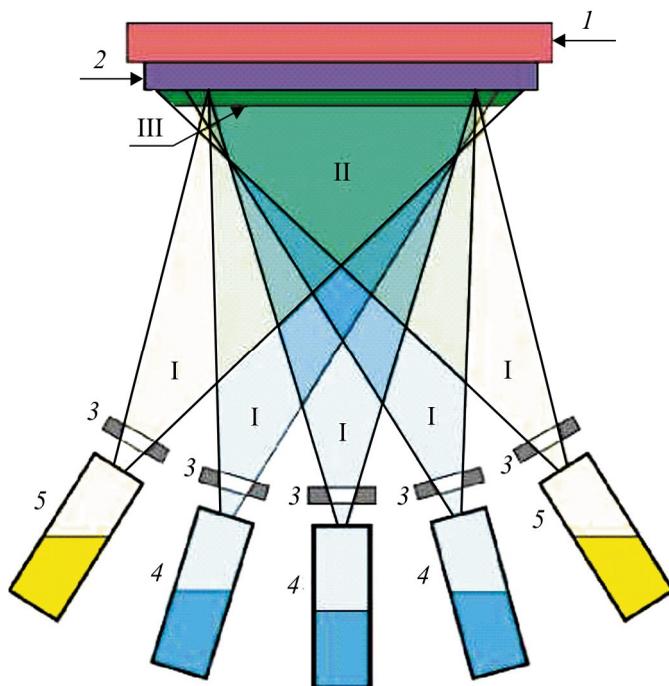


Рис. 9.17. Схема МЛЭ-установки: 1 – подложка; 2 – растущая пленка; 3 – заслонки; 4 – эффузионные ячейки основных компонентов; 5 – эффузионные ячейки легирующих примесей; I – зона генерации молекулярных пучков; II – зона смешивания пучков; III – зона кристаллизации на подложке (зона роста)

Управление составом выращиваемого материала и концентрацией легирующих примесей осуществляется с помощью заслонок, которые перекрывают тот или иной поток частиц. Для повышения однородности во многих случаях подложка с растущей пленкой постоянно вращается.

В зоне роста одного моноатомного слоя, в пределах одной секунды, атом совершает несколько тысяч диффузионных прыжков, пока не займет свое окончательное положение в решетке (рис. 9.18). Так как химические связи в различных материалах разные, то различаются и энергии активации поверхностной диффузии атомов, входящих в состав этих соединений. В связи с этим качество гетерограниц может существенно отличаться в зависимости от того, какое из соединений при выбранном температурном режиме растет первым. Границы называют *нормальными*, если компонент с более низкой температурой плавления растет первым. Если последовательность роста обратная, то такие границы называют *инвертированными*. Иллюстрация границ этих типов, обозначенных индексами n и i , дается на рис. 9.18 на примере структуры $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$

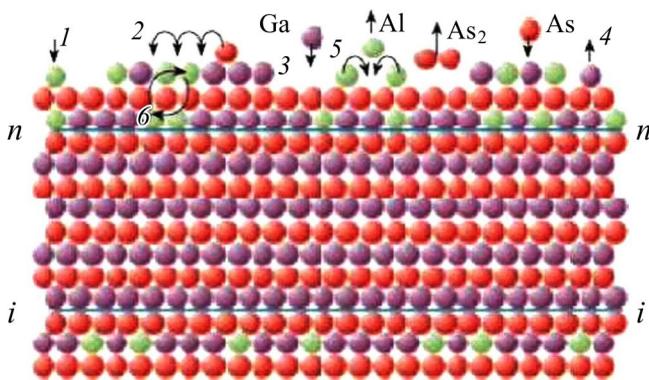


Рис. 9.18. Элементарные процессы в зоне роста: 1 – адсорбция атомов из зоны смешивания на поверхности; 2 – миграция адсорбированных атомов по поверхности; 3 – встраивание адсорбированных атомов в кристаллическую решетку; 4 – термическая десорбция; 5 – образование поверхностных зародышей; 6 – взаимная диффузия. Над растущей поверхностью показаны атомы газовой смеси компонентов в приповерхностной области. Буквами n - n и i - i показаны нормальная и инвертированная поверхности раздела растущей гетероструктуры

Таким образом, имеет место своего рода самоорганизация растущей структуры, поэтому МЛЭ является самым эффективным методом получения многослойных полупроводниковых систем, *гетероструктур* и *тонких пленок* с контролем толщины на атомном уровне. Большие размеры, сложность конструкции, высокая стоимость (до десятков миллионов долларов) делает комплексы МЛЭ «пирамидами» века нанотехнологий.

9.4.2. Нанолитография

Различные методы микрогравировки слоев обобщенно называются литографией. Различают фотолитографию, рентгеновскую литографию, электронно-лучевую литографию и т.д. Постоянное совершенствование методов литографии открыло возможность перехода от микроэлектроники к наноэлектронике, в рамках которой литография продолжает успешно развиваться.

Суть методов микрогравировки можно уяснить на примере фотолитографии, простейшего вида литографии. *Фотолитография представляет собой метод фотохимической микрогравировки металлических, диэлектрических и полупроводниковых слоев.* Основные этапы фотолитографии на пластине кремния представлены на рис. 9.19:

- нанесение на пластину слоя диэлектрика, обычно диоксида кремния SiO_2 (рис. 9.19, *а*);
- нанесение на слой диэлектрика фоточувствительного слоя – фоторезиста (рис. 9.19, *б*);
- наложение (при контактной фотолитографии) на слой фоторезиста фотошаблона, который отображает соответствующую часть формируемой топологической схемы; фотошаблон представляет собой непрозрачную пластину с прозрачными участками, дублирующими форму и местоположение будущей схемы (рис. 9.19, *в*);
- экспонирование фоторезиста (в простейшем варианте видимым или ультрафиолетовым светом); экспонирование из-

меняет скорость последующего растворения фоторезиста в специальном травителе (на рис. 9.19, *в* экспонирование отображено системой стрелок);

- удаление фотошаблона;
- проявление (травление) фоторезиста; участки, подвергнутые воздействию света, вытравливаются до слоя окисла (рис. 9.19, *з*);

- вытравливание отверстий («окон») в слое диэлектрика через отверстия в фоторезисте (1 и 2 на рис. 9.19, *д*);

- удаление фоторезиста (рис. 9.19, *е*).

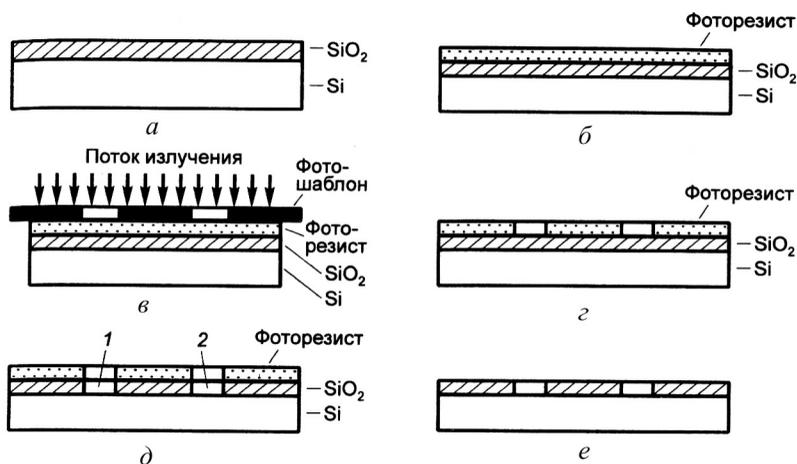


Рис. 9.19. Основные этапы контактной фотолитографии

Полученные окна в диэлектрике используются для формирования соответствующих элементов схемы на пластине. Если данный элемент формируется диффузией примесей, то вещество диэлектрического слоя выбирается так, чтобы скорость диффузии данной примеси в нем была значительно меньше, чем в кремнии. В таком случае легированными окажутся лишь участки кремния, не укрытые слоем диэлектрика. Диффузию большинства типичных доноров и акцепторов в кремниевой технологии хорошо блокирует диоксид кремния (SiO_2).

Разрешающая способность литографии ограничивается как техническими, так и физическими факторами. Принципиальные физические ограничения связаны главным образом с явлением дифракции излучения на деталях фотолитографического рисунка при экспонировании. Дифракционное размытие изображения отверстия тем меньше, чем меньше длина волны используемого излучения. Оптическая литография обеспечивает разрешение с размерами до 0,5 мкм. Использование ультрафиолетовых лучей

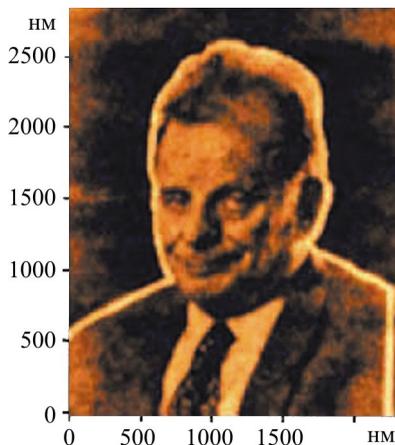


Рис. 9.20. Анодно-окислительная литография портрета Ж.И. Алферова с использованием атомного силового микроскопа

позволяет сместить этот предел до 0,1 мкм. Рентгено-, ионолитография сдвигает указанный предел в нанобласть с разрешающей способностью до 10 нм.

В качестве примера реализации этой технологии на рис. 9.20 представлен литографический портрет российского ученого, лауреата Нобелевской премии Ж.И. Алферова, выполненная на сверхтонкой титановой пленке с использованием атомно-силового микроскопа методом локального зондового электрического окисления.

9.4.3. Сканирующая туннельная и атомно-силовая микроскопия

Методы создания наноразмерных структур с применением сканирующих точечных зондов берут свое начало от *сканирующей туннельной микроскопии* (СТМ) и *атомно-силовой микроскопии* (АСМ). Роль этих методов быстро эволюционировала от фундаментальной (научной) к прикладной (технологической). Они сочетают в себе возможность исследовать по-

верхность образца с атомным разрешением и возможности манипулирования.

Принципы и конструкцию первого устройства для *сканирующей туннельной микроскопии* предложили в 1981 году Герд Бинниг и Генрих Рорер, работавшие в то время в Цюрихе. Позже, в 1986 году, за работы по сканирующей туннельной микроскопии они были удостоены Нобелевской премии по физике.

Физическую основу сканирующей туннельной микроскопии составляют явления, определяемые туннелированием электронов в зазоре между атомарно-острым зондом и поверхностью анализируемого образца. Туннельный ток через зазор очень чувствителен к структурным неоднородностям поверхности образца. Поэтому, перемещая зонд вдоль поверхности и контролируя протекающий по нему туннельный ток, можно анализировать топологию поверхности с атомным разрешением.

Металлический зонд, обычно изготавливаемый из вольфрама (рис. 9.21), закрепляют в держателе, пространственное положение которого регулируется управляющим напряжением. Зонд подводят к образцу на расстояние, обеспечивающее протекание туннельного тока, и пьезоэлементами, задающими его положение в плоскости образца, сканируют вдоль поверхности.

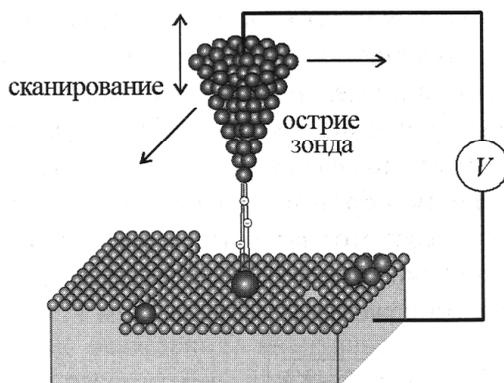


Рис. 9.21. Относительное расположение зонда и подложки в сканирующем туннельном микроскопе

Таким образом, удастся не только «увидеть» расположение атомов на поверхности, но и различить области с разным химическим составом. Вертикальное разрешение при этом достигает 0,01–0,05 нм, а горизонтальное – 0,3 нм. Размер анализируемой поверхности обычно составляет сотни микрометров. Ограничением метода является требование высокой электропроводности исследуемого материала, что необходимо для протекания достаточного для регистрации туннельного тока.

В *атомной силовой микроскопии* для анализа поверхности вместо туннельного тока регистрируется сила взаимодействия между зондом и подложкой. Для определения этой силы острый зонд закрепляют на упругой консоли, как показано на рис. 9.22. Зондовый датчик АСМ называют *кантилевером* (англ. *cantilever* – консоль).

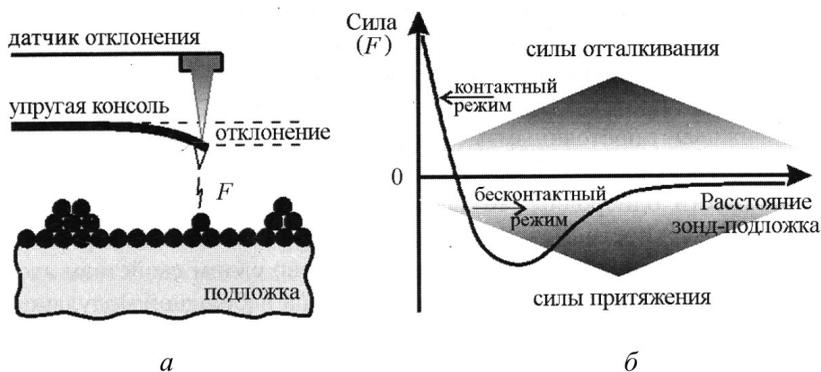


Рис. 9.22. Относительное положение зонда и подложки (а) и сила взаимодействия зонд-подложка, регистрируемая в атомном силовом микроскопе (б)

Отклонение консоли пропорционально действующей на нее силе. Это отклонение регистрируется с высокой точностью оптическими (например, посредством лазерной интерференции) или электронными (например, зондом сканирующего туннельного микроскопа) методами. При сканировании зонда вдоль анализи-

руемой поверхности сигнал об отклонении консоли дает информацию о распределении атомных и молекулярных сил по поверхности образца, а следовательно, и о расположении и природе поверхностных атомов.

В качестве кантилевера эффективно используются нанотрубки (рис. 9.23), они изгибаются при ударе о поверхность и не ломаются, а затем возвращаются в исходное состояние. Склонность нанотрубок к складыванию вместо разрушения делает повреждение кончика маловероятным. Благодаря малому поперечному сечению и большому отношению длины к диаметру такого зонда он может проникать в глубокие канавки на поверхности, которые недоступны для обычных зондов. Электропроводные нанотрубки могут использоваться и как зонды для сканирующей туннельной микроскопии.

Атомная силовая микроскопия, в отличие от сканирующей туннельной микроскопии, не чувствительна к электронным свойствам подложки. Поэтому она может быть использована для анализа поверхности как проводящих, так и диэлектрических материалов.

Перемещение атомов параллельно поверхности подложки может быть осуществлено в процессе *полевой диффузии* (*field diffusion*) или *скольжения* (*sliding*). В обоих случаях связи между перемещаемыми атомами и подложкой не разрываются. Адсорбированный (находящийся на поверхности) атом всегда находится в потенциальной яме. Энергия, необходимая для его перемещения, соответствует энергетическому барьеру для диффузии по поверхности. Она обычно находится в диапазоне 0,01–1,0 эВ.

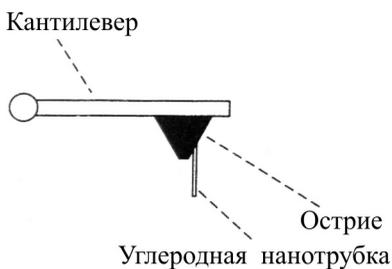


Рис. 9.23. Схема размещения однослойной углеродной нанотрубки на кантилевере атомного силового микроскопа

Полевая диффузия адсорбированных на поверхности атомов инициируется сильно неоднородным электрическим полем, создаваемым между острием зонда и поверхностью. Напряженность этого поля может достигать 30–50 В/нм. Этого вполне хватает для ионизации и десорбции (удаления) атомов. Потенциальная энергия этого поля добавляется к периодическому потенциалу поверхности (рис. 9.24, *а*), образуя потенциальный рельеф, благоприятный для направленного движения адсорбированного атома в область, находящуюся непосредственно под острием зонда. В зависимости от особенностей взаимодействия рассматриваемого атома и зонда возможно два варианта результирующего потенциального рельефа.

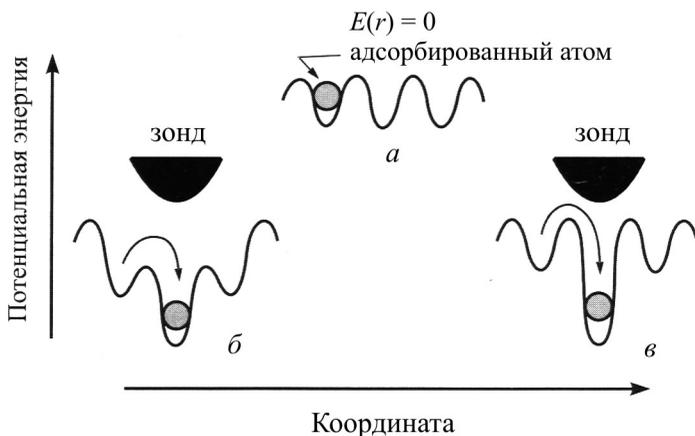


Рис. 9.24. Потенциальная энергия атома, адсорбированного на поверхности кристаллической подложки, как функция его положения относительно зонда сканирующего туннельного микроскопа

При слабом взаимодействии обычно формируется широкая потенциальная яма (рис. 9.24, *б*) с рельефом, модулированным периодическим потенциалом поверхности. В случае же сильного ориентированного взаимного притяжения атома и зонда, связанного с их химической природой, потенциальная яма

для атома сужается и локализуется строго под зондом (рис. 9.24, в). В процессе диффузии по поверхности адсорбированный атом «проваливается» в эту яму и задерживается в ней.

Классический пример, иллюстрирующий возможности манипулирования атомами на поверхности твердого тела, приведен на рис. 9.25. Изображения представляет собой распределение плотности электронных состояний, соответствующих атомам исследуемых материалов и подложки.

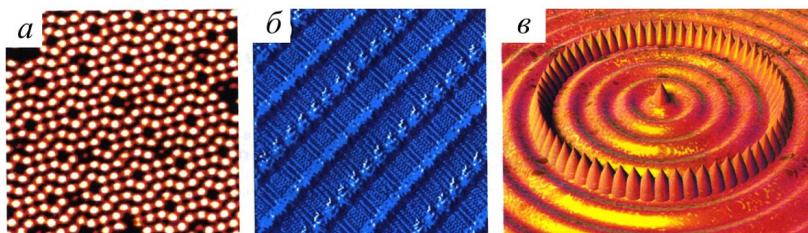


Рис. 9.25. Изображения поверхности Si (111) (а); монослоя динонадекан-бензена (б); квантового коралла из 48 атомов Fe на подложке Cu (111), радиус коралла 7,13 нм (в), полученные в сканирующем туннельном микроскопе

Рассмотренные подходы к манипулированию атомами на поверхности подложки составляют основу *атомной инженерии*. Они позволяют создавать наноразмерные структуры с заданным атомным составом. Реальные возможности, ограничения и практическая применимость для различных комбинаций атом-подложка сегодня являются предметом интенсивных научных исследований.

9.5. Применение нанотехнологий в технике

Нанотехнология обещает большие возможности при разработке новых материалов, совершенствовании связи, развитии биотехнологии, микроэлектроники, энергетики и вооружений. Среди наиболее вероятных научных прорывов эксперты называют увеличение производительности компьютеров, восстано-

ление человеческих органов с использованием вновь созданной ткани, получение новых материалов, созданных напрямую из заданных атомов и молекул, и появление новых открытий в химии и физике, способных оказать революционное воздействие на развитие цивилизации.

В соответствии с существующей классификацией все известные в настоящее время достижения практической нанотехнологии подразделяются на три группы: *инкрементные*, *эволюционные* и *радикальные*. Рассмотрим их более подробно.

Инкрементная нанотехнология подразумевает промышленное применение существующих наноструктур, а также специфических эффектов и феноменов, характерных для области перехода между атомным и мезоуровнями в целях значительного усовершенствования существующих классических материалов.

Наибольшее развитие инкрементные нанотехнологии получили в области получения композиционных конструкционных материалов с различными свойствами, защитных самоочищающихся покрытий, препаратов автохимии и некоторых других.

В качестве примера бытового применения нанотехнологий можно привести полироль, создающую «умную» поверхность с многофункциональными свойствами. Она представляет собой уникальную комбинацию бразильского карнаубского воска, синтетических восков, силиконов и неабразивных наноалмазов, обеспечивающую эффективную защиту, восстановление цвета и блеска лакокрасочных покрытий. Наноразмерные алмазы позволяют получить совершенную сотовую структуру пленки полироли с повышенной прочностью, износостойкостью, адгезией к поверхности, фотохимической и химической стойкостью, дополнительно обеспечивает эффект самоочищения поверхности при эксплуатации. Полироль эффективна при эксплуатации автомобиля в сложных погодных и климатических условиях, например в сырую и дождливую погоду, при езде по грязной дороге, а также в жаркую солнечную погоду – для защиты лакокрасочного покрытия от теплового и ультрафиолетового излучения.

Эволюционная нанотехнология связана с наномеханизмами, работы над которыми находятся на начальном этапе.

Как отмечено выше, по идее К.Э. Дрекслера, выдвинувшего концепцию универсальных молекулярных роботов, методами самосборки могут быть воссозданы любые объекты (в том числе и себе подобные) из подручных молекул, фуллеренов, нанотрубок и других аналогичных структур. Могут быть собраны структуры, имеющие форму разнообразных нанодеталей – зубчатых колес, штоков, деталей подшипников и других узлов, роторов молекулярных турбин, подвижных узлов манипуляторов и т.д.

Изделия нанотехнологии, созданные на основе оптимальной сборки атомов и молекул, позволят получить их предельно высокие характеристики.

На рис. 9.26–9.28 приведены примеры механических конструкций от простейших до довольно сложных, рассчитанных методами молекулярной динамики и собранных из наноконпонентов.

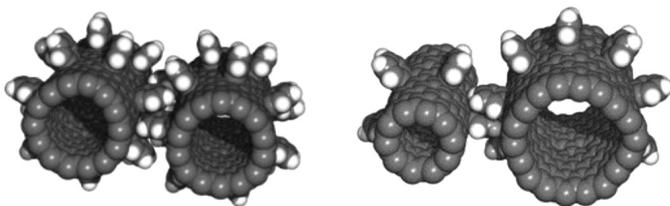


Рис. 9.26. Простейшие шестеренчатые передачи

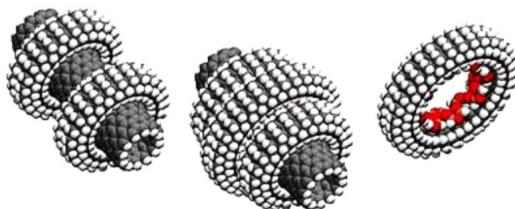


Рис. 9.27. Сборочные единицы наноподшипника

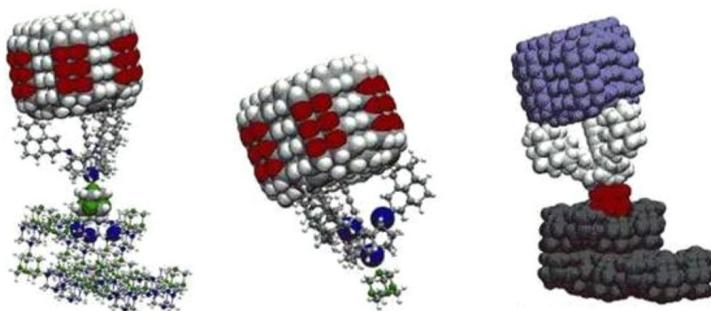


Рис. 9.28. Сборка конструкций наноманипуляторами

Наибольшего прогресса в этом направлении достигли японские ученые, создавшие микроскопический «подшипник», в котором потери на трение близки к нулю. Материалом для «вечного» миниподшипника послужили синтетические молекулы – фуллерены. В данном случае они состоят из 60 атомов углерода, расположенных в виде правильных пяти- и шестиугольников, которые вместе составляют шар. Эти вращающиеся «шарики» после сложного технологического процесса поставили в ряд между двумя удлиненными пластинками из графита.

Реализован принцип безызносного подшипника, простейшая схема работы которого представлена на рис. 9.29. Это достижение

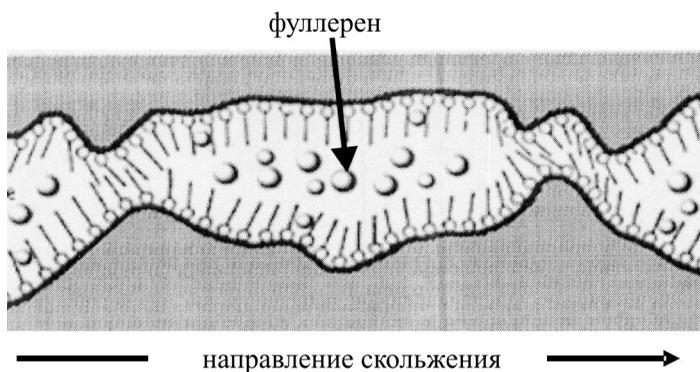


Рис. 9.29. Схема антиизносного механизма фуллеренов

нанотехнологий, по мнению ученых, планируется использовать в производстве миниатюрных роботов и микромеханизмов, детали которых практически не будут изнашиваться.

Радикальная нанотехнология – нанороботы (предполагаемые конструкции и результаты их использования в настоящее время существуют лишь в фантастических рассказах и кинофильмах).

Они могут быть способны к перемещению в окружающей среде и снабжены бортовой системой управления. Нанороботы смогут быть использованы для решения широкого круга задач, включая диагностику и лечение болезней, в том числе борьбу со старением, для перестройки организма человека «по заказу», изготовления сверхпрочных конструкций вплоть до лифтов «Земля – орбита» и т.д.

Вопросы для самоконтроля

1. Классифицируйте наноматериалы по геометрической форме и размерности структурных элементов.

2. Что такое гидрофильные и гидрофобные коллоидные кластеры? Какие типы мицелл могут образовываться в растворах с участием ПАВ?

3. Каковы различия между кристаллическими модификациями углерода – графитом, алмазом и фуллереном? Какие молекулы фуллерена называются фуллеритами?

4. Нанотрубки, их характеристики и области применения.

5. Какие технологии получения нанопленок вам известны?

6. В чем состоит технология получения нанопленок Ленгмюра–Блоджетт?

7. Приведите примеры природных нанокристаллов и практического применения искусственных нанокристаллов.

8. Приведите примеры эффективного применения наноповрхности. Почему они обладают феноменом самоочистки?

9. Что такое оптические метаматериалы? Каковы области их практического использования?

10. Какова суть отрицательной рефракции метаматериалов?

11. Метод молекулярно-лучевой эпитаксии пленок, его достоинства и недостатки.
12. Методы нанолитографии получения пленочных структур.
13. Метод сканирующей туннельной микроскопии формирования наноструктур.
14. Метод атомно-силовой микроскопии формирования наноструктур.
15. Приведите примеры применения эволюционной нанотехнологии в технике.

10. ЛАЗЕРЫ

Генерация света *лазером* основана на эффекте усиления вынужденного излучения в квантовой системе. Само английское слово *Laser* представляет собой аббревиатуру фразы «*Light Amplification by Stimulated Emission of Radiation*», которая дословно переводится как «Усиление света за счет вынужденного испускания излучения». Если излучение попадает в сверхвысокочастотный (микроволновый) диапазон, то усилитель такого типа называют *мазерным*. Соответствующий акроним (англ. *maser*) получен из предыдущего заменой слова *Light* на *Microwave*. Согласно смыслу акронима «лазер» его следует применять только по отношению к генераторам видимого излучения. Однако название «лазер» используется для обозначения устройства, испускающего любое вынужденное излучение. Чтобы уточнить тип излучения, говорят, соответственно, о лазерах инфракрасного, видимого, ультрафиолетового или рентгеновского диапазонов.

В полном виде концепцию лазера разработали Н.Г. Басов, А.М. Прохоров в России и Ч. Таунс, А. Шавлов в США. Первый лазер на кристалле рубина создал в 1960 году американский исследователь Т. Мейман.

Лазер (оптический квантовый генератор) представляет собой устройство, преобразующее различные виды энергии (электрическую, световую, химическую, тепловую и др.) в энергию

когерентного электромагнитного излучения оптического диапазона. Любой лазер, работающий как генератор когерентного излучения, должен состоять из трех элементов: источника энергии (*системы накачки*), *активной среды*, которая поглощает эту энергию и переизлучает ее в виде когерентного излучения, и *оптического резонатора* – устройства, осуществляющего обратную связь. В простейшей форме оптический резонатор образован двумя зеркалами (резонатор Фабри-Перо), от которых отражается излучаемая волна, вновь возвращаясь в активную среду, вызывая индуцированные переходы. Одно из зеркал делается полупрозрачным для выхода части излучения (рис. 10.1).

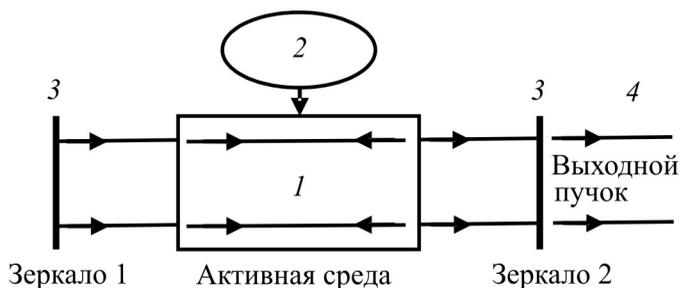


Рис. 10.1. Принципиальная схема лазера:
 1 – активная среда; 2 – система накачки; 3 – оптический резонатор; 4 – генерируемое излучение

Обратимся теперь к вопросу о том, как и при каких условиях возможно усиление света за счет вынужденного излучения.

10.1. Спонтанное и вынужденное излучение, поглощение

Для описания явления спонтанного излучения (рис. 10.2, а) рассмотрим два энергетических состояния 1 и 2 некоторого атома или молекулы данного вещества с энергиями соответственно E_1 и E_2 ($E_1 < E_2$). С точки зрения последующего рассмотрения это может быть любая пара из неограниченного набора состояний,

характерных для данного атома. Удобно, однако, принять состояние 1 за основное. Предположим, что первоначально атом находится в состоянии 2. Поскольку $E_2 > E_1$, то атом стремится перейти в состояние 1. В результате такого перехода атом должен выделить энергию, равную разности $(E_2 - E_1)$, называемую *энергией перехода*. Когда эта энергия выделяется в виде электромагнитной волны, процесс называют *спонтанным излучением*. При этом частота ν_0 излученной волны выражается известным соотношением

$$\nu_0 = (E_2 - E_1)/h, \quad (10.1)$$

в котором h – постоянная Планка. Таким образом, спонтанное излучение характеризуется испусканием фотона с энергией $h\nu_0 = (E_2 - E_1)$ при переходе атома из состояния 2 в состояние 1 (рис. 10.2, а). Отметим, что излучение фотона является для атома только одним из двух возможных способов перейти из одного состояния в другое. Такой переход может произойти также и без излучения фотона. В этом случае энергия перехода $(E_2 - E_1)$ выделяется в иной, отличной от электромагнитного излучения, форме (например, избыток энергии может перейти в кинетическую или внутреннюю энергию окружающих атомов или молекул). Такой процесс называют *безызлучательным переходом*, или *безызлучательной дезактивацией*.

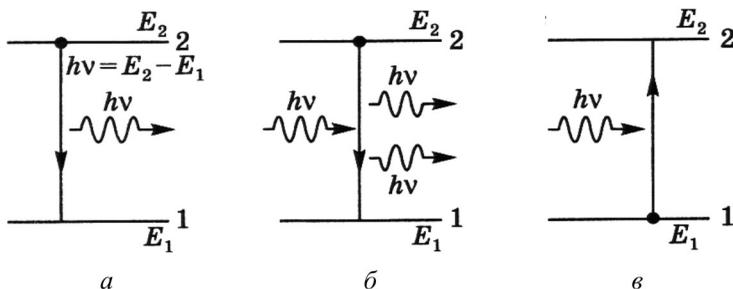


Рис. 10.2. Схематическая иллюстрация трех процессов: (а) спонтанного излучения, (б) вынужденного излучения, (в) поглощения

Предположим теперь, что атом первоначально находится в состоянии 2 и при этом на среду падает электромагнитная волна с частотой $\nu = \nu_0$, равной частоте волны, которая испускалась бы при спонтанном переходе $2 \rightarrow 1$ (рис. 10.2, б). Поскольку частоты этих двух волн одинаковы, оказывается, что существует конечная вероятность того, что падающая волна вызовет переход атома из состояния 2 в состояние 1. В этом случае энергия ($E_2 - E_1$) выделится в виде электромагнитного излучения, которое добавится к падающему. В этом и заключается явление *вынужденного излучения*, иногда называемого также *индуцированным излучением*.

Между процессами спонтанного и вынужденного излучения существует принципиальное различие. В случае спонтанного излучения различные атомы испускают электромагнитные волны, никак не связанные по фазе друг с другом. Более того, каждая из этих волн может быть испущена в любом направлении. В случае же вынужденного излучения, поскольку этот процесс вызывается падающей электромагнитной волной, волна, испущенная любым из атомов, добавляется к падающей, имея одинаковую с ней фазу и распространяясь в том же направлении.

Предположим теперь, что атом первоначально находится в состоянии 1 (рис. 10.2, в). Если это состояние – основное, то атом будет оставаться в нем до тех пор, пока не появится какое-либо действующее на него внешнее возмущение. Пусть на среду падает электромагнитная волна с частотой $\nu = \nu_0$. В этом случае существует конечная вероятность того, что атом перейдет в верхнее состояние 2. Энергия ($E_2 - E_1$), которая потребуется атому, чтобы осуществить этот переход, будет при этом заимствована из энергии падающей электромагнитной волны. В этом заключается процесс *поглощения*.

Таким образом, в каждом акте вынужденного излучения происходит рождение, а в каждом акте поглощения – уничтожение одного фотона.

10.2. Принцип работы лазера

Рассмотрим два произвольных энергетических состояния 1 и 2 некоторого вещества и обозначим соответственно через N_1 и N_2 их *населенности* (числа атомов, молекул в единице объема). Если в данном веществе вдоль оси z распространяется плоская электромагнитная волна с плотностью потока фотонов F (рис. 10.3), то малое изменение плотности потока dF в тонком слое толщиной dz (заштрихованная область на рис. 10.3) будет обусловлено как процессами вынужденного излучения, так и поглощения. Пусть через S обозначена площадь поперечного сечения пучка. Разность между числами фотонов, покидающих выделенный объем и поступающих в него в единицу времени, будет при этом равна SdF . Поскольку в каждом акте вынужденного излучения рождается, а в каждом акте поглощения уничтожается один фотон, то величина SdF должна равняться разнице между числом актов вынужденного излучения и поглощения, произошедших в рассматриваемом объеме в единицу времени.

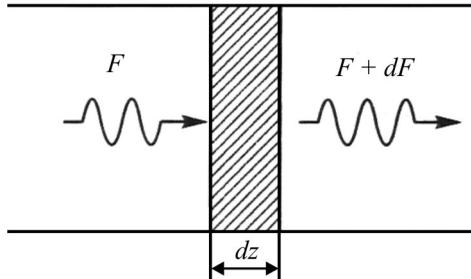


Рис. 10.3. Малое изменение плотности потока фотонов dF в плоской электромагнитной волне при прохождении через тонкий слой вещества толщиной dz

Изменение плотности потока

$$dF = \sigma F (N_2 - N_1) dz, \quad (10.2)$$

где σ – *поперечное сечение перехода*. Если реализуются неравновесные условия, при которых $N_2 > N_1$, то среда ведет себя как

усилитель света. Имея в виду то, что разность населенностей $N_2 - N_1$ противоположна по знаку той, которая существует при тепловом равновесии, говорят об *инверсии населенностей*.

Для превращения усилителя в генератор необходимо дополнить его положительной обратной связью, помещая активную среду в объемный резонатор между двумя зеркалами (см. рис. 10.1). Световая волна, распространяющаяся вдоль оси резонатора, многократно проходит через активную среду, непрерывно набирая энергию. Для самовозбуждения лазера необходимо достижение *пороговых* условий, при которых усиление света за счет вынужденного излучения в инвертированной среде начинает превышать величину потерь. Данное условие называют *амплитудным* условием самовозбуждения лазера. Другое условие самовозбуждения, называемое *фазовым*, состоит в том, что на длине резонатора должно укладываться целое число полуволн генерируемого излучения. Благодаря частичному пропусканию света одним из зеркал резонатора излучение выводится наружу, образуя лазерный луч.

10.3. Схемы накачки

В стационарных условиях теплового равновесия процессы поглощения и вынужденного излучения компенсируют друг друга ($N_2 = N_1$), эта ситуация называется *насыщением* двухуровневой системы. В этой системе невозможно создать инверсию населенностей.

В *трехуровневом лазере* (рис. 10.4, а) атомы «накачивают», с уровня 1 (основного) на уровень 3. Если среда такова, что атом, возбужденный на уровень 3, быстро переходит на уровень 2 (в результате быстрого безызлучательного перехода), то в этой среде можно получить инверсию населенностей уровней 2 и 1.

В *четырёхуровневом лазере* (рис. 10.4, б) атомы также переводят с основного уровня на уровень 4. Если после этого атом быстро переходит на уровень 3 (в процессе быстрой безызлучательной дезактивации), то можно и в этом случае получить ин-

версию населенностей уровней 3 и 2. Когда в таком четырехуровневом лазере начинается генерация, атомы за счет вынужденного излучения переходят с уровня 3 на уровень 2. Поэтому для работы четырехуровневого лазера в непрерывном режиме необходимо, чтобы переходы $2 \rightarrow 1$ также происходили очень быстро (обычно они обусловлены быстрыми безызлучательными переходами).

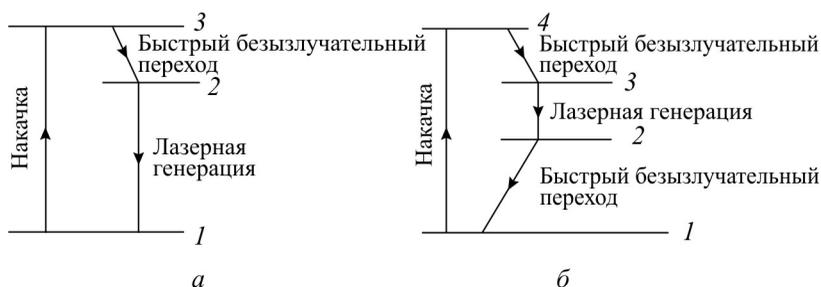


Рис. 10.4. Трехуровневая (а) и четырехуровневая (б) схемы лазера

Для того чтобы обеспечивалось выполнение пороговых условий генерации, скорость накачки должна достигнуть *пороговой* или *критической* величины.

10.4. Свойства лазерных пучков

Лазерное излучение характеризуется чрезвычайно высокой степенью *монохроматичности*, *когерентности*, *направленности* и *яркости*. К этим свойствам можно добавить еще одно, а именно – возможность генерации *сверхкоротких импульсов* света.

Свойство *монохроматичности* определяется следующими двумя обстоятельствами: во-первых, может быть усилена только электромагнитная волна с частотой ν_0 , определяемой выражением (10.1), во-вторых, генерация в резонаторе может происходить только на собственных частотах этого резонатора. Последнее обстоятельство приводит к тому, что ширина спек-

тра лазерного излучения значительно меньше (до 10 порядков), чем обычная ширина линии перехода $2 \rightarrow 1$, наблюдаемая при спонтанном излучении.

Когерентность в первом приближении для любой электромагнитной волны характеризуется двумя независимыми характеристиками: *пространственной когерентностью* и *временной когерентностью*.

Пространственную когерентность связывают с разностью фаз напряженностей электрического поля. Если эта разность фаз остается равной нулю в любой момент времени, то говорят, что между этими двумя точками имеется полная когерентность. Если такая когерентность существует между любыми парами точек волнового фронта, то говорят, что данная волна характеризуется полной пространственной когерентностью.

Понятие временной когерентности поясняется на рис. 10.5. Если разность фаз напряженностей поля остается постоянной при таких задержках времени τ , что $0 < \tau < \tau_0$, то говорят о частичной временной когерентности волны с характерным временем когерентности τ_0 .

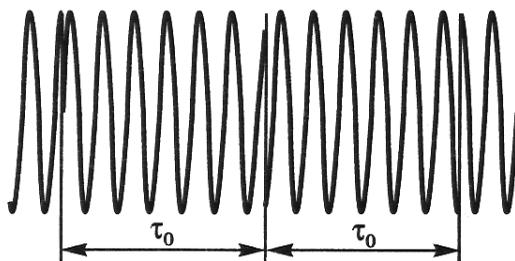


Рис. 10.5. Пример электромагнитной волны с временем когерентности порядка τ_0

Понятия временной и пространственной когерентности позволяют охарактеризовать только когерентность лазерного излучения первого порядка. Лазерное излучение принципиально от-

личается от света обычных источников за счет различий в соответствующих свойствах когерентности высших порядков.

Свойство *направленности* является простым следствием того, что активную среду помещают в резонатор. В случае плоскопараллельного резонатора, изображенного на рис. 10.1, только волны, распространяющиеся в направлении, перпендикулярном к плоскостям зеркал (или очень близком к нему), будут оставаться в резонаторе. В результате суммирования когерентных пучков в резонаторе выходной пучок будет иметь такие же поперечные размеры, что и пучок, приходящий из одной зоны когерентности. Это отличает направленность лазерного пучка от луча обычного источника света, подверженного дифракционной расходимости.

Яркость источника электромагнитных волн определяется как мощность излучения, испускаемого с единицы поверхности источника в единичный телесный угол. Яркость является важным параметром любого источника света. Если с помощью оптической системы формируется в воздушной среде изображение какого-либо источника света, то справедливо следующее: яркость изображения всегда меньше или равна яркости источника, при этом равенство сохраняется только в том случае, если в оптической системе нет потерь света, испущенного источником. Яркость лазера даже небольшой мощности на порядки превосходит яркость обычных источников света из-за высокой направленности лазерного пучка. Интенсивность сфокусированного лазерного излучения может достигать огромных величин, что используется во многих областях применения лазеров.

Применяя специальную технику, которая называется *синхронизацией мод*, возможно генерировать *сверхкороткие импульсы* света, длительность которых примерно равна обратной ширине линии лазерного перехода $2 \rightarrow 1$. Так, длительность импульса газовых лазеров может достигать $\sim 0,1$ – 1 нс. Такие длительности не считаются достаточно малыми, даже некоторые лампы-вспышки могут излучать импульсы длительностью менее

1 нс. С другой стороны, ширина спектра излучения у некоторых твердотельных и жидкостных лазеров может быть в 10^3-10^5 раз больше, чем у газовых лазеров; в этом случае можно получать гораздо более короткие импульсы (вплоть до ~ 10 фс). Характерный вид этих импульсов представлен на рис. 10.6.

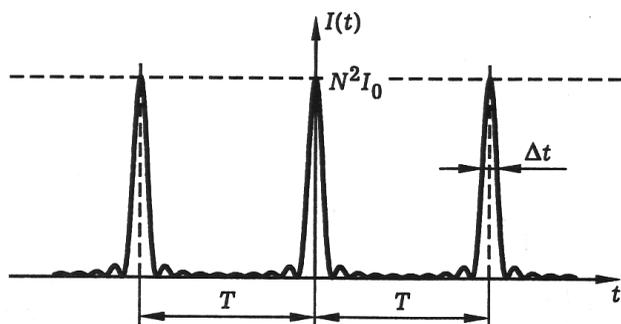


Рис. 10.6. Временной ход излучения лазера в режиме синхронизации мод

Свойство малой длительности импульса, предполагающее концентрацию энергии во времени, открывает новые возможности для исследовательских и технологических применений лазеров.

10.5. Типы лазеров

Различные типы лазеров, созданные к настоящему времени, демонстрируют широкий спектр физических и рабочих характеристик. Действительно, если лазеры классифицируют по физическому состоянию активной среды, то их называют *твердотельными*, *жидкостными* или *газовыми* лазерами. К особому типу относят такие, в которых активная среда состоит из свободных электронов, движущихся с релятивистскими скоростями через пространственно-периодическое магнитное поле (*лазеры на свободных электронах*).

Если лазеры классифицируют по длине волны генерируемого излучения, то говорят о лазерах *инфракрасного* (ИК), *видимого*, *ультрафиолетового* (УФ) или *рентгеновского* диапазонов. Соответствующие длины волн изменяются в пределах от ≈ 1 мм (т.е. от области миллиметровых волн) до ≈ 1 нм (т.е. до верхнего предела жесткого рентгеновского излучения). Разброс длин волн может достигать, таким образом, 10^6 (напомним, что видимый диапазон отвечает менее чем двукратному изменению длин волн – примерно от 400 до 700 нм).

Мощности излучения на выходе из лазеров перекрывают еще более широкий диапазон величин. У *непрерывных* лазеров типичные мощности составляют от нескольких мВт – в лазерах, используемых в качестве источников оптических сигналов, до нескольких МВт (≈ 5 МВт к настоящему времени) – в лазерах, необходимых для некоторых военных применений. У *импульсных* лазеров пиковая мощность может быть гораздо выше, чем у непрерывных, достигая таких гигантских величин, как 1000 ТВт (10^{15} Вт). Кроме того, длительность импульса лазерного излучения может изменяться в широких пределах – от нескольких мс, типичных для так называемого режима свободной генерации, до порядка 10 фс ($1 \text{ фс} = 10^{-15} \text{ с}$), что типично для некоторых лазеров с синхронизацией мод.

Сильно могут изменяться и геометрические размеры лазеров. В терминах длины резонатора, например, эта длина может быть малой, ~ 1 мкм, для наиболее коротких лазеров, и огромной, порядка нескольких километров, для наиболее протяженных (например, лазер для проведения геодезических исследований длиной 6,5 км, который был установлен в пещере).

Широта диапазона физических или рабочих характеристик лазеров является как достоинством, так и недостатком. Если говорить о применении, то широкий спектр параметров обеспечивает огромный потенциал возможного использования лазеров в различных областях фундаментальных и прикладных исследований. С другой стороны, с точки зрения рынка большое

разнообразие устройств и систем может служить препятствием для их массового производства и соответствующего удешевления продукции.

Рассмотрим примеры некоторых конкретных типов лазеров.

Лазеры на углекислом газе

Этот лазер относится к семейству молекулярных лазеров, в которых излучение генерируется в результате переходов между колебательными энергетическими уровнями молекул. CO_2 -лазер генерирует инфракрасное излучение на длине волны 10,6 мкм. Данный лазер отличается высоким КПД (15–20 %) и высокой мощностью генерации в непрерывном режиме, которая может достигать 10^6 Вт.

На рис. 10.7 показаны типы колебаний молекулы CO_2 (симметричное, деформационное, антисимметричное колебания), а на рис. 10.8 – квантовые уровни энергии, соответствующие этим колебаниям.

Возбуждение колебаний молекулы CO_2 осуществляется путем электронного удара в электрическом разряде. Рабочее вещество CO_2 -лазера состоит из смеси углекислого газа, азота

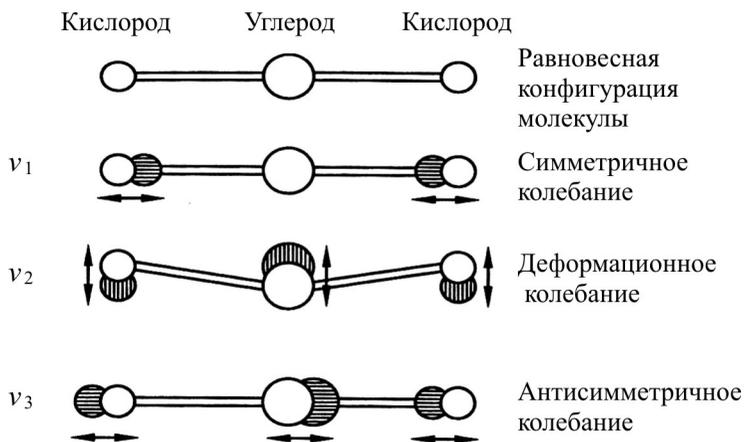


Рис. 10.7. Типы колебаний молекулы углекислого газа

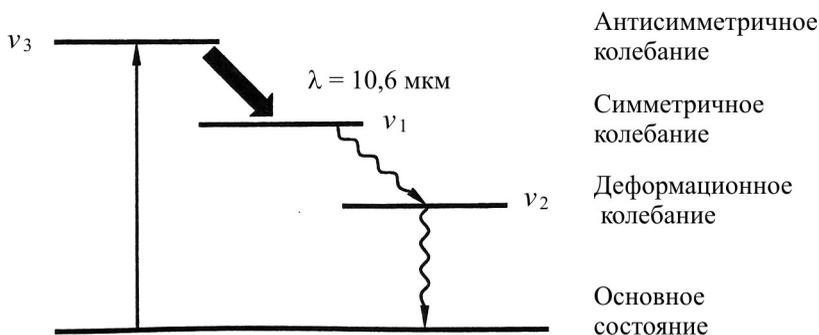


Рис. 10.8. Уровни энергии, соответствующие колебаниям молекулы углекислого газа и схема работы CO_2 -лазера

и гелия. Столкновения электронов газового разряда, атомов и молекул в такой смеси приводят к эффективному заселению верхнего и быстрому опустошению нижнего уровней лазерного перехода, что увеличивает инверсию населенностей и мощность генерируемого излучения.

Типичные лазеры на углекислом газе имеют длину около двух метров и могут создавать непрерывное лазерное излучение мощностью около 150 Вт. Ничто не препятствует созданию очень длинного лазера для излучения значительно большей выходной мощности, так как она линейно увеличивается с увеличением длины.

На установках для лазерного термоядерного синтеза созданы лазеры, генерирующие импульсы длительностью порядка 10^{-9} с энергией 10^5 Дж и мощностью до 10^{14} Вт. Фокусировка такого излучения позволяет получить интенсивность света на уровне 10^{19} Вт/см², при этом напряженность поля в световой волне достигает 10^{11} В/см, т.е. превышает напряженность внутриатомного электрического поля.

На рис. 10.9 показан внешний вид CO_2 -лазера, в котором газ проходит по разрядным трубкам (длина каждой около 20 см) с поперечным возбуждением, горячий газ охлаждается в теплообменнике.

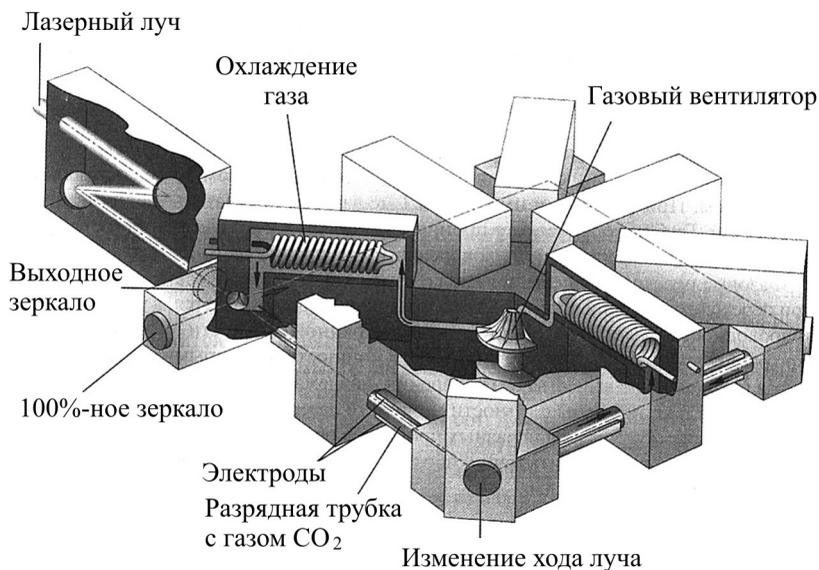


Рис. 10.9. Внешний вид CO_2 -лазера мощностью около 10 кВт для обработки материалов

CO_2 -лазеры являются важнейшим источником излучения при обработке материалов, однако их инфракрасное излучение пока еще не удастся эффективно проводить по гибким оптоволоконкам.

Полупроводниковые лазеры

В полупроводниковых лазерах излучение генерируется в области p - n -переходов в соединениях элементов III группы периодической таблицы (Al, Ga, In) с элементами V группы (N, P, As, Sb) и др. Резонаторами здесь могут быть грани кристалла, поэтому полупроводниковые лазеры компактны. Малые размеры, высокий КПД, продолжительный срок службы и простое обслуживание полупроводниковых лазеров – все это позволяет в большом объеме производить электрооптические устройства, например, для измерительной и аналитической техники, а также для бытовой электроники ежедневного пользования.

Их излучение с длиной волны около 800 нм весьма удобно при передаче информации по стекловолоконкам.

В 60-х годах появились полупроводниковые лазеры на *гетероструктурах* (рис. 10.10), в которых активной средой является узкозонный слой с высоким квантовым выходом. Благодаря особенностям формы активной зоны такие лазеры легко присоединяются к волноводам. Ширина активной области лазера при такой геометрии может быть уменьшена до 1 мкм, вследствие чего пороговые токи могут достигать 1 кА/см^2 .

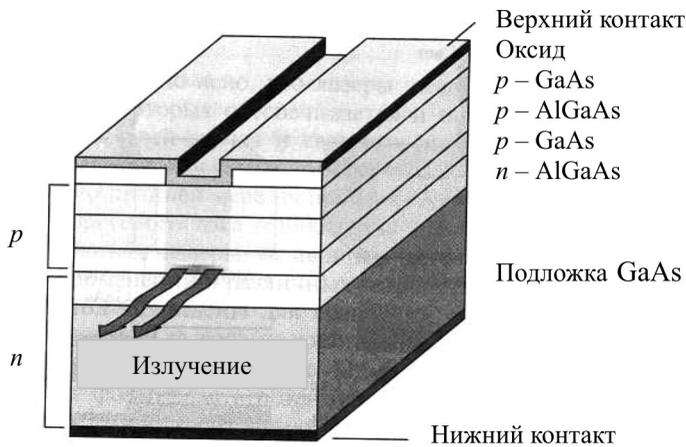


Рис. 10.10. Полосковый полупроводниковый лазер на двойной гетероструктуре

В конце 80-х годов стало ясно, что гетеропереходы позволяют формировать потенциальные ямы для электронов и дырок, повышая концентрацию носителей, увеличивая инверсную заселенность электронов и дырок, снижая плотность порогового тока до $\sim 50 \text{ А/см}^2$.

Самый первый лазер с резонатором Фабри–Перо с *самоорганизующимися квантовыми точками* был создан в 1994 году на основе точек из InGaAs в матрице GaAs (рис. 10.11). В настоящее время лазеры на квантовых точках (с излучением в ви-

димом и инфракрасном диапазонах) уже являются коммерческим продуктом. Миллионы таких лазеров, излучающих с поверхности, могут быть объединены на одном чипе, что, естественно, создает множество возможностей их практического использования в дисплеях и системах обработки оптических сигналов.

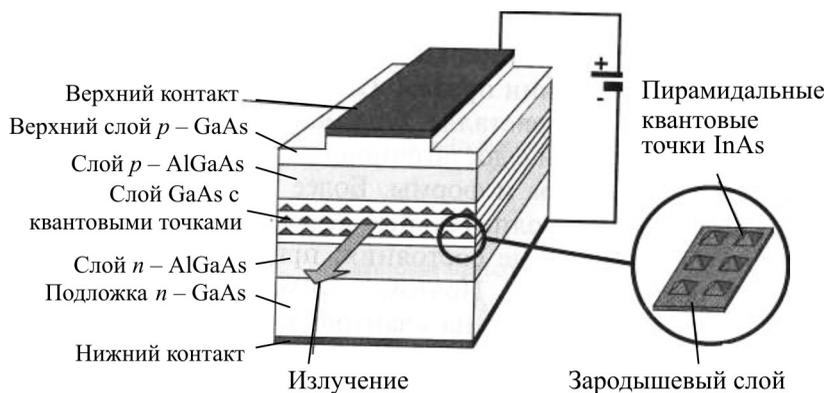


Рис. 10.11. Схема устройства лазера с краевым излучением на самоорганизованных квантовых точках (на вставке показан зародышевый слой с пирамидальными квантовыми точками)

Представленные на рис. 10.12 значения плотности порогового тока для разнообразных лазерных структур за последние десятилетия показывают, что на лазерах с квантовыми точками уже достигнуты самые низкие значения порогового тока.

Сравнение идеальных систем разной размерности с квантовой локализацией по коэффициенту усиления показывает (рис. 10.13), что квантовые точки позволяют добиться максимальной остроты пиков спектра при самых высоких значениях коэффициента усиления. Линия излучения идеального лазера на квантовой точке должна быть исключительно узкой. Узкий энергетический интервал излучения лазера на квантовых точках не зависит от температуры, т.е. квантовые точки обеспечивают температурную стабильность и не требуют охлаждения.

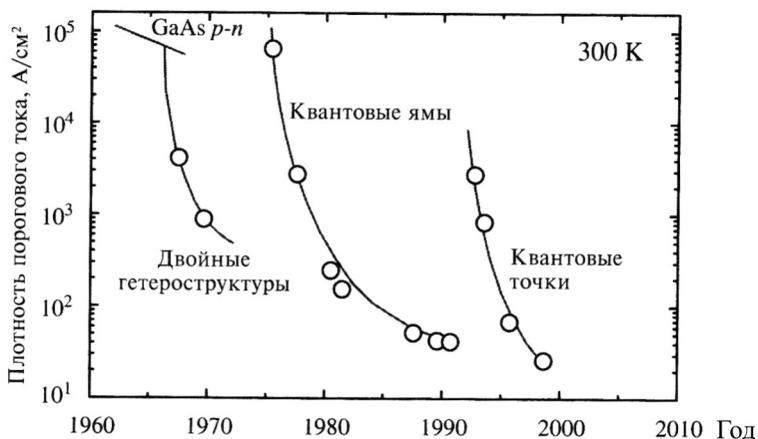


Рис. 10.12. Плотность порогового тока для лазерных структур с различной локализацией, достигнутая за последние десятилетия

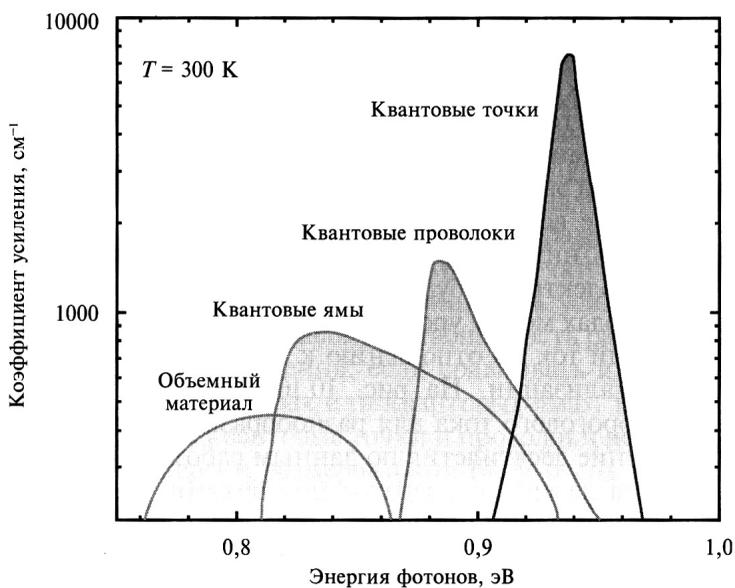


Рис. 10.13. Спектр коэффициентов усиления для лазеров на идеальных объемных полупроводниках, квантовых ямах, квантовых проволоках и квантовых точках

Явным преимуществом лазеров на квантовых точках является и то, что их электронная структура может управляться путем подбора материала, а также размеров и формы точек. Однако настоящий прогресс в создании лазеров на квантовых точках связан с развитием методов создания квантовых точек на основе так называемой *самоорганизации* (самосборки). Эта область нанотехнологий продолжает оставаться одной из перспективных.

Прорыв в лазерной физике ознаменовало появление *волоконных лазеров*, выходная мощность которых за последние 10 лет была увеличена с нескольких сотен ватт до нескольких десятков киловатт (рис. 10.14). На Всероссийской конференции по волоконной оптике в 2011 году сообщалось о передаваемой мощности 50 кВт. Беспрецедентно стремительный рост выходной мощности волоконных лазеров сдерживается не возможностями реализации, а наличием платежеспособного спроса.

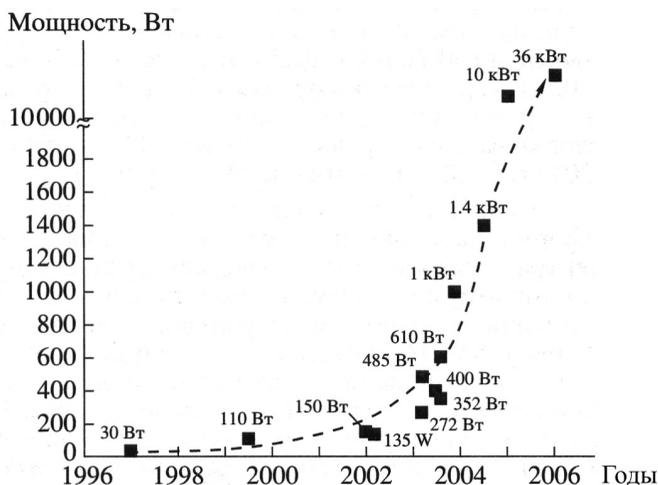


Рис. 10.14. Рост максимальной выходной мощности непрерывных волоконных лазеров за последние 10 лет

Разработаны эффективные волоконные лазеры на основе редкоземельных элементов – неодима, европия, гольмия и тулия, генерирующие в ближней инфракрасной области спектра. Спек-

тральные области лазерной генерации существующих эффективных редкоземельных волоконных лазеров не перекрывают спектрального диапазона 1150–1500 нм (рис. 10.15), перспективного, в частности, для волоконно-оптических систем связи. При этом наибольшую ширину спектральной области имеют *висмутовые волоконные лазеры* на основе алюмосиликатного стекла.

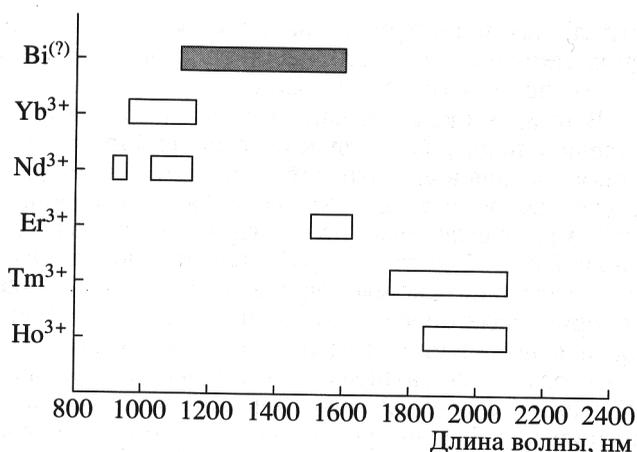


Рис. 10.15. Спектральные области генерации волоконных лазеров на редкоземельных элементах и потенциальная область генерации висмутового волоконного лазера

До недавнего времени при обработке материалов (резка металлических и пластиковых листов, сварка, сверление и т.д.) в основном применялись мощные CO_2 -лазеры, однако они громоздки, не имеют волоконного выхода, не обладают высокой эффективностью и качеством пучка, то есть по этим параметрам уступают волоконным лазерам. Кроме того, на волне излучения CO_2 -лазера (10,6 мкм) металлы являются хорошим зеркалом, что снижает эффективность их обработки. Коэффициент отражения от металлов падает по мере того, как укорачивается длина волны. И с этой точки зрения более предпочтительны *иттербиевые волоконные лазеры*, генерирующие на волне длиной 1 мкм. Ожида-

ется, что по мере дальнейшего совершенствования элементной базы и снижения стоимости лазерных диодов, используемых в качестве источников накачки, волоконные лазеры будут вытеснять CO₂-лазеры из сферы обработки материалов.

10.6. Области применения лазеров

Рассмотрим на конкретных примерах некоторые области применения лазерной техники.

Передача информации по стекловолоконкам

Генерируемая полупроводниковым лазерным диодом (инжекционным лазером) как передатчиком световая волна способна передавать информацию на расстояния до 100 км через тонкое стекловолокно диаметром в несколько тысячных долей миллиметра (рис. 10.16). Соединенные промежуточными усилителями стекловолоконные кабели осуществляют связь между континентами через моря и океаны.

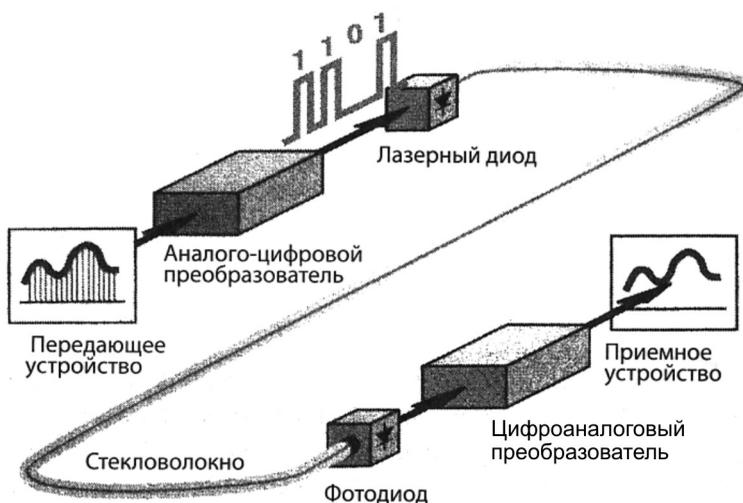


Рис. 10.16. Передача информации по стекловолоконкам

В 1988 году был проложен первый подводный стекловолоконный кабель для осуществления трансатлантической связи между восточным побережьем США и Европой. Проведенный в 1989 году транстихоокеанский стекловолоконный кабель был рассчитан на скорость передачи данных 280 Мбит/с. Эта подводная стекловолоконная линия содержала оптические усилители, благодаря которым удалось значительно снизить число требуемых оптоэлектронных преобразований и регенераций электрических сигналов.

Объединив в кабели 100 стекловолокон, можно одновременно включить свыше 1000 телевизионных каналов или, например, 1 млн телефонных линий, можно передавать информацию со скоростью до 2 Тбит/с. Подобные способы применения лазерной техники и современной оптики в информационных технологиях находятся в настоящее время в центре развития.

Лазерная обработка материалов

При использовании энергии лазера для обработки материалов плотность мощности и время облучения должны соответствовать выбранному способу обработки. При низкой плотности мощности и длительном времени облучения большой объем вещества нагревается теплопроводностью. При высокой плотности мощности и коротком времени облучения материал нагревается только в зоне падающего лазерного луча.

На рис. 10.17 приведены плотности мощности лазерного излучения в зависимости от времени облучения для разных процессов обработки. Границы диапазонов здесь невозможно обозначить точно, ибо они зависят от типа используемого лазера и обрабатываемого материала.

При *лазерной резке* листового металла, например, для корпусов машин материал доводится до расплавленного состояния, после чего жидкое вещество выталкивается струей газа. Преимущество применения лазерного луча состоит в том, что он может

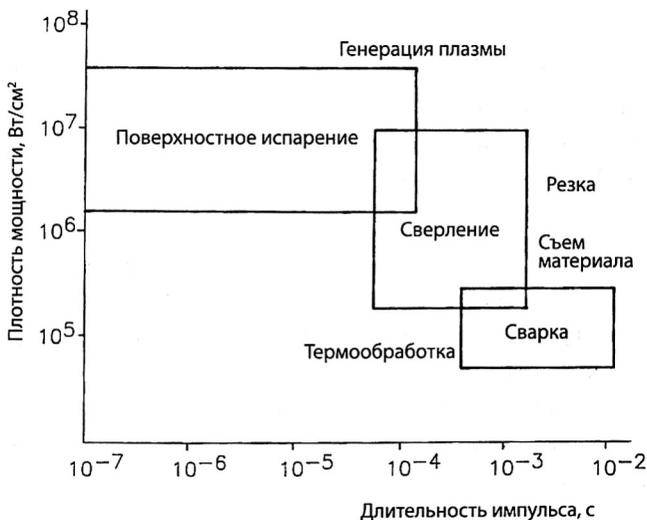


Рис. 10.17. Плотности мощности и длительности импульсов для обработки материалов с помощью лазеров

следовать по сложной пространственной траектории, позволяет практически без износа инструментов легко обрабатывать даже очень твердые материалы, которые обычно поддаются только линейному резанию с применением алмазных шлифовальных кругов.

При *лазерной сварке* стальных листов достигаются скорости порядка нескольких метров в минуту (рис. 10.18). В режиме теплопроводной сварки лазерный луч так проводится по поверхности материала, что температура кипения не достигается, но начинается едва заметное испарение. Под действием лазерного излучения расплавляются подлежащие соединению детали, приложенные друг к другу по сварному шву. Сварной шов при этом не очень глубок и примерно соответствует его 1,5-кратной ширине. Более сложными представляются операции по сварке с глубоким проплавлением. Здесь температура выше точки кипения, поэтому расплав перемещается под давлением пара в сварном шве. Расплав циркулирует и частично даже поднимается на поверхности материала, образуя валик.

Сварной шов отличается мелкозернистой кристаллической структурой с минимумом загрязнений. Нередко его прочность превышает прочность сваренного материала.

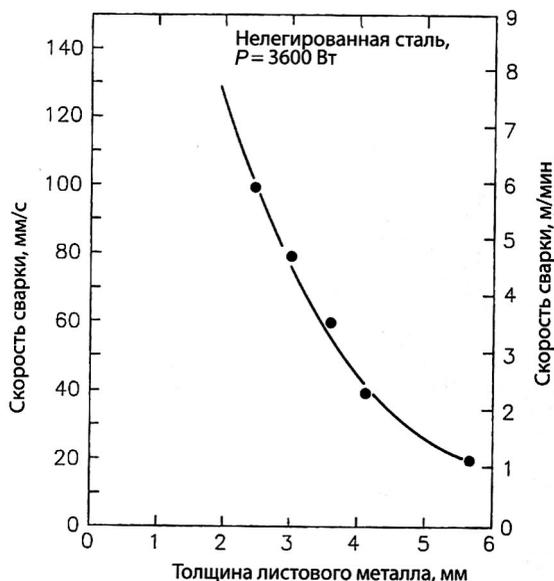


Рис. 10.18. Скорость сварки с применением лазера высокой энергии излучения

Лазерная сварка очень популярна и в электронной промышленности, где ее используют для соединения разнородных веществ.

Лазерной термообработке, закалке, легированию подвергают металл с целью повышения механических и химических свойств, снижения циклической усталости, а также улучшения внешнего вида изделия. Рис. 10.19 демонстрирует в поперечном разрезе поверхность, закаленную на ширину 4 мм лучом лазера. При этом было достигнуто существенное повышение ее прочности.

Новые свойства поверхности металла, недоступные традиционным методам диффузионного насыщения поверхности металла примесью, дает технология лазерного легирования. Поверхностное легирование началось в промышленном масштабе

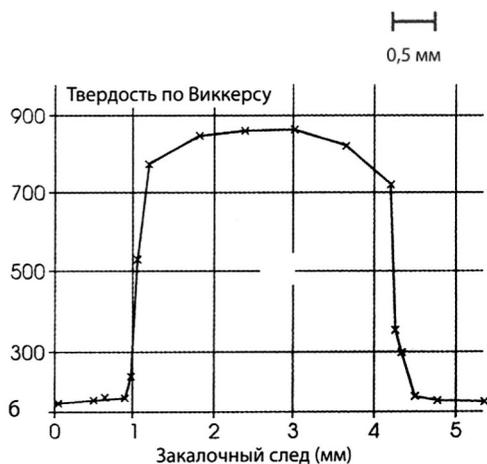
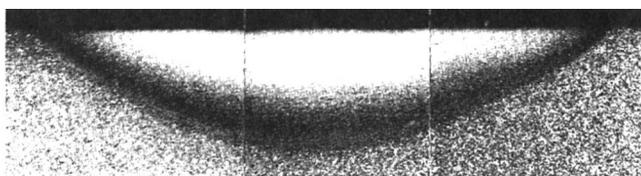


Рис. 10.19. Закалка поверхности лазерным излучением: вверху – разрез перпендикулярно поверхности разной структуры; внизу – характеристика закалки

только в последние годы именно благодаря применению лазеров. При этом на заготовку наносится порошкообразное вещество (бор или карбид), которое лазерным излучением сплавляется с поверхностью. В результате образуется новый сплав, который при быстром охлаждении отличается чрезвычайно мелкозернистой и прочной структурой. При кратковременном лазерном оплавлении поверхности металла возникающие интенсивные гидродинамические потоки термокапиллярной конвекции ускоряют процессы массопереноса по всей зоне оплавления. В итоге тонкая корочка поверхностно упрочненного металла при мягкой сердцевине повышает стойкость металла к истиранию и циклическим нагрузкам. Результаты моделирования этой технологии представлены на рис. 10.20.

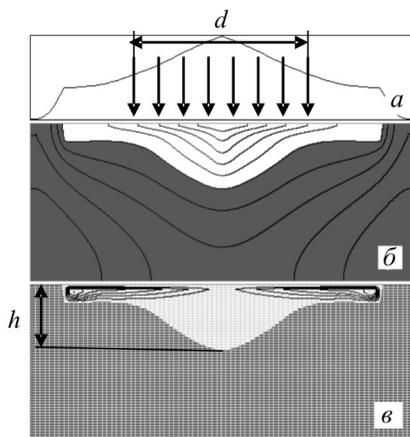


Рис. 10.20. Оплавление титана через 14 мс после начала импульсного лазерного воздействия: *a* – распределение температуры на поверхности; *б* – изотермы; *в* – траектории течения расплава. Энергия в импульсе – 8 Дж; плотность потока энергии – $3,3 \cdot 10^8$ Вт/м²; длительность импульса – $6 \cdot 10^{-3}$ с; диаметр лазерного пучка $d \approx 1,2$ мм; глубина проплавления $h \approx 0,25$ мм

Лазерные способы скрабирования и разделения полупроводниковых чипов, производимых в большом количестве на кремниевой подложке или кристаллической пластине диаметром в несколько сантиметров, а также пайка и подгонка резисторов путем съема материала с помощью лазеров уже давно утвердились в области производства микроэлектроники.

Измерение расстояний

Измерения расстояний с применением лазерного луча сводятся к измерениям времени. Благодаря лазерному излучению любую длину можно определить как временной интервал, необходимый световому лучу для прохождения измеряемого участка.

Важную роль в развитии измерительной техники играет генерация сверхкоротких световых импульсов, которые позволяют исследовать с высоким временным разрешением быстро проте-

кающие химические реакции. Как электронный микроскоп в пространственном диапазоне, лазер обеспечивает во временном диапазоне возможность изучения элементарных структур. С помощью лазеров создаются так называемые «микроскопы времени», эффективность которых почти на 12 порядков превосходит традиционные «лупы времени» (устройства растяжения сигнала во времени), широко применяемые в кино- и телевизионной технике. Время пробега коротких импульсов находит свое применение в геодезии для точного и удобного определения расстояний.

Лазерные лучи распространяются с отличной фокусировкой и очень хорошей прямолинейностью, поэтому могут служить в качестве направленных лучей при строительстве дорог, каналов, туннелей, зданий, а также для точной центровки при сборке механизмов, установке оборудования.

Применение лазеров в медицине

Лазеры успешно используются в медицине – как в диагностике, так и для терапевтических целей.

Тепловое действие лазерного излучения на живую ткань сводится к повышению ее температуры. При температуре примерно 60 °С происходит свертывание белка, при 100 °С из ткани испаряется вода, а при дальнейшем повышении температуры ткань обугливается. Эти этапы воздействия на биологический материал используются в хирургии. Обладая высокой точностью, лазерная хирургия дает преимущество бесконтактного асептического вмешательства и возможность – в случае сильно кровоточащих тканей – почти бескровного разрезания благодаря закупорке сосудов на основе коагуляции.

Наряду с тепловым воздействием лазерного излучения на ткани существуют (с учетом длительности облучения и спектральной плотности мощности) и другие механизмы, широко применяемые в медицине. Например, для фотоабляции (отслойки ткани) требуются короткие импульсы высокой мощности. Эффект наступает, когда глубина проникновения луча в ткань

находится в пределах микрометров, а длительность импульсов столь коротка, что никакой значимой теплопроводности в этот момент не отмечается. В результате пораженная ткань отслаивается под действием коротких импульсов без термического повреждения окружающей здоровой ткани. Фотоабляция широко применяется в фоторефрактивной хирургии роговицы глаза. С помощью лазера на эксимере удастся путем соответствующего съема так изменить кривизну роговицы, что пациент сможет обходиться без очков. При дальнейшем повышении мощности и укорочении длительности импульсов в ткани происходит оптическая перфорация. В офтальмологии такой эффект применяют для разрушения мутной пленки вторичной катаракты. С помощью лазерного луча хирург проникает во внутриглазную область через хрусталик глаза, не повреждая его, и производит необходимую операцию. Восстановление сетчатки глаза с помощью этой техники стало обычным делом и применяется во многих клиниках.

На рис. 10.21 показано удаление отложений в кровеносных сосудах, так называемое обызвествление артерий (ангиопластика). При этом лазерный луч с помощью катетера и стекловолокна вводится в артерию и производит съем отложений.

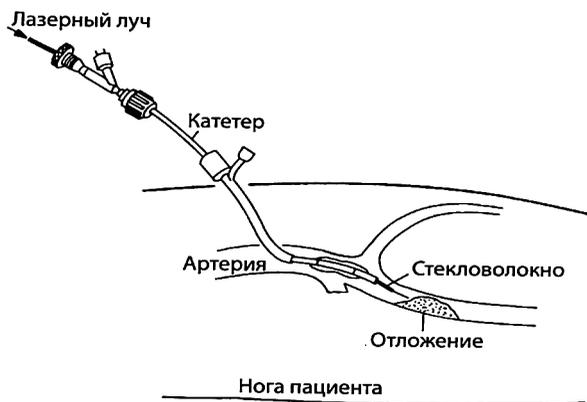


Рис. 10.21. Удаление отложений в артериях с помощью лазерного луча

Дальнейший прогресс применения лазеров в медицине связывают с развитием полупроводниковых технологий, снижающих себестоимость лазерной техники.

Термоядерная реакция с применением лазеров

С 50-х годов в лабораториях разных стран предпринимаются попытки использовать управляемые процессы ядерного синтеза на Земле в целях получения энергии. Чтобы начать процессы синтеза, необходимо в достаточной степени нагреть газообразный водород, который переходит в плазменное состояние. Для этого требуются температуры в сотни миллионов градусов Кельвина на период всего нескольких секунд. Ядра атомов водорода должны обладать достаточно мощной энергией, чтобы произошло их слияние. При столь высоких температурах требуется специальное оборудование, чтобы удержать плазму: в проводимых до сих пор экспериментах для этой цели использовались магнитные поля, в которых устойчивого удержания плазмы осуществить не удалось.

Альтернативный путь к решению этой проблемы связан с применением лазера. Основная идея при этом заключается в том, чтобы из изотопов водорода, дейтерия и трития изготовить маленький твердый шарик и нагреть его путем кратковременного облучения светом лазера. Энергия лазера должна быть достаточно большой, чтобы достичь требуемых температур ядерного синтеза. При этом используется короткий импульс возбуждения, исключаящий разброс плазмы при начавшейся ядерной реакции. Таким образом, плазма при таком процессе удерживается не под действием внешнего поля, а в результате своей собственной инерционности. Прежде чем частицы плазмы разлетятся во все стороны, уже должны произойти ядерные реакции. Это так называемое инерциальное удержание плазмы требует наличия экстремальных лазеров. Здесь нужна энергия лазерного излучения выше 106 Дж, причем такая энергия должна за сверхкороткое время, около 10–9 с, войти в водородную мишень с диаметром не более 1 мм.

Схема новой лазерной установки для термоядерных экспериментов показана на рис. 10.22, данные лазера представлены в табл. 10.1. Этот лазер размещен в многоэтажном здании длиной 200 м. Там же находятся 192 цепи лазерных систем, включенных параллельно и приводимых в действие единственным лазерным генератором, обеспечивающим синхронизацию разных лазерных усилителей.

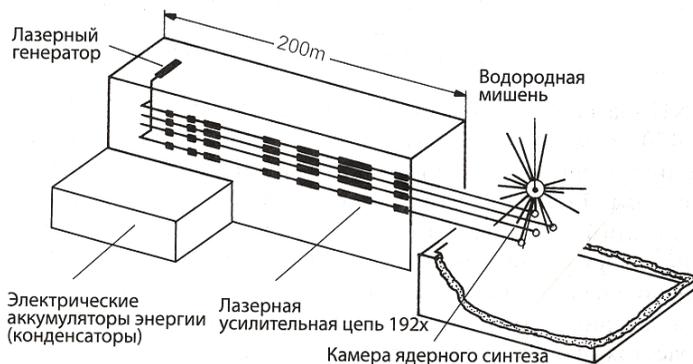


Рис. 10.22. Конструкция лазера для получения ядерного синтеза в лаборатории *Lawrence-Livermore*, США, выходная энергия 1,8 МДж. Эта лазерная система содержит в общей сложности 192 параллельных усилительных цепи

Таблица 10.1

Рабочие характеристики лазера для термоядерного синтеза

Выходная энергия в импульсе	1,8 МДж = $1,8 \cdot 10^6$ Дж = = 450 г взрывчатого вещества в тротиловом эквиваленте
Длительность импульсов	3 нс = 0,000 000 003 с
Пиковая мощность	500 ТВт = $5 \cdot 10^{14}$ Вт = 500 000 электростанций
Длины волн	350 нм (УФ-область спектра, утроение частоты 1060 нм)
Материал	Стекло с неодимом
Затраты	Несколько миллиардов \$ США, постройка 1998–2004 гг.

При успешном проведении экспериментов разработчики планируют к 2030 году запуск первой опытной электростанции на термоядерном синтезе.

Лазерные принтеры

Параллельно с успехами в электронной обработке информации в последние десятилетия достигнут значительный прогресс в технике вывода ее на печать. Классические печатающие машинки с литерными рычагами уступили место матричным, термографическим, струйным и электрофотографическим печатающим устройствам. Эти устройства первоначально печатали буквы последовательно друг за другом, затем построчно и, наконец, постранично, что позволило резко ускорить процесс печати и повысить его качество.

Среди постранично печатающих устройств огромную роль играет электрофотографическая система, которую не совсем правильно называют лазерным принтером. Дело в том, что наряду с лазерами в качестве источников света здесь используются еще светоизлучающие диоды и галогенные лампы, управляемые жидкокристаллической шиной.

Функция лазерного принтера наглядно представлена на рис. 20.23. Печать осуществляется путем передачи оттиска на обладающий фотопроводимостью барабан, сохраняющий в темноте отрицательные электрические заряды на своей поверхности. При вводе света поверхностный заряд исчезает. Собственно процесс печати осуществляется в несколько этапов.

На первом этапе отрицательные заряды из электрического разряда попадают на поверхность барабана (рис. 10.23, *а*).

Затем барабан в результате засветки лазерным лучом или экспонирования посредством другого источника света разряжается с образованием скрытого, невидимого изображения подлежащей выводу на печать страницы (рис. 10.23, *б*). Для этого лазерный луч перемещается параллельно оси барабана.

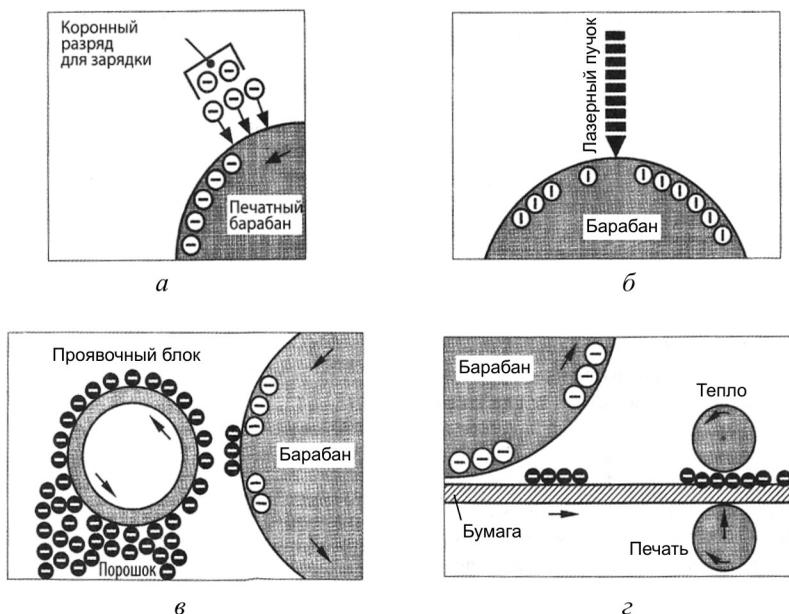


Рис. 10.23. Функциональная схема лазерного принтера: а – зарядка печатного барабана; б – экспонирование лазерным лучом, формирование потенциального рельефа; в – проявление, формирование изображения частицами порошка; г – процесс печати на бумаге

На следующем этапе в результате вращения барабана происходит запись по всей его поверхности, разряженные зоны на барабане движутся мимо проявочного блока, забирая от него отрицательно заряженные частицы порошка для электростатической печати (рис. 10.23, в).

И, наконец, бумага вступает в контакт с вращающимся барабаном и запечатывается (рис. 10.23, г). Частицы упомянутого выше специального порошка в результате нагревания и давления прочно соединяются с бумагой, барабан очищается от избытка порошка. Под действием лазерного излучения он равномерно разряжается, и после полного оборота барабана процесс печати вновь начинается с первого шага.

Для лазерного принтера требуется электронное управление, обеспечивающее в растровой сетке разрешение для запечатываемой страницы на уровне 300 dpi (точек на дюйм). Таким образом, одна сторона изображения будет представлена в 8 миллионах точек. В полиграфической промышленности при фотографическом изготовлении печатных матриц достигается разрешение выше 1200 dpi.

Скорость лазерного принтера определяется, прежде всего, предварительной электронной обработкой. Само печатающее устройство действует чрезвычайно быстро. Индивидуальные или офисные принтеры могут работать со скоростью 10–20 страниц в минуту, а в полиграфической промышленности минутная скорость печати достигает 200 страниц. Лазерные принтеры позволяют также изготавливать цветные копии, для чего последовательно друг за другом включаются сразу три барабана с порошком красного, зеленого и синего цветов.

Немногочисленные примеры показывают, что лазеры успешно используются в разных сферах – научной, технической, медицинской. Эта область современной фотоники успешно развивается. В перспективе можно ожидать более широкого и разнообразного применения лазерных устройств.

Полупроводниковые лазеры, как и светоизлучающие диоды (СИД), представляют собой оптоэлектронные приборы, в которых осуществляется преобразование электрической энергии в оптическую (световую). В основе этого процесса лежит излучение света, обусловленное электронным переходом из зоны проводимости полупроводника в валентную зону. Для возбуждения электронов в зону проводимости, или, как говорят, для накачки, используется инжекция носителей. С этой целью обычно формируют p - n -переход, для чего в n -область вводят больше донорной примеси, а в p -область больше акцепторной. С помощью инжекции обеспечивается создание неравновесных носителей заряда, что обеспечивает генерацию оптического излучения в p - n -переходе полупроводника. Получение оптической энергии на выходе

оптоэлектронных устройств оказывается очень простым: для этого достаточно подать напряжение на вход прибора и обеспечить протекание по нему тока. Путем изменения тока инжекции можно менять оптическую энергию на выходе прибора, т.е. простыми средствами осуществлять оптическую модуляцию. Этот способ называется прямой (непосредственной) модуляцией. Простота осуществления оптической модуляции является одной из причин использования полупроводниковых лазеров и СИД в системах оптической связи.

Типичные примеры зависимости между током и мощностью оптического излучения на выходе оптоэлектронного прибора приведены на рис. 10.24. Как видно из рис. 10.24, *а*, полупроводниковый лазер является «пороговым прибором»: если увеличивать ток инжекции, то при превышении некоторого порогового значения $I_{\text{п}}$ возникает резко линейное увеличение оптического выхода лазера. В окрестности порогового значения тока наблюдается качественное изменение процесса: медленный рост вынужденного излучения переходит в режим генерации излучения. При $I < I_{\text{п}}$ излучение лазера подобно свету обычной электрической лампы и представляет собой сумму случайных световых потоков или некогерентный свет. СИД конструируют

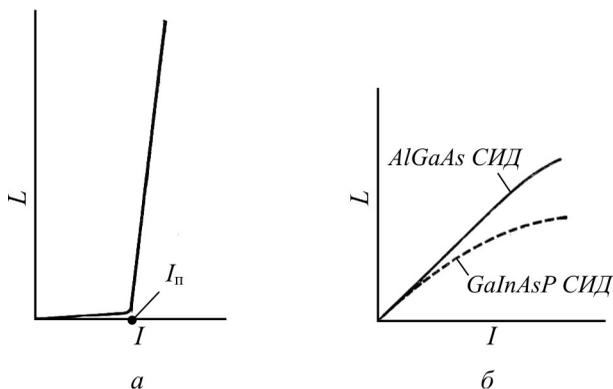


Рис. 10.24. Зависимость мощности оптического излучения L на выходе полупроводникового лазера (*а*) и СИД (*б*) от тока

таким образом, чтобы в них не возникал режим генерации. Поэтому в них, как показано на рис. 10.24, б, по мере увеличения тока инжекции происходит монотонное нарастание оптического выхода. При этом насыщение оптического выхода СИД связано с выделением тепла, т.е. по мере увеличения концентрации носителей в области светового излучения падает светоотдача диода.

Вопросы для самоконтроля

1. Назовите основные обязательные элементы любого лазера.
2. Какие энергетические переходы происходят при спонтанном, вынужденном излучении и поглощении в атоме.
3. При каких условиях происходит амплитудное и фазовое самовозбуждения лазера?
4. Как создается инверсия населенностей в трех- и четырехуровневом лазере?
5. Чем обусловлены основные свойства лазерных пучков: монохроматичность, когерентность, направленность, яркость, возможность генерации сверхкоротких импульсов света.
6. Классификация лазеров по физическому состоянию активной среды и по длине волны генерируемого излучения.
7. Особенности работы и сфера применения лазеров на углекислом газе.
8. Каковы особенности работы полупроводниковых лазеров на квантовых структурах?
9. Приведите примеры практического использования лазеров при обработке материалов.
10. Каков принцип работы лазерного принтера?
11. В чем состоит отличие полупроводниковых приборов – лазеров и СИД, использующих $p-n$ -переход для получения оптического излучения?

11. ОПТИЧЕСКИЕ ВОЛОКНА

11.1. Общие сведения

Первые волоконные световоды (ВС) или оптические волокна (ОВ) состояли из однородного по составу оптически прозрачного цилиндрического стержня (сердцевины), окруженного оптически прозрачным материалом (оболочкой). В простейшем

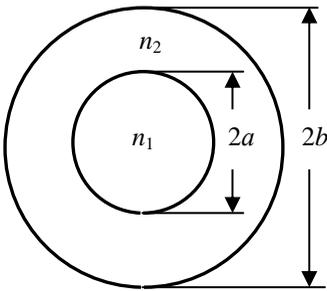


Рис. 11.1. Поперечное сечение ОВ

варианте ОВ представляет собой тонкую нить из оптически прозрачного материала, сердцевина которой диаметром $2a$ имеет показатель преломления n_1 , а оболочка диаметром $2b$ имеет показатель преломления $n_2 < n_1$ (рис. 11.1). Поскольку показатель преломления сердцевины выше, чем оболочки, направленное распространение света происходит благодаря многократным полным внутренним отражениям от границы между ними.

Величины $2a$ и $n_1 - n_2$ определяют число типов волн (мод), которые могут распространяться по ОВ при заданной длине волны света λ . Выбирая $2a$ и $\Delta n = n_1 - n_2$ достаточно малыми, можно добиться, чтобы ОВ работал в *одномодовом* режиме. Одномодовые ОВ, в которых $2a < 10$ мкм, а Δn составляет несколько десятых долей процента, применяются в межконтинентальных линиях связи и других магистральных линиях, где требуется чрезвычайно высокое качество передаваемой информации. С увеличением диаметра сердцевины до десятков и сотен мкм, а Δn до 1–2 % ОВ работает в *многомодовом* режиме. Из-за потерь, связанных с межмодовым взаимодействием, многомодовые волокна применяются в основном для передачи мощности или для связи, но на короткие расстояния. Одно- и многомодо-

вые оптические волокна мирно сосуществуют: одномодовые волокна используются в основном для магистральных коммуникаций, а многомодовые – для коротких линий (в домах и офисах): их легче монтировать, они более надежны в эксплуатации.

На рис. 11.2 показан случай ввода гауссова пучка света в световод с отклонением от его центральной оси. Входная волна, возбуждает сначала основную 0-ю моду, затем 1-ю моду, после чего поочередно следуют моды более высоких порядков. Все моды должны распространяться одновременно. Та часть волноводных мод, которая не укладывается в световод, образует вытекающую моду и излучается за пределами световода. Эта часть света образует потери при вводе излучения в волновод.

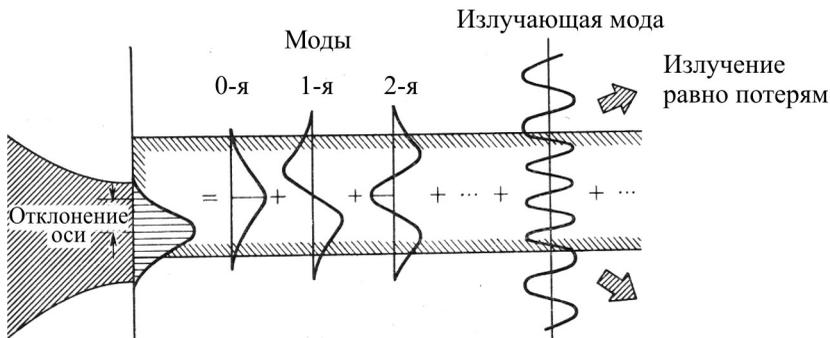


Рис. 11.2. Развитие собственных мод, соответствующих одной падающей волне. Показаны три моды: 0, 1 и 2-го порядка с близким распределением интенсивности, а также моды высшего порядка и мода излучения

ОВ применяются в системах оптической связи, в датчиках, в оптических компьютерах, для канализации мощного лазерного излучения. Важнейшими характеристиками ОВ являются *оптические потери, дисперсия групповой скорости, оптическая нелинейность и механическая прочность.*

Оптические потери обусловлены поглощением света по длине ОВ и оцениваются в децибелах (дБ/км). Если на входе ли-

нии протяженностью L мощность сигнала равна P_0 , а на выходе P_1 , потери α на 1 км длины линии равны $(10/L)\lg(P_0/P_1)$.

Спектры потерь в изогнутых и прямых волокнах могут существенно различаться. В прямых волокнах потери ограничиваются в основном рэлеевским рассеянием (происходит на оптических неоднородностях, размеры которых значительно меньше длины волны) и уменьшаются с увеличением длины волны $\sim \lambda^{-4}$, например, на $\lambda = 1550$ нм потери всегда меньше, чем на $\lambda = 1310$ нм.

В то же время в изогнутых волокнах потери могут, наоборот, увеличиваться с увеличением длины волны. Изгибы световода, как и различные неровности границы раздела, приводят к выходу излучающей моды за его пределы.

Потери при изгибании волокна возникают по двум причинам. Во-первых, потери возникают в месте соединения прямого и изогнутого волокна. Обусловлены они тем, что в изогнутом волокне центр модового пятна смещен относительно оси волокна на некую величину d , зависящую от радиуса изгиба волокна (рис. 11.3). В результате модовые пятна прямого и изогнутого волокна в месте их соединения оказываются смещенными друг

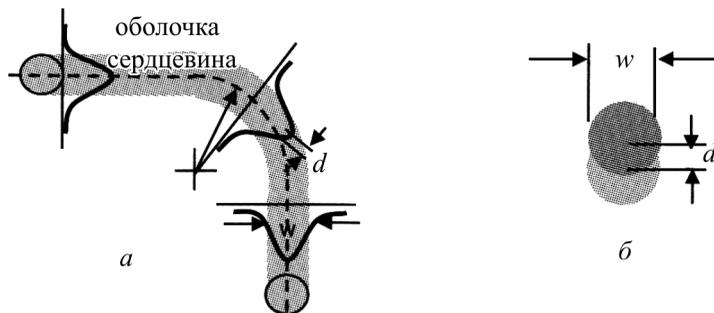


Рис. 11.3. Схема, поясняющая причину возникновения потерь в месте соединения прямого и изогнутого волокна: a – в изогнутом волокне центр модового пятна смещен относительно оси волокна на величину d ; $б$ – в месте соединения «прямого» и изогнутого волокна их модовые пятна смещены друг относительно друга на величину d

относительно друга также на величину d . Поэтому только часть мощности моды «прямого» волокна (диаметром w) передается моде изогнутого волокна, остальная же мощность преобразуется в оболочечные моды и теряется.

Во-вторых, мощность теряется и непосредственно в изогнутом волокне. Происходит это из-за того, что в изогнутом волокне периферийная часть моды распространяется со скоростью больше скорости света в среде (в оболочке). Эта часть моды излучается в оболочку волокна и в конечном счете теряется (рис. 11.4). Величина этих потерь тем больше, чем больше число витков волокна и чем меньше радиус изгиба волокна.

Дисперсия (разброс, отклонение) – это эффект зависимости фазовой скорости световых колебаний от длины волны. Она приводит к искажению формы и длительности светового импульса. Поскольку при цифровой передаче информация кодируется последовательностью импульсов, то чем они короче и больше их число в единицу времени, тем больше информации можно передать. Если же импульсы расплываются настолько, что приемник не может их различить, то приходится понижать плотность их следования, при этом уменьшается и пропускная способность канала. Различают *материальную* и *модовую* дисперсию, обусловленную разными факторами. Первая зависит от состава материала, а вторая – от геометрии волокна.

Материальная дисперсия возникает из-за того, что спектр оптического сигнала имеет конечную ширину и разные спектральные компоненты сигнала движутся в волокне с разной ско-

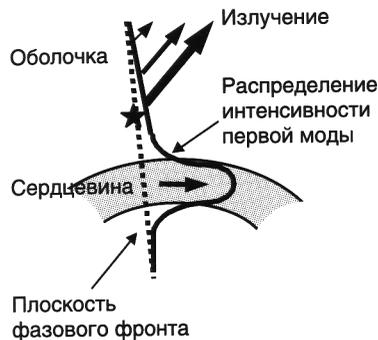


Рис. 11.4. Схема, поясняющая возникновение потерь в изогнутом волокне. Звездой обозначена точка, где скорость распространения фазового фронта превысила скорость света в среде

ростью (рис. 11.5). В результате световой импульс после прохождения через дисперсионную среду уширяется.

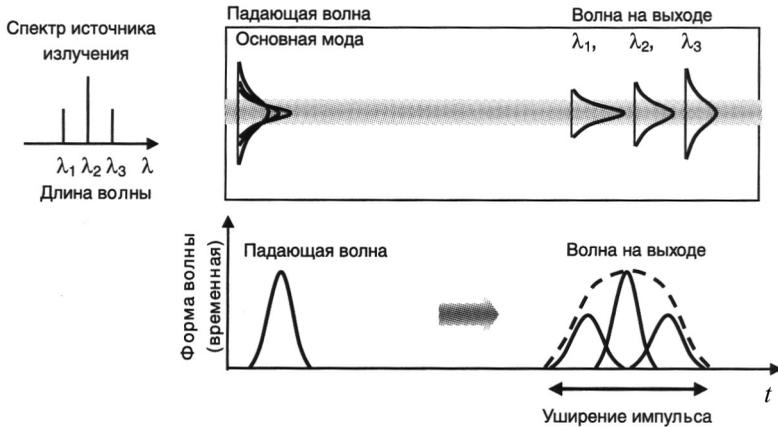


Рис. 11.5. Материальная дисперсия в одномодовом волокне

Появление модовой дисперсии обусловлено изменением скорости распространения волны из-за зависимости показателя преломления среды от длины волны. В волокне волна распространяется в двух средах – частично в сердцевине, а частично – в кварцевой оболочке, и для нее показатель преломления принимает некое среднее значение между значением показателя преломления сердцевины и кварцевой оболочки (рис. 11.6).

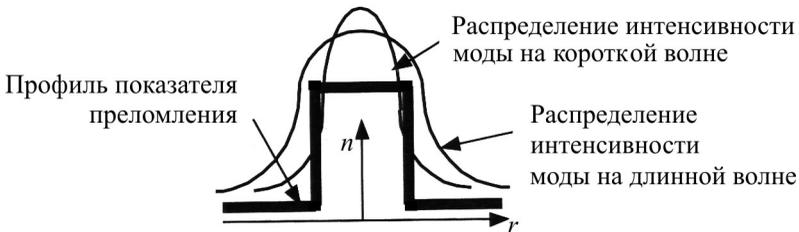


Рис. 11.6. Модовая дисперсия возникает из-за того, что усредненный по диаметру моды показатель преломления изменяется при изменении длины волны

Этот средний показатель преломления может изменяться по двум причинам. Во-первых, из-за того, что показатели преломления сердцевины и кварцевой оболочки примерно одинаково зависят от длины волны. Во-вторых, потому, что при изменении длины волны, меняется глубина проникновения поля в кварцевую оболочку и, соответственно, меняется среднее значение показателя преломления. Это чисто межмодовый эффект, и поэтому возникающую из-за него дисперсию называют межмодовой (волноводной).

Волноводная дисперсия может возникнуть и в одномодовом волокне при поляризации света и двулучепреломлении. Она наводится в номинально круглом волокне при его изготовлении из-за неизбежного появления небольшой эллиптичности сердцевины и внутренних напряжений, не обладающих аксиальной симметрией (рис. 11.7). Поскольку наведенные в волокне напряжения не имеют выделенного направления, величина и азимут двулучепреломления изменяется случайным образом вдоль оси волокна.



Рис. 11.7. Причины возникновения поляризации (двулучепреломления) в оптических волокнах

Как правило, превалирует материальная дисперсия, а модовая и волноводная дисперсия начинает проявляться при высоких скоростях передачи и расстоянии между ретрансляторами в несколько сот километров.

Представим свет, распространяющийся в одномодовом волокне, в виде суммы двух ортогональных поляризационных мод. Возбужденные быстрая и медленная поляризационные моды распространяются вдоль волокна, не обмениваясь при этом мощностью. Это приводит, как видно из рис. 11.8, к появлению

разности фазовых запаздываний поляризационных мод $\Delta\tau$ и, соответственно, к уширению импульсов.



Рис. 11.8. Уширение импульсов в поляризованном волокне

Началом современного этапа разработки волоконно-оптических систем связи принято считать 1970 год, когда впервые были изготовлены ВС из кварцевого стекла с потерями порядка 20 дБ/км. Первая волоконно-оптическая система связи была создана в 1970 году, а уже к 1979 году системы, работающие в диапазоне длин волн 0,82–0,85 мкм, прошли стадию экспериментальных исследований и начали вводиться в эксплуатацию на междугородных и крупных городских магистралях. На этих длинах волн потери в световодах уменьшились до 2–3 дБ/км. Позднее оказалось, что диапазоны (окна) около 1,3 и 1,5 мкм в стеклянных волноводах обладают большими преимуществами. На длине волны 1,3 мкм хроматическая дисперсия кварцевых стекол, определяющая максимальную скорость передачи информации, вообще отсутствует. При дальнейших исследованиях выяснилось, что абсолютный минимум оптических потерь лежит на длине волны 1,5 мкм. И постепенно, особенно для очень длинных линий, межконтинентальных, системы связи стали конструироваться именно на эту длину волны, позволяющую передавать информацию на большие расстояния без ретрансляторов.

Спектральная область современных кварцевых световодов, в которой возможна передача оптических сигналов с относительно низкими потерями (до 0,2 дБ/км), очень широка (рис. 11.9). Сейчас же в основном используются только два участка спектра:

в районе 1,3 и 1,5 мкм. Возможность использования всей этой области (и, соответственно, существенного увеличения пропускной способности) связана с решением проблемы широкополосного усиления в ближайшей перспективе.

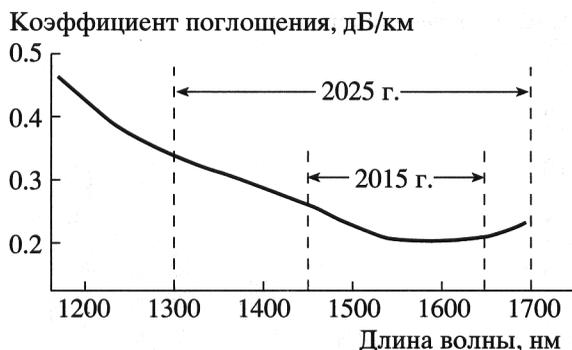


Рис. 11.9. Спектр потерь современных волоконных световодов на основе кварцевого стекла и прогноз расширения области для передачи информации в 2015 и 2025 гг. В настоящее время для передачи информации используется сравнительно узкая спектральная область 1530–1610 нм

Кварцевое стекло является по уровню прочности, стабильности, распространенности в природе очень хорошим материалом. Достигнутые в кварцевых волокнах потери соизмеримы с потерями в области стыковки, изгиба и других деформаций при монтаже волокон. Более низкие потери порядка одной сотой дБ/км можно получить во фторидных стеклах, но технология изготовления этих стекол крайне сложна.

Принципиальным преимуществом ОВ при передаче информации является не только большая широкополосность при низких оптических потерях, но и высокая скорость передачи информации. Медные провода в электрических линиях связи позволяют передавать информацию со скоростью до 2–10 Мбит/с. При передаче информации со скоростью 10 Гбит/с электроника уже не работает. В отличие от электрических линий связи, где потери зави-

сят от частоты передаваемых сигналов, в оптических волокнах потери не зависят от скорости передачи данных. Поэтому при низкой скорости передачи предельно допустимое расстояние между ретрансляторами ограничивается потерями в волокнах, а при высоких скоростях – дисперсией.

В волоконных линиях дальней связи, построенных в России, скорость передачи, как правило, не превышает 2,5 Гбит/с, без использования оптических усилителей. Поэтому в них расстояние между ретрансляторами (~100 км) ограничивается потерями в волокне. В этих линиях используются стандартные одномодовые волокна. Потери в лучших промышленных образцах таких волокон на длине волны 1550 нм составляют 0,18...0,19 дБ/км.

В большинстве зарубежных линий дальней связи используются *оптические усилители*, и в этих линиях расстояние между ретрансляторами уже не лимитируется потерями в волокнах. Так, в наземных линиях связи это расстояние может достигать 1000 км, в подводных линиях – и 10 000 км. Скорость передачи данных в большинстве таких линий составляет до 40 Гбит/с.

По одному волокну можно передавать гигантское количество информации – около 1 Тбит/с. Но на одной длине волны передавать потоки более 10 Гбит/с практически нецелесообразно, оказываются существенными ограничения, возникающие из-за дисперсии оптических волокон. Значительно проще передавать 2,5 Гбит/с на одной длине волны, но использовать большое число несущих частот (длин волн). Реально используется около ста длин волн. В одном из экспериментов вводили 132 длины волны в световод и передавали по 20 Гбит/с на каждой, то есть скорость передачи информации получалось более 2 Тбит/с. Реализация *технологии спектрального уплотнения* представлена на рис. 11.10. На длинных морских линиях начали использовать усилители, накачиваемые излучением мощного эрбиевого лазера. Оптические усилители – это очень важный компонент для спектрального уплотнения каналов, поскольку они пропускают и усиливают сразу все используемые длины волн без преобразования в электронную форму.

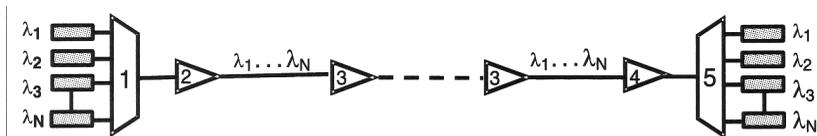


Рис. 11.10. Принцип работы системы связи со спектральным уплотнением по длинам волн: 1 – мультиплексор; 2 – оптический усилитель мощности; 3 – линейные оптические усилители; 4 – оптический предусилитель; 5 – демультиплексор

На рис. 11.11 представлен один из вариантов оптического кабеля связи, содержащего несколько ОВ.

Системы связи, имеющие такие огромные скорости передачи информации, до недавнего времени оставались недогруженными. Складывалось впечатление, что полоса пропускания волоконных световодов бесконечна и насыщение пропускной способности наступит нескоро. Однако в последние годы глобальный поток информации начал резко возрастать. Этот, на первый взгляд, несколько неожиданный результат объясняется, во-первых, растущим объемом услуг, предоставляемым в настоящее время Интернетом, во-вторых, низкой стоимостью производства и установки волоконно-оптического телекоммуникационного оборудования. Эти два обстоятельства привели к тому, что массовое распространение получил широкополосный (волоконный) доступ населения к информационным услугам, то есть к Интернету, причём распространение этой услуги в последнее время приобрело взрывной характер (рис. 11.12).

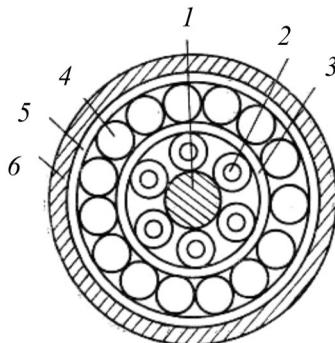


Рис. 11.11. Подводный оптический кабель связи с однослойным броневым покровом: 1 – силовой элемент; 2 – ОВ; 3 – алюминиевая трубка; 4 – стальная броня; 5 – медное покрытие; 6 – полиэтиленовая оболочка



Рис. 11.12. Число потребителей услуги «Волоконный световод в каждый дом» в 2004–2008 гг.

Среди стран – потребителей услуги «Волоконный световод в каждый дом» в настоящее время всех опережает Япония, в которой число подписчиков этой услуги достигло 15 млн при численности населения Японии 128 млн человек.

В связи с прогнозом увеличения глобального потока информации пропускная способность трансокеанических волоконно-оптических систем связи должна обеспечивать ее передачу со скоростью около 100 Тбит/с к 2025 г. Достижение указанных скоростей передачи информации возможно за счёт увеличения числа спектральных каналов в одном световоде. Однако рост числа каналов и расширение спектральной области, где оптические потери световодов больше (см. рис. 11.9), приводят к увеличению мощности сигналов, вводимых в волоконный световод. А это с неизбежностью влечёт за собой нелинейное взаимодействие каналов. Разработка волоконно-оптических систем связи нового поколения потребует проведения обширных фундаментальных исследований.

Использование оптических кабелей увеличивает пропускную способность и надежность волоконно-оптических систем, способствует экономии дефицитных цветных металлов (медь, свинец). Внешний диаметр кабеля обычно не превышает 10 мм, что особенно важно при его прокладке в существующей кабельной канализации сетей ГТС. Наконец, невосприимчивость оптического кабеля к любым внешним электромагнитным помехам позволяет прокладывать его в непосредственной близости от мощных радиопередатчиков, энергетических установок и других линий связи.

Вместе с тем на пути широкого внедрения в практику оптических кабелей связи имеется еще немало трудностей и проблем. В частности, стоимость оптических кабелей в настоящее время еще весьма велика, а параметры передачи значительно хуже, чем у экспериментальных образцов, и нестабильны, что связано с несовершенством технологии промышленного производства. Определенные трудности вызываются специфическими особенностями прокладки, монтажа и эксплуатации волоконно-оптических линий связи, а также сложностью измерений характеристик и испытаний оптических кабелей.

11.2. Типы оптических волокон

Основным элементом оптического кабеля является оптический волновод – круглый стержень из оптически прозрачного диэлектрика, структура которого обеспечивает распространение вдоль него световых сигналов. Оптические волноводы из-за малых размеров поперечного сечения обычно называют *волоконными световодами* (ВС) или *оптическими волокнами* (ОВ). Первый термин обычно применяется при исследовании вопросов передачи информации с помощью законов оптики, второму отдается предпочтение при рассмотрении конструктивных и технологических особенностей оптических кабелей.

В тех случаях, когда длина волны излучения λ гораздо меньше размеров поперечного сечения ВС, для описания процес-

са распространения света можно пользоваться приближенными методами геометрической (лучевой) оптики. Если же указанное условие не выполняется или законы геометрической оптики не позволяют получить правильные результаты, для исследования волновых явлений в ВС необходимо решать уравнения Максвелла, что требует применения аппарата математической физики. Это сложная задача даже при рассмотрении простейших ВС. Рассмотрим процесс распространения световых волн методами геометрической оптики.

Если световая волна из оптически более плотной среды падает на границу раздела с оптически менее плотной средой ($n_1 > n_2$), то согласно закону Снеллиуса всегда существует критический (предельный) угол падения

$$\varphi_{\text{кр}} = \arcsin \frac{n_2}{n_1}, \quad (11.1)$$

при котором прошедшая волна распространяется вдоль границы раздела сред ($\varphi_{\text{пр}} = \pi/2$). При всех углах падения $\varphi > \varphi_{\text{кр}}$ преломленная волна отсутствует и свет полностью отражается от поверхности оптически менее плотной среды. Это явление называется *полным внутренним отражением*. Отраженная волна при этом приобретает фазовый сдвиг, зависящий от угла падения. При полном внутреннем отражении электромагнитное поле световой волны частично заходит во вторую среду, где экспоненциально убывает.

Круглый диэлектрический стержень

Простейшим ВС является круглый диэлектрический стержень, показатель преломления которого n_1 выше, чем у окружающей среды n_0 (рис. 11.13).

Пучок световых лучей реального источника конечных размеров преобразуется на торце ВС в два типа лучей: *меридио-*

нальные, которые пересекают ось световода, и *косые*, которые эту ось не пересекают. Часть меридиональных и косых лучей, испытывая полное внутреннее отражение в местах падения на границу стержень – окружающая среда, распространяется вдоль ВС по зигзагообразным путям. Световые волны, которые изобращаются этими лучами, многократно отражаясь от границы, накладываются сами на себя и образуют *направляемые волны (моды)*. Поле направляемых мод частично проникает в окружающую среду, где экспоненциально убывает. Остальные лучи, которые падают на границу под углами $\varphi < \varphi_{кр}$, уходят из стержня в окружающую среду. Так образуются волны (моды) излучения – рассеянный свет.

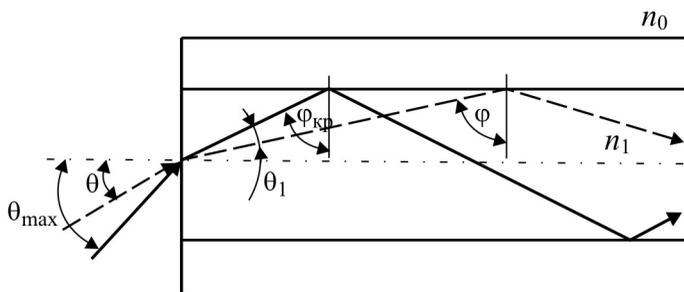


Рис. 11.13. Прохождение меридиональных лучей по диэлектрическому стержню

Рассмотрим ход меридиональных лучей, падающих на входной торец ВС и распространяющихся затем вдоль стержня (рис. 11.13). Луч, падающий под углом θ , преломляется под углом θ_1 , который в соответствии с законом преломления определяется выражением

$$n_0 \sin \theta = n_1 \sin \theta_1. \quad (11.2)$$

Преломленный луч падает на боковую поверхность под углом $\varphi = \pi/2 - \theta_1$. Для полного внутреннего отражения лучей от боковой поверхности необходимо выполнение условия $\varphi \geq \varphi_{кр}$,

где $\varphi_{\text{кр}}$ – критический угол падения, определяемый выражением (11.1). Следовательно, $\sin \varphi = \cos \theta_1 \geq n_0/n_1$. Из (11.2)

$$\cos \theta_1 = \sqrt{1 - \sin^2 \theta_1} = \sqrt{1 - \left(\frac{n_0}{n_1}\right)^2 \sin^2 \theta}.$$

Таким образом,

$$\sqrt{1 - \left(\frac{n_0}{n_1}\right)^2 \sin^2 \theta} \geq \frac{n_0}{n_1} \quad \text{или} \quad \left(\frac{n_1}{n_0}\right)^2 - \sin^2 \theta \geq 1.$$

Однако $\sin^2 \theta \leq 1$, т.е. $(n_1/n_0)^2 \geq 2$. Следовательно, если

$$n_1/n_0 \geq \sqrt{2}, \quad (11.3)$$

то любой меридиональный луч, падающий на входной торец ВС, распространяется вдоль стержня, образуя направляемую моду.

Условие (11.3), которое справедливо и для косых лучей, легко реализуется на практике. Например, кварцевый стержень ($n_1 = 1,46$) в воздухе ($n_0 = 1$) будет направлять все световые лучи, падающие на его торец. Однако, несмотря на это, казалось бы, очевидное преимущество, такой световод в оптических кабелях не применяется. Обусловлено это следующими причинами. Стекланные ОВ, имеющие диаметр порядка сотых долей миллиметра (сотен микрон), для повышения механической прочности и защиты от внешних воздействий обычно помещают внутрь защитного полимерного покрытия. Иногда внутрь такого защитного чехла помещают пучок ОВ. В местах соприкосновения волокон между собой и с защитным покрытием нарушается условие полного внутреннего отражения, и возникают моды излучения, которые являются причиной дополнительных потерь. Кроме того, на боковой поверхности ОВ возможны различные царапины. Любое повреждение поверхности приводит к возник-

новению рассеянного света, поскольку лучи в областях повреждений не испытывают полного внутреннего отражения, а выходят из волокна.

Указанные нежелательные эффекты можно предотвратить или существенно ослабить, если покрыть волокно *оболочкой* из оптически прозрачного материала с показателем преломления n_2 более низким, чем показатель преломления центральной части, которую обычно называют *сердцевинной*. При соответствующем выборе радиуса оболочки электромагнитное поле на ее внешней границе будет практически отсутствовать. Этим достигается почти полная концентрация в волокне передаваемой световой энергии и исключается искажение поля соседними волокнами и защитным покрытием. Наличие оболочки на ОВ позволяет обеспечить оптимальное соотношение между n_1 и n_2 . Выбор этого соотношения в значительной мере зависит от назначения и области применения оптического кабеля и определяется различными факторами. Например, увеличивая разность между n_1 и n_2 , можно существенно повысить эффективность ввода излучения источника в световод и в соответствии с (11.3) при $n_1 / n_0 \geq \sqrt{2}$ сделать ее максимальной. С другой стороны, искажение (уширение) импульсных сигналов в световоде оказывается тем меньше, чем ближе к единице отношение n_1/n_2 .

Ступенчатый волоконный световод

На рис. 11.14 представлена конструкция волокна в оболочке, получившая наибольшее распространение на практике. Оптическое излучение распространяется в ОВ при выполнении условия $n_1 > n_2 > n_0$. Соотношение между n_1 и n_2 принято характеризовать относительной разностью показателей преломления $\Delta = (n_1 - n_2)/n_1 = \Delta n/n_1$. Для большинства ОВ $\Delta = 10^{-2} \dots 10^{-3}$.

Показатель преломления оболочки имеет постоянное значение, а показатель преломления сердцевинны может оставаться

постоянным или изменяться вдоль радиуса по определенному закону. Изменение показателя преломления ОВ вдоль радиуса $n(r)$ называется *профилем показателя преломления*.

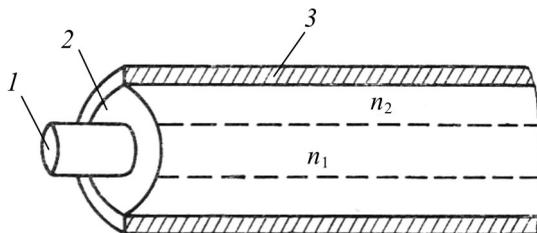


Рис. 11.14. Оптическое волокно: 1 – сердцевина; 2 – оболочка; 3 – защитное покрытие

В зависимости от профиля показателя преломления ОВ подразделяются на *ступенчатые* и *градиентные*. Волокно (световод) называется ступенчатым, если значение показателя преломления остается постоянным в пределах сердцевины. В случае градиентного волокна профиль показателя преломления является монотонно убывающей функцией радиуса в пределах его сердцевины.

Наличие оболочки, естественно, усложняет волновой процесс в ВС. В случае ступенчатого ВС (рис. 11.15) часть меридиональных и косых лучей, многократно отражаясь от границы сердцевина – оболочка, распространяется вдоль сердцевины и образует моды сердцевины или направляемые моды. Остальные лучи, которые падают на эту границу под углами $\varphi < \varphi_{кр}$, уходят из сердцевины. Однако в отличие от диэлектрического стержня не все лучи, покинувшие сердцевину, образуют моды излучения. Некоторые распространяются в оболочке за счет полного внутреннего отражения от границы оболочка – окружающая среда и образуют моды оболочки. Если защитное покрытие ОВ выполнено из сильно поглощающего материала, то оно будет устранять перекрестные помехи между ОВ в оптическом кабеле, которые обусловлены модами излучения. Кроме того, защитное покрытие будет рассеивать энергию мод оболоч-

ки аналогично тому, как это происходит с направляемыми модами диэлектрического стержня.

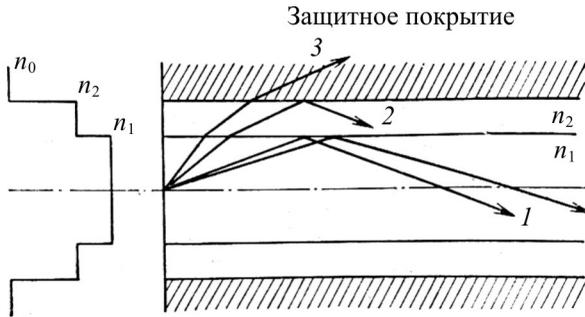


Рис. 11.15. Распространение лучей в ступенчатом волоконном световоде: 1 – моды сердцевины (направляемые моды); 2 – моды оболочки; 3 – моды излучения

Рассмотрим ход меридиональных лучей, падающих из свободного пространства с показателем преломления n_0 на входной торец ступенчатого ВС и распространяющихся затем вдоль сердцевины (рис. 11.16). Луч, падающий под углом θ к оси световода, преломляется на торцевой поверхности под углом θ_1 и затем падает на границу сердцевина – оболочка под углом $\varphi = \pi/2 - \theta_1$. Поскольку $\varphi \geq \varphi_{\text{кр}}$, этот луч распространяется вдоль сердцевины, многократно претерпевая полное внутреннее отражение на границе с оболочкой. При $\varphi = \varphi_{\text{кр}}$ получается максимальный угол падения (ввода, при котором луч еще удерживается сердцевиной). Значение этого угла θ_{max} можно указать в радианах или градусах, но обычно этот угол характеризуют величиной

$$A = n_0 \sin \theta_{\text{max}}, \quad (11.4)$$

которая называется *числовой апертурой*. В соответствии со вторым законом Снеллиуса можно записать:

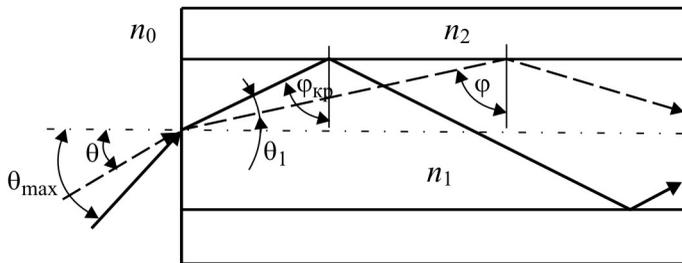


Рис. 11.16. Прохождение меридиональных лучей по ступенчатому световоду

$$n_0 \sin \theta_{\max} = n_1 \sin \theta_1 = n_1 \sin \left(\pi/2 - \varphi_{\text{кр}} \right).$$

Подставляя сюда значение $\varphi_{\text{кр}}$ (11.1), после преобразований находим значение числовой апертуры:

$$A = \sqrt{n_1^2 - n_2^2} \approx n_1 \sqrt{2\Delta}. \quad (11.5)$$

Здесь учтено, что при $\Delta = (n_1 - n_2)/n_1 \ll 1$, $n_1 + n_2 \approx 2n_1$. Таким образом, числовая апертура ступенчатого ВС определяет синус половины угла при вершине конического пучка лучей, которые захватываются и направляются ВС. Из (11.5) видно, что с увеличением разности показателей преломления сердцевины и оболочки значение A возрастает, что улучшает эффективность ввода излучения в световод. Например, при $n_1 = 1,51$ и $n_2 = 1,13$ все меридиональные лучи, падающие из воздуха ($n_0 = 1$) на торец ВС, входят в сердцевину [$A = 1$, $\theta_{\max} = \pi/2$]. Однако возрастание A приводит к увеличению дисперсии импульсов. Поэтому для ступенчатых ВС, используемых в системах связи, числовая апертура обычно равна 0,18...0,23 и лишь для отдельных типов световодов может достигать 0,4...0,55.

Формула (11.5) учитывает только меридиональные лучи ВС. Однако в реальных условиях лишь ограниченная часть пучка световых лучей источника преобразуется в меридиональные

лучи. В основном же в ВС преобладают косые лучи, которые не пересекают его ось, а распространяются по ломаным или плавным право- или левовинтовой спиральям (рис. 11.17). Законы распространения косых лучей сложнее, чем меридиональных, и вывести простое окончательное выражение для числовой апертуры косых лучей не удастся. Поэтому отметим лишь, что числовая апертура, подсчитанная для меридиональных лучей ступенчатого ВС (11.5), меньше действительной числовой апертуры A_d , учитывающей все лучи.

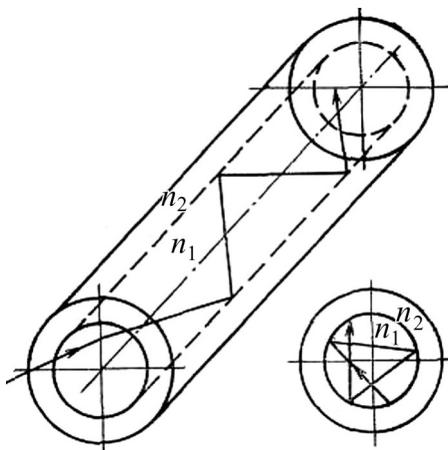


Рис. 11.17. Прохождение косого луча в ступенчатом световоде

Однако такой подход, основанный на законах геометрической оптики, не учитывает свойств света как электромагнитной волны и во многих случаях не позволяет получить правильные результаты. Например, из рассмотренного выше понятия числовой апертуры следует, что вся бесконечная совокупность лучей конического пучка, определяемого углом θ_{\max} , образует также бесконечную совокупность направляемых мод. Однако это не так. Учитывая волновой характер света, можно показать, что только конечное число лучей конического пучка с определен-

ными углами падения на торец могут образовывать направляемые моды ВС. В ступенчатом световоде с точки зрения геометрической оптики это объясняется тем, что при полном внутреннем отражении от границы с оболочкой волна приобретает *фазовый сдвиг*, зависящий от угла падения. Если в сердцевине многократно отраженные волны складываются в фазе, образуется направляемая мода. В противном случае поля волн взаимно компенсируются.

Градиентный волоконный световод

В градиентных волоконных световодах (оптических волокнах) в отличие от ступенчатых профиль показателя преломления является монотонной убывающей функцией радиуса в пределах сердцевины. Вид этой функции может быть более или менее сложным. В настоящее время наиболее изучены характеристики ВС со степенным профилем:

$$n(r) = \begin{cases} n_1 \left[1 - \Delta (r/a)^g \right] & \text{при } 0 \leq r \leq a, \\ n_2 & \text{при } a \leq r \leq b, \end{cases} \quad (11.6)$$

где n_1 – значение показателя преломления на оси ВС (т.е. при $r = 0$); n_2 – показатель преломления оболочки; $\Delta = (n_1 - n_2)/n_1$ – относительная разность показателей преломления; g – показатель степени, определяющий изменение $n(r)$.

Показатель степени g в принципе может принимать любые значения от 1 до ∞ . Нетрудно видеть, что при $g \rightarrow \infty$ получается ВС со ступенчатым профилем. Наибольшее распространение на практике получили градиентные ВС с $g = 2$, называемые параболическими:

$$n(r) = \begin{cases} n_1 \left[1 - \Delta (r/a)^2 \right] & \text{при } 0 \leq r \leq a, \\ n_2 & \text{при } a \leq r \leq b. \end{cases} \quad (11.7)$$

При изготовлении градиентных ОВ трудно, а порой невозможно получить необходимый профиль показателя преломления. Например, по технологическим причинам часто в центре сердцевины получается область с уменьшенным значением показателя преломления, а максимальное его значение находится вблизи периферии сердцевины. Такие ОВ получили название волокон с осевым провалом в профиле, или кольцевых.

Моды оболочки и излучения волоконного градиентного ВС (рис. 11.18), как и ступенчатого, образуются меридиональными и косыми лучами, покинувшими сердцевину, и подавляются защитным покрытием. Что касается направляемых мод, то здесь вместо полного внутреннего отражения на границе сердцевина – оболочка лучи плавно изгибаются в направлении градиента показателя преломления. По аналогии с (11.4) и (11.5) введем для градиентного ВС локальную числовую апертуру:

$$A(r) = n_0 \sin \theta_{\max} = \sqrt{n^2(r) - n_2^2}. \quad (11.8)$$

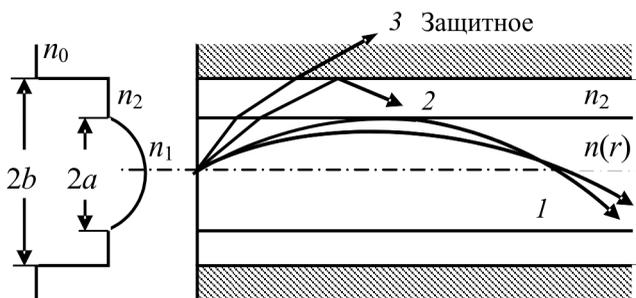


Рис. 11.18. Распространение лучей в градиентном волоконном световоде: 1 – моды сердцевины (направляемые моды); 2 – моды оболочки; 3 – моды излучения

Формула (11.8) показывает, что максимальный угол падения меридионального луча из свободного пространства на входной торец градиентного ВС, при котором он еще удерживается

сердцевиной, зависит от того, в какой точке сердцевины находится этот луч. Вблизи границы с оболочкой локальная числовая апертура стремится к нулю, а на оси световода достигает максимального значения $A = \sqrt{n_1^2 - n_2^2}$, которое называется числовой апертурой градиентного ВС.

При расчете эффективности ввода излучения в градиентный ВС удобно рассматривать этот световод как ступенчатый и характеризовать его эффективной числовой апертурой. Для ВС с параболическим профилем показателя преломления эффективная числовая апертура

$$A_{\text{эф}} = \frac{1}{\sqrt{2}} \sqrt{n_1^2 - n_2^2}. \quad (11.9)$$

Определение числовой апертуры косых лучей градиентного ВС представляет еще более сложную задачу, чем в случае ступенчатого.

Сравнение по одинаковой эффективной апертуре показывает преимущество градиентного ВС по сравнению со ступенчатым, особенно ярко эти преимущества проявляются при передаче информации в многомодовом режиме.

На рис. 11.19 представлена сравнительная картина распространения света в различных световодах. В световоде со ступенчатым профилем показателя преломления свет распространяется, испытывая многократное полное отражение от границы раздела между сердцевинкой и оболочкой световода (рис. 11.19, *а*). Распространяющийся свет содержит два типа мод: моду, которая распространяется почти вдоль оптической оси 1 и группу мод 2 , которые распространяются под углом к оптической оси, многократно отражаясь от границы раздела. Эти моды имеют различную групповую скорость, поэтому, возбуждаясь одновременно на входе, они разделяются так, что на выходе между ними образуется временной сдвиг. Следовательно, при распространении короткого светового импульса в таком световоде ширина импульса

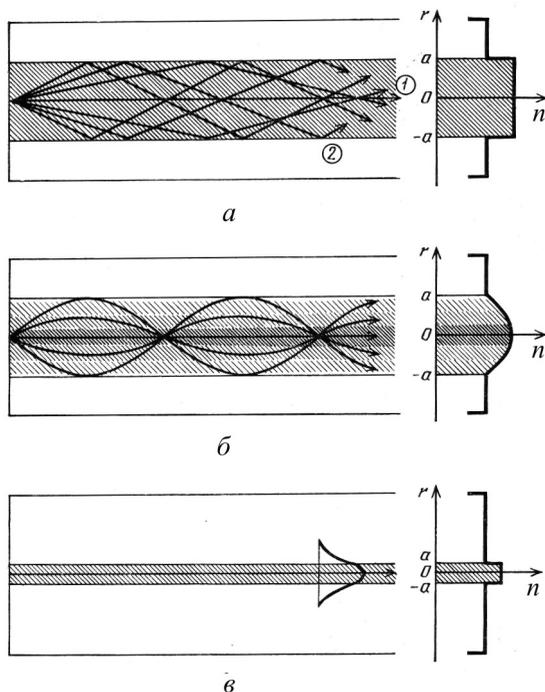


Рис. 11.19. Распространение света: *а* – в волоконном световоде со ступенчатым профилем показателя преломления; *б* – в градиентном световоде; *в* – в одномодовом световоде

возрастает. Межмодовая дисперсия практически отсутствует в градиентном световоде (рис. 11.19, *б*). В одномодовом световоде импульс передается без искажения (рис. 11.19, *в*).

11.3. Материалы для изготовления оптических волокон

В настоящее время ВС изготавливают главным образом из кварцевого стекла (волокна «кварц – кварц») и многокомпонентных стекол соответствующего состава (волокна «стекло – стекло»). В некоторых случаях применяют полимерные ОВ. Иногда сердцевину выполняют из кварцевого или многокомпонентного стек-

ла, а оболочку из полимера (например, волокна «кварц – полимер»). Из кварцевого стекла изготавливают ВС высокого качества. Достоинство его перед другими видами оптически прозрачных диэлектриков состоит в том, что он обладает наименьшими потерями на поглощение.

Для создания необходимой разности показателей преломления сердцевины и оболочки ОВ кварцевое стекло легируют соответствующими веществами, например оксидами германия, фосфора, бора и др. Так, для увеличения показателя преломления сердцевины двухслойного ОВ в состав SiO_2 вводят такие легирующие добавки, как окислы GeO_2 , P_2O_5 и TiO_2 .

Требуемую разность показателей преломления сердцевины и оболочки ОВ можно обеспечить, уменьшая показатель преломления кварца путем легирования его веществами, понижающими показатель преломления, например двуокисью бора B_2O_3 . Другая возможность понижения показателя преломления заключается в добавлении фтора в плавный кварц. Разность показателей преломления чистого SiO_2 и материала с добавкой фтора увеличивается линейно с увеличением концентрации фтора.

Высокая температура плавления кварца затрудняет производство ОВ. Для облегчения технологического процесса применяют различные добавки к SiO_2 , позволяющие не только варьировать показатель преломления, но и снижать температуру плавления.

ОВ становится световодом только в том случае, если сердцевина имеет более высокий показатель преломления, чем окружающая область. Поэтому большинство ОВ изготавливают с оболочкой из чистого SiO_2 , а показатель преломления сердцевины повышают присадкой оксида германия или фосфора. Иногда делают наоборот: из чистого кварца изготавливают сердцевину, а снижение показателя преломления оболочки достигают легированием оксидом бора. В последнее время все более широкое применение находят ОВ из многокомпонентных стекол и полимеров. Многокомпонентные стекла (натрий-боросиликаты, силикаты калия, алюмосиликаты натрия) имеют значительно

более низкую температуру плавления, чем кварцевое стекло, что существенно упрощает процесс вытяжки ОВ.

Однако при существующей технологии изготовления таких стекол, связанной с плавлением исходных материалов, невозможно обеспечить высокую чистоту стекломассы. Поэтому многокомпонентные стекла целесообразно использовать при дешевом массовом производстве ОВ среднего качества, например для обеспечения связи на короткие расстояния на внутренних сетях.

Полимерные ВС имеют также значительно более высокие потери, чем стеклянные. Например, в лучших ВС из полиметилметакрилата затухание составляет около 20 дБ/км. Однако полимерные ОВ очень дешевы и отличаются высокими механическими характеристиками. Это позволяет широко использовать полимерные ОВ в оптических линиях протяженностью в сотни метров, где они способны успешно конкурировать с ОВ из многокомпонентного стекла.

11.4. Технология изготовления оптических волокон

Одним из начальных этапов технологического процесса производства ОВ является изготовление из стекла *заготовки*, которой придают в поперечном сечении нужную геометрию волокна. Поэтому заготовку, позволяющую получать ОВ длиной более 10 км при диаметре 125 мкм, достаточно подвергнуть вытяжке.

Технологический процесс изготовления заготовок для кварцевых ОВ может базироваться на *методе парофазного осаждения*. В основе этого метода лежит реакция окисления высокочистых компонентов (например, SiCl_4 , GeCl_4), в результате которой образуются частицы SiO_2 и GeO_2 . В зависимости от того, где образуются частицы стекла – на внешней или внутренней поверхности исходной трубки – при указанном методе происходит внешнее и внутреннее осаждение.

Процесс *внешнего парофазного осаждения* осуществляется горелкой, в которую подают смесь примесей в виде хлоридов с чистым кислородом и горючим газом. Образующиеся пары

гидролизуются в пламени, образуя малые порошкообразные частицы высокочистого стекла. Поток этих частиц стекла осаждается со скоростью 0,5...1,0 г/мин на вращающемся стержне, образуя пористую заготовку со средним размером пор приблизительно 0,3 мкм, общей пористостью около 75 %.

Полученную пористую заготовку после зонного стеклования при прохождении через горячую зону печи (1500 °С) превращают в сплошную заготовку, т.е. стеклянный стержень с сечением, подобным сечению будущего ОБ. Рассмотренным методом можно изготавливать ОБ как со ступенчатым, так и с градиентным профилем показателя преломления.

Метод *внутреннего парофазного осаждения* базируется на реакциях окисления галогенидов (SiCl_4 , GeCl_4 , BCl_3) в паровой фазе с последующим осаждением на внутренней стенке стеклянной трубки, находящейся в высокотемпературной зоне.

В трубку вводится смесь необходимых газов (SiCl_4 , O_2 и примеси), и реакция происходит на стенках, нагретых до 1300...1600 °С, что вызывает осаждение стекла на внутренней поверхности. За один оборот трубки над горелкой наращивается слой толщиной в несколько микрометров. Изменяя концентрацию основных добавок (германий, бор, фосфор и др.), можно варьировать показатель преломления от слоя к слою и получать не только двухслойные, но и градиентные ОБ. После того как толщина слоев стекла достигнет требуемого размера, процесс химического осаждения прекращается.

По окончании процесса осаждения стекла на внутреннюю поверхность опорной трубки полученную трубчатую заготовку видоизменяют в стержневую. Для этого ее конец нагревают до 1900 °С, за счет чего он размягчается и под действием сил поверхностного натяжения расплава сжимается в сплошной стержень, из которого в дальнейшем получают ОБ.

Метод внутреннего парофазного осаждения является более чистым по сравнению с методом внешнего парофазного осаждения и позволяет получать ОБ с затуханием менее 0,5 дБ/км.

Принцип вытяжки волокон из заготовки (*преформы*) достаточно прост – конец заготовки нагревают до температуры плавления кварцевого стекла, а затем из него вытягивают тонкую нить. Охлаждаясь на воздухе, тонкая кварцевая нить быстро стеклется и перестает удлиняться, что и обеспечивает возможность получения волокон с постоянным по длине диаметром. Температура подбирается так, чтобы можно было производить принудительную вытяжку. В этом случае форма профиля показателя преломления волокна получается близкой к форме профиля показателя преломления исходной заготовки.

Для нагревания конца заготовки в настоящее время применяются в основном циркониевые печи с индукционным нагревателем. Нагреть конец кварцевой заготовки до нужной температуры можно также с помощью более доступных источников тепла: кислородно-водородного пламени или графитовой печи. Однако при нагреве с помощью кислородно-водородного пламени нестабильность пламени увеличивает флуктуации диаметра оболочки волокна, а образующаяся как побочный продукт вода ухудшает состояние поверхности волокна.

Нагрев кварцевых заготовок с помощью графитовой печи также может приводить к их загрязнению, так как температура, необходимая для размягчения кварца, настолько высока, что способна вызвать взаимодействие SiO_2 с углеродом. Циркониевые же печи с индукционным нагревателем, как и графитовые печи, обладают теплотворной способностью, достаточной для вытягивания волокон из больших заготовок, но приводят к меньшему загрязнению волокон.

Преформа устанавливается на вершине колонки для вытяжки волокон и медленно вдвигается в печку, разогретую до температуры порядка 2100 °С. Нижний конец преформы размягчается так, что из него можно вытягивать нити нужного диаметра. При этом форма профиля показателя преломления сохраняется, несмотря на огромное изменение масштаба в поперечном направлении.

На выходе из печи установлено лазерное устройство, с помощью которого измеряется диаметр кварцевой оболочки волокна (рис. 11.20). Сигнал с выхода этого устройства используется для коррекции скорости вытяжки волокна.

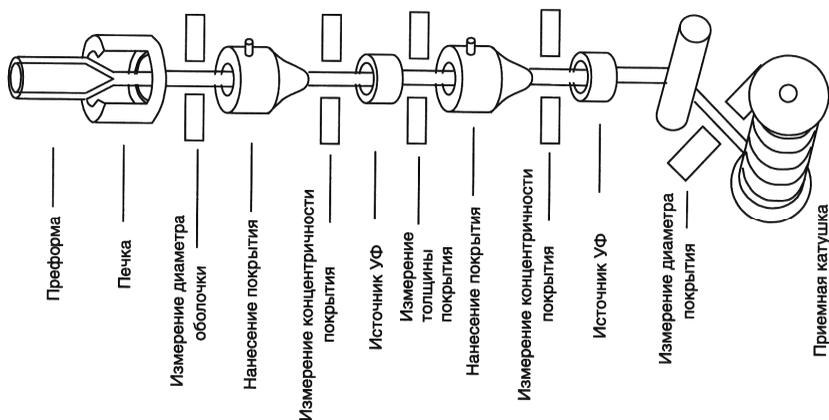


Рис. 11.20. Схема установки для вытяжки волокон

Сразу после измерения диаметра кварцевой оболочки на волокно наносится двухслойное покрытие, отверждаемое ультрафиолетовым излучением. Первый слой покрытия мягкий, а второй более твердый. Назначение покрытия состоит в механическом заполнении микротрещин и предохранении ОВ от влаги. В результате наложения покрытия прочность ОВ на разрыв повышается, что имеет первостепенную важность для дальнейших технологических операций (нанесение защитной оболочки, сборка кабеля). Концентричность наносимых на волокно покрытий контролируется по картине дифракции излучения He-Ne лазера.

11.5. Механическая прочность оптических волокон

Основными физико-механическими свойствами ОВ являются механическая прочность и минимальный радиус изгиба. Указанные физико-механические свойства определяют пригод-

ность использования ОВ в оптических кабелях, поэтому проблемам улучшения физико-механических характеристик ОВ уделяет особое внимание.

Физико-механические свойства ОВ довольно существенно отличаются от свойств исходного материала, из которого они изготавливаются. Прежде всего это относится к механической прочности, которая у ОВ больше. Так, если предел прочности при растяжении для стекла составляет $(3,9...9,8) \cdot 10^7$ Па, то ОВ диаметром 3...9 мкм, изготовленное из того же стекла, имеет этот предел $(1,47...1,9) \cdot 10^9$ Па. Одной из причин повышения прочности ОВ является так называемый масштабный фактор (увеличение прочности для изделий малого сечения).

Механическая прочность ОВ зависит от химического состава материала сердцевинки и оболочки, отношения их площадей в поперечном сечении, диаметра волокна и окружающих условий. Прочность ОВ также в значительной степени зависит от технологических условий его получения. Основным фактором, снижающим прочность ОВ в процессе изготовления, являются случайные поверхностные дефекты. Уже при малых (начальных) деформациях появляются микротрещины, приводящие, в свою очередь, к повышению локальных напряжений. Поэтому прочность зависит как от размеров дефектов, так и их распределения по длине. С увеличением длины ОВ его прочность снижается, так как число поверхностных дефектов увеличивается. В конечном счете, когда концентрация напряжений в месте расположения дефекта достигнет критического значения данного материала, ОВ обрывается.

Другим фактором, влияющим на прочность ОВ, является окружающая среда. В сухой среде прочность ОВ наибольшая, а с увеличением влажности она снижается. Дело в том, что ОВ, находящиеся в напряженном состоянии в присутствии влаги, подвержены статической усталости – процессу развития трещин во времени. В результате разрыв ОВ наступает при значительно меньших напряжениях. Оптическое волокно, находящееся в воде, снижает свою механическую прочность на 20...30 %.

Основной причиной излома ОВ является их малая гибкость. С уменьшением диаметра ОВ прочность на изгиб повышается. Так, например, если ОВ диаметром 50...70 мкм являются достаточно гибкими, то при увеличении диаметра до 100 мкм и более гибкость существенно падает. На гибкость ОВ существенно оказывает влияние и окружающая среда. С повышением влажности воздуха сопротивление изгиба резко снижается.

Для повышения прочности и гибкости ОВ в процессе изготовления покрывают защитной оболочкой. Это позволяет защитить поверхность ОВ от механических и климатических воздействий, снижающих физико-механические характеристики.

11.6. Принцип работы волоконного оптического гироскопа

Гироскоп (от греч. *gyréiō* – кружусь, вращаюсь и *skopéō* – смотрю, наблюдаю) – быстро вращающееся симметричное твердое тело, ось вращения которого (ось симметрии) может сохранять свое направление в пространстве. Это устройство является основным элементом приборов, применяемых для управления движением самолетов, ракет и в ряде других систем гироскопической стабилизации для целей навигации.

Волоконный оптический гироскоп (ВОГ) – скоростной квантовый гироскоп, основанный на использовании *эффекта Саньяка* (*G. Sagnac*, опыт 1913 г.) – смещения интерференционных полос во вращающемся кольцевом интерферометре. Это смещение возникает вследствие зависимости времени обхода светом вращающегося контура от скорости вращения и направления обхода. Согласно общей теории относительности, разность времени обхода вращающегося контура $\Delta\tau$ в рамках нерелятивистской кинематики равна

$$\Delta\tau = 4S\Omega\cos\varphi/c^2, \quad (11.10)$$

где S – площадь контура; Ω – угловая скорость вращения; φ – угол между осью вращения и нормалью к плоскости кон-

тура; c – скорость света. В результате величина сдвига интерференционных полос Δz определяется выражением

$$\Delta z = 4S \cos \varphi / \lambda_0 c, \quad (11.11)$$

где λ_0 – длина волны света в вакууме. Регистрация малых угловых скоростей вращения требует большой площади контура, поэтому практическое использование эффекта Саньяка стало осуществимым лишь с появлением волоконных световодов.

Сдвиг интерференционных полос пропорционален числу витков световода в катушке, не зависит от положения оси вращения относительно центра катушки, от формы площади катушки S , от показателя преломления света (без учета дисперсии) и записывается в виде

$$\Delta z = 2L_c R \Omega \cos \varphi / \lambda_0 c, \quad (11.12)$$

где L_c – длина оптического волокна; R – радиус катушки.

Многовитковая катушка с волоконным световодом, обеспечивающая стабильность поляризации и разности фаз интерферирующих волн, является чувствительным элементом ВОГ. Ее контур образован нитью оптического волокна длиной L_c , намотанного на цилиндр радиусом R . В соответствии с рис. 11.21 излучение *лазерного диода* подается на светоделитель и разделяется на два луча. Эти лучи, обошедшие контур в противоположных направлениях, рекомбинируют на светоделителе и смешиваются в фотодетекторе, преобразующем оптический сигнал в электрический – выходной ток фотодетектора повторяет изменения интенсивности (мощности) входного излучения. Таким образом, значения выходного тока пропорциональны в конечном счете угловой скорости вращения контура Ω . *Электронное устройство обработки информации* может вычислять угол поворота контура θ , скорость угла поворота и другие характеристики объекта, на котором установлен ВОГ.

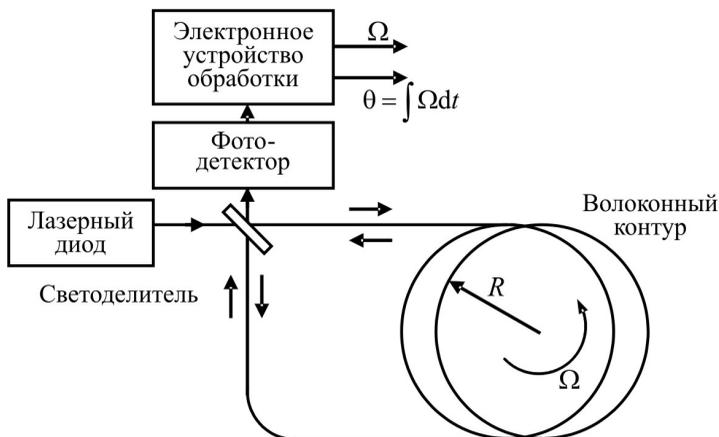


Рис. 11.21. Принципиальная схема волоконного оптического гироскопа

Для увеличения точности ВОГ используется ряд методов. Так, например, флуктуации интерференционных полос из-за рэлеевского рассеяния и невзаимные сдвиги фаз за счет разности интенсивностей встречных волн могут быть уменьшены посредством использования источников излучения с широким спектром – *полупроводниковых лазеров* (лазерных диодов). Влияние эффектов из-за изменения двойного лучепреломления в волокне при различных внешних воздействиях (механических, тепловых, акустических и др.) может быть ослаблено посредством использования *одномодовых световодов*. Прямое измерение сдвига интерференционной полосы (фазы Саньяка) ограничивает точность и динамический диапазон, поэтому в реальных ВОГ применяют более сложные методы регистрации. Например, используется *фотодетектирование* – нелинейное преобразование оптического излучения в электрический сигнал в виде последовательности импульсов электрического тока.

По сравнению со сложными и дорогостоящими электро-механическими гироскопами, применяемыми в качестве чувствительного элемента вращения в инерциальных системах нави-

гации, управления и стабилизации, ВОГ обладает следующими преимуществами:

- малые габариты и масса конструкции благодаря возможности создания ВОГ полностью на интегральных оптических схемах;

- невысокая стоимость производства и конструирования при массовом изготовлении и относительная простота технологии;

- предельно малое потребление энергии, что имеет немаловажное значение при использовании ВОГ на борту;

- отсутствие вращающихся механических элементов (роторов) и подшипников, что повышает надежность и удешевляет производство;

- практически мгновенная готовность к работе, поскольку не затрачивается время на раскрутку ротора;

- высокая чувствительность (точность) прибора, составляющая около 10^{-4} град/ч;

- большой динамический диапазон измеряемых угловых скоростей (от 1 град/ч до 300 град/с);

- нечувствительность к большим линейным ускорениям и, следовательно, работоспособность в условиях высоких механических перегрузок;

- высокую помехоустойчивость, нечувствительность к мощным внешним электромагнитным воздействиям из-за диэлектрической природы волокна;

- слабую подверженность проникающей гамма-нейтронной радиации, особенно в диапазоне 1,3 мкм.

Вопросы для самоконтроля

1. Почему на практике применяется волоконный световод, состоящий из сердцевины и оболочки?

2. Что такое «профиль показателя преломления волоконного световода»?

3. Какой волоконный световод (оптическое волокно) называется ступенчатым и какой градиентным?

4. Для чего на оптическое волокно наносят полимерное покрытие?
5. Каковы причины возникновения световых потерь в волоконном световоде и в каких единицах они измеряются?
6. Какой режим работы волоконного световода называется одномодовым и какой – многомодовым?
7. Чем определяется число направляемых мод в волоконных световодах?
8. Как определить границы одномодового режима?
9. В каких пределах находятся величины фазовых и групповых скоростей направляемых мод и чем объясняется их зависимость от длины волны излучения?
10. Как проявляется в оптическом волокне хроматическая и межмодовая дисперсия?
11. В чем состоит технология спектрального уплотнения в волоконно оптических линиях связи?
12. Как определяется числовая апертура ступенчатого и градиентного волоконного световода? Что характеризует эффективная числовая апертура?
13. Какие материалы используют для изготовления волоконных световодов? Факторы, влияющие на механическую прочность оптического волокна.
14. Какими методами изготавливают заготовки для оптических волокон?
15. Как осуществляется вытяжка оптических волокон и наносится защитное полимерное покрытие?
16. Эффект Саньяка и принцип работы волоконного оптического гироскопа.
17. Каковы преимущества волоконного оптического гироскопа по сравнению с электромеханическим гироскопом?

ЧАСТЬ 4. ОСНОВЫ ОПТОИНФОРМАТИКИ И ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА

12. ОСНОВЫ ОПТОИНФОРМАТИКИ

12.1. Предмет и задачи информатики

Информатика – это техническая наука, систематизирующая приемы создания, хранения, воспроизведения, обработки и передачи данных средствами вычислительной техники, а также принципы функционирования этих средств и методы управления ими.

Из этого определения видно, что информатика очень близка к технологии, поэтому ее предмет нередко называют *информационной технологией*.

Предмет информатики составляют следующие понятия:

- аппаратное обеспечение средств вычислительной техники;
- программное обеспечение средств вычислительной техники;
- средства взаимодействия аппаратного и программного обеспечения;
- средства взаимодействия человека с аппаратными и программными средствами.

Как видно из этого списка, в информатике особое внимание уделяется вопросам *взаимодействия*. Для этого даже есть специальное понятие – *интерфейс*. Методы и средства взаимодействия человека с аппаратными и программными средствами называют *пользовательским интерфейсом*. Соответственно существуют *аппаратные интерфейсы, программные интерфейсы и аппаратно-программные интерфейсы*.

Основной задачей информатики является систематизация приемов и методов работы с аппаратными и программными средствами вычислительной техники. Цель систематизации состоит

в выделении, внедрении и развитии передовых, наиболее эффективных технологий, в автоматизации этапов работы с данными, а также в методическом обеспечении новых технологических исследований.

Информатика – практическая наука. Ее достижения должны проходить подтверждение практикой и приниматься в тех случаях, когда они соответствуют критерию повышения эффективности. В составе основной задачи информатики сегодня можно выделить следующие направления для практических приложений:

- архитектура вычислительных систем (приемы и методы построения систем, предназначенных для автоматической обработки данных);
- интерфейсы вычислительных систем (приемы и методы управления аппаратным и программным обеспечением);
- программирование (приемы, методы и средства разработки компьютерных программ);
- преобразование данных (приемы и методы преобразования структур данных);
- защита информации (обобщение приемов, разработка методов и средств защиты данных);
- автоматизация (функционирование программно-аппаратных средств без участия человека);
- стандартизация (обеспечение совместимости между аппаратными и программными средствами, а также между форматами представления данных, относящихся к различным типам вычислительных систем).

На всех этапах технического обеспечения информационных процессов для информатики ключевым понятием является *эффективность*. Для аппаратных средств под эффективностью понимают отношение производительности оборудования к его стоимости (с учетом стоимости эксплуатации и обслуживания). Для программного обеспечения под эффективностью понимают производительность лиц, работающих с ними (пользователей). В программировании под эффективностью понимают

объем программного кода, создаваемого программистами в единицу времени.

В информатике все жестко ориентировано на эффективность. Вопрос, *как выполнить ту или иную операцию*, для информатики является важным, но не основным. Основным же является вопрос, *как сделать данную операцию эффективной*.

Слово *информатика* происходит от французского слова *Informatique*, образованного в результате объединения терминов *Information* (информация) и *Automatique* (автоматика), что выражает ее суть как науки об автоматической обработке информации. Кроме Франции, термин *информатика* используется в ряде стран Восточной Европы. В то же время в большинстве стран Западной Европы и США используется другой термин – *Computer Science* (наука о средствах вычислительной техники).

Одним из источников информатики является *кибернетика* (от греч.: *kyberneticos* – искусный в управлении). Основы близкой к информатике технической науки *кибернетики* были заложены трудами по математической логике американского математика Норберта Винера, опубликованными в 1948 году.

Впервые термин *кибернетика* ввел французский физик Андре Мари Ампер в первой половине XIX века. Он занимался разработкой единой системы классификации всех наук и обозначил этим термином гипотетическую науку об управлении, которой в то время не существовало, но которая, по его мнению, должна была существовать. Сегодня предметом кибернетики являются принципы построения и функционирования систем автоматического управления, а основными задачами – методы моделирования процесса принятия решений техническими средствами, связь между психологией человека и математической логикой, связь между информационным процессом отдельного индивидуума и информационными процессами в обществе, разработка принципов и методов искусственного интеллекта. На практике кибернетика во многих случаях опирается на те же программные и аппаратные средства вычислительной техники,

что и информатика, а информатика, в свою очередь, заимствует у кибернетики математическую и логическую базу для развития этих средств.

12.2. История информационных технологий

Изначально носителем информации была речь. Развитие речи, языка – объективный процесс в развитии общества и является *первой информационной революцией* на заре формирования человека разумного (40 тыс. лет до н.э.). Развитие речи помогало общению, передаче накопленного опыта и знаний. В дальнейшем возникла потребность в передаче информации знаковым образом – появилось первобытное искусство – целые галереи наскальных рисунков с изображением животных и охоты сохранились в пещерах. К ранним знаковым информационным системам можно отнести приметы, гадания, изобразительное искусство, музыку, графику, танец и т.д. Постепенно развивался процесс обособления и подъема духовной сферы – изобразительного и музыкального искусства, архитектуры.

Изобретение и освоение письменности стало *второй информационной революцией* (около 5 тыс. лет до н.э.). К каменному веку относятся первые примеры информационной символики – пиктографическое письмо (рисунки) на камне. В бронзовом веке появились изображения повторяющихся систем понятий – идеограмм, которые к концу IV в. до н. э. превратились в рисуночное иероглифическое письмо. В этот же период благодаря развитию производства и торговли совершенствуется числовая символика, которая сначала возникла в виде счета из двух чисел – 1 и 2. Дальнейшее развитие счета произошло благодаря физиологическим особенностям человека – наличию пальцев на руках (счет с 5 до 10). В III тыс. до н. э. в Вавилоне возникла клинописная запись счета. Позднее появились другие способы записи счета, например вавилонская, критская, латинская, арабская.

Добумажные информационные технологии (ИТ) характеризуются постоянным совершенствованием носителя информации.

Запись на камне впервые позволила добиться эффекта обезличения процесса передачи информации. Запись на глиняных табличках и деревянных дощечках дала возможность перейти к информационным коммуникациям, изобретение папируса (III тыс. до н. э.) означало значительное повышение емкости носителя информации, а применение пергамента завершило добумажную фазу: появился оптимальный носитель информации – книга (IV в. до н.э.). Александрийская библиотека была основана в 3 веке до н.э. и хранилища одного из крупнейших собраний книг той эпохи насчитывали по разным сведениям от 40 до 500 тыс. свитков. Весь объем информации, заключавшийся в свитках Александрийской библиотеки, можно оценить в 10^6 – 10^8 слов (0,1...1 Гбайт).

В те же времена начали развиваться и ИТ передачи информации. Качественно новый, более динамичный и открытый характер информационные коммуникации приобрели, когда в крупных государствах (Греция, Персия, Египет IV–III в. до н.э.) возникла хорошо налаженная почтовая связь.

Первая ИТ-система обработки информации появилась в то время, когда с увеличением объема вычислений от счета на предметах человек естественно перешел на счет на абаке. Древнегреческий абак (доска, или «саламинская доска» – по имени острова Саламин в Эгейском море) представлял собой посыпанную морским песком дощечку (рис. 12.1). На песке проводились бороздки, на которых камешками обозначались числа. Римляне усовершенствовали абак, перейдя от деревянных досок, песка и камешков к мраморным доскам с выточенными желобками и мраморными шариками.

Абак, или в дальнейшем счеты, сохранился до эпохи Возрождения, а в ви-

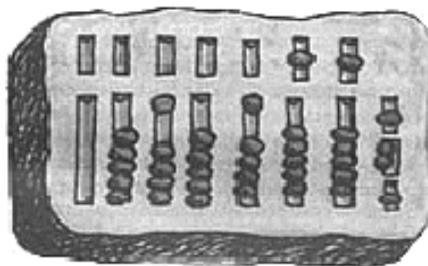


Рис. 12.1. Древнеримский абак

доизмененном виде, сначала как «дощатый счет» и как русские счеты, – до наших дней. Абак удобно использовать для выполнения операций сложения и вычитания, умножение и деление выполнять при помощи абака гораздо сложнее. Революцию в области механизации умножения и деления и соответственно в области ИТ-обработки информации совершил шотландский математик Джон Непер (John Naper). Изобретение логарифмов в 1614 году – крупнейшее достижение Джона Непера. При помощи логарифмических таблиц легко было выполнять умножение и деление больших чисел. Джон Непер также создал прибор для умножения с использованием логарифмов, названный счетными палочками (рис. 12.2). Особенно интересно изобретение Непером счетной доски для умножения, деления, возведения в квадрат, извлечения квадратного корня в двоичной системе счисления. В 1622 году, используя принцип действия этого устройства, Вильям Оугтред (William Oughtred) разработал логарифмическую линейку, которая в XIX–XX веках стала основным инструментом инженеров.

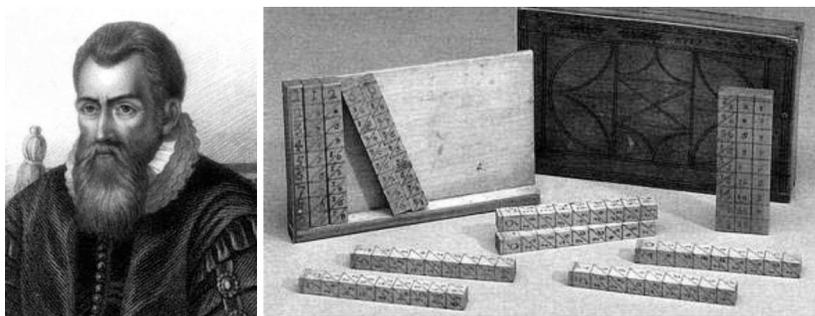


Рис. 12.2. Джон Непер (1550–1617) создал деревянную машину для выполнения простейших вычислений – счетные палочки

В дневниках гениального итальянца Леонардо да Винчи (1452–1519) уже в наше время был обнаружен ряд рисунков, которые оказались эскизным наброском первой вычислительной

машины – тринадцатирядного десятичного суммирующего устройства на основе колес с десятью зубцами (рис. 12.3).

Считается, что первая машина, способная автоматически выполнять четыре арифметических действия, была создана в 1623 году Вильгельмом Шиккардом (1592–1635). Причиной, побудившей Шиккарда разработать счетную машину для суммирования и умножения шестирядных десятичных чисел, было его знакомство с польским астрономом И. Кеплером. Ознакомившись с работой великого астронома, связанной в основном с вычислениями, Шиккард загорелся идеей оказать ему помощь в нелегком труде. В письме на его имя, отправленном в 1623 году, он приводит рисунок машины и рассказывает, что она устроена на базе шестирядного десятичного вычислителя, состоявшего также из зубчатых колес, рассчитанного на выполнение сложения, вычитания, а также табличного умножения и деления. Об изобретениях Леонардо да Винчи и Вильгельма Шиккарда стало известно лишь в наше время, современникам они были неизвестны.

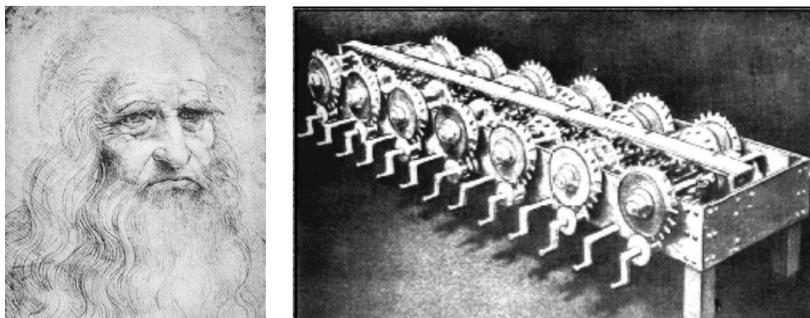


Рис. 12.3. Автопортрет и модель счетного устройства Леонардо да Винчи

В 1642 году великий французский ученый Блез Паскаль (1623–1662) механизировал канцелярские расчеты по налогообложению, соорудив настольный арифмометр на основе зубчато-

го колеса. 18-летний сын французского сборщика налогов изобрел механический калькулятор, чтобы помочь отцу в расчетах с пошлинами. В медной прямоугольной коробке, получившей название «*Pascaline*», были размещены восемь подвижных дисков (рис. 12.4).

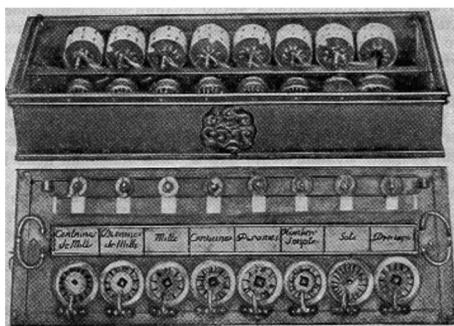


Рис. 12.4. Блез Паскаль и его машина для суммирования чисел

В 1673 году другой великий европеец, немецкий ученый Вильгельм Готфрид Лейбниц (1646–1716), создает счетную машину (арифметический прибор, по словам Лейбница) для сложения и умножения двенадцатиразрядных десятичных чисел. К зубчатым колесам он добавил ступенчатый валик, позволяющий осуществлять умножение и деление.

В 1821 году француз Карл Томас организовал серийное производство арифмометров, основанных на применении ступенчатого валика Лейбница. В дальнейшем петербургским ученым В.Т. Однером был создан арифмометр и организован его массовый выпуск. Арифмометры распространились по всему миру. Несколько десятков лет это была самая распространенная вычислительная машина. Однер заменил ступенчатые валики Лейбница зубчатым колесом с меняющимся числом зубцов.

В 1876 году был создан первый арифмометр Чебышева, который является 10-разрядной суммирующей машиной с непрерывной передачей десятков, где колесо высшего разряда

продвигается на одно деление, в то время как колесо низшего разряда переходит с 9 на 0. На основе арифмометра Чебышева в 1935 году в СССР был выпущен клавишный полуавтоматический арифмометр КСМ-1 (клавишная счетная машина). Эта машина имела два привода: электрический (со скоростью 300 оборотов в минуту) и ручной.

Считается, что первым ученым, предложившим использовать принцип программного управления для автоматического выполнения арифметических вычислений, был английский профессор математики Чарльз Бэббидж (1791–1871). Разочарованный большим количеством ошибок в вычислениях Королевского астрономического общества, Бэббидж пришел к мысли о необходимости автоматизации вычислений. Первая попытка реализации такой машины была предпринята Бэббиджем в 1822 году, когда он создал машину, предназначенную для решения дифференциальных уравнений, названную «разностной машиной» (рис. 12.5).

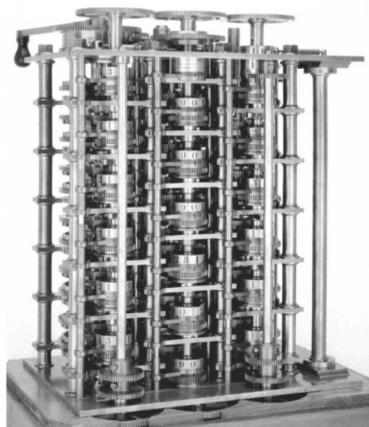
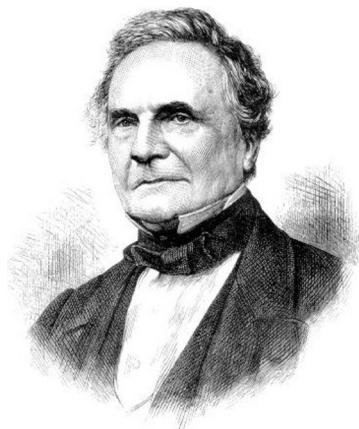


Рис. 12.5. Чарльз Бэббидж и модель аналитической машины с 25 тыс. деталей

Работа модели основывалась на принципе, известном в математике как «метод конечных разностей». Аналитическая

машина (так назвал ее Бэббидж), проект которой он разработал в 1836–1848 годах, стала механическим прототипом появившихся спустя столетие ЭВМ. В ней предполагалось иметь те же, что и в ЭВМ, пять основных устройств: *арифметическое, памяти, управления, ввода, вывода*. Для арифметического устройства Ч. Бэббидж использовал зубчатые колеса, подобные тем, что использовались ранее. На них же Ч. Бэббидж намеревался построить устройство памяти из 1000 50-разрядных регистров (по 50 колес в каждом). Программа выполнения вычислений записывалась на перфокартах (пробивками), на них же записывались исходные данные и результаты вычислений. В число операций, помимо четырех арифметических, была включена операция условного перехода и операции с кодами команд. Автоматическое выполнение программы вычислений обеспечивалось устройством управления. Время сложения двух 50-разрядных десятичных чисел составляло, по расчетам ученого, 1 с, умножения – 1 мин. Движение механических частей машины должен был обеспечивать паровой двигатель. Большая, как локомотив, машина должна была автоматически выполнять вычисления и печатать результаты. Программы вычислений на машине Бэббиджа, составленные дочерью Байрона Адой Августой Лавлейс (1815–1852), поразительно схожи с программами, составленными впоследствии для первых ЭВМ. Не случайно именем этой женщины назвали одну из первых систем программирования. Большая разностная машина так и не была построена до конца. В 1871 году Бэббидж изготовил опытный образец арифметического устройства («завода») аналитической машины и принтера. Технические трудности, с которыми пришлось встретиться при реализации, не позволили осуществить проект, поэтому Бэббидж не опубликовал проект полностью, а ограничился описанием его в своих лекциях, чертежах и рисунках.

В 1847 году английский математик Джордж Буль (1815–1864) опубликовал работу «Математический анализ логики». Появился

новый раздел математики, получивший название «Булева алгебра». Каждая величина в ней может принимать только одно из двух значений: истина или ложь, 1 или 0. Буль изобрел своеобразную алгебру – систему обозначений и правил, применимую к различным объектам, от чисел до предложений. Пользуясь правилами алгебры, он мог закодировать высказывания (утверждения, истинность или ложность которых требовалось доказать) с помощью символов своего языка, а затем манипулировать ими, подобно тому как в математике манипулируют числами. Основными операциями булевой алгебры являются *конъюнкция* (И), *дизъюнкция* (ИЛИ), *отрицание* (НЕ). Через некоторое время стало понятно, что система Буля хорошо подходит для описания переключательных схем. Ток в электрической цепи может либо протекать, либо отсутствовать, подобно тому, как утверждение может быть либо истинным, либо ложным. Разработанная им алгебра логики (алгебра Буля) нашла применение лишь в следующем веке, когда понадобился математический аппарат для проектирования схем ЭВМ, использующих двоичную систему счисления.

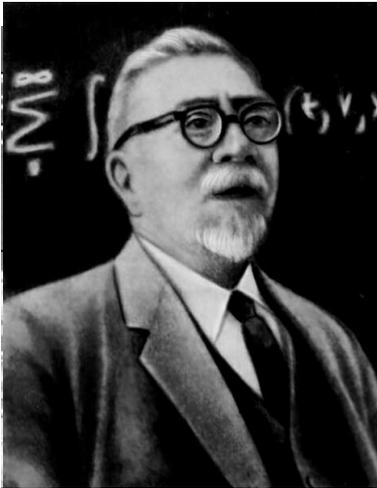
Через 63 года после смерти Ч. Бэббиджа немецкий студент Конрад Цузе (1910–1985) взял на себя задачу создать машину, подобную по принципу действия той, которой отдал жизнь Ч. Бэббидж. Работу по созданию машины он начал в 1934 году, за год до получения инженерного диплома. Конрад ничего не знал ни о машине Бэббиджа, ни о работах Лейбница, ни об алгебре Буля, которая словно создана для того, чтобы проектировать схемы с использованием элементов, имеющих лишь два устойчивых состояния. Тем не менее он оказался достойным наследником В. Лейбница и Дж. Буля, поскольку вернул к жизни уже забытую двоичную систему исчисления, а при расчете схем использовал нечто подобное булевой алгебре. В 1937 году машина Z1 (что означало «Цузе 1») заработала. Она была подобно машине Бэббиджа чисто механической. Использование двоичной системы сотворило чудо – машина занимала всего два квадратных метра на столе в квартире изобретателя! Длина слов составляла 22 двоичных разряда. Выполнение операций производилось с использованием

плавающей запятой. Для мантиссы и ее знака отводилось 15 разрядов, для порядка – 7. Память (тоже на механических элементах) содержала 64 слова (против 1000 у Беббиджа, что тоже уменьшило размеры машины). Числа и программа вводились вручную. Еще через год в машине появилось устройство ввода данных и программы, использовавшее киноленту, на которую перфорировалась информация, а механическое арифметическое устройство заменило аналогичное устройство последовательного действия на телефонных реле.

Почти одновременно, в 1943 году, американец Говард Эйкен с помощью работ Бэббиджа на основе техники XX века – электро-механических реле – смог построить на одном из предприятий фирмы ИВМ легендарный гарвардский «Марк-1». «Марк-1» имел в длину 15 метров и в высоту 2,5 метра, содержал 800 тысяч деталей, располагал 60 регистрами для констант, 72 запоминающими регистрами для сложения, центральным блоком умножения и деления, мог вычислять элементарные трансцендентные функции.

Работа по созданию первой электронно-вычислительной машины была начата, по-видимому, в 1937 году в США профессором Джоном Атанасовым, болгариним по происхождению. Эта машина была специализированной и предназначалась для решения задач математической физики. В ходе разработок Атанасов создал и запатентовал первые электронные устройства, которые впоследствии применялись довольно широко в первых компьютерах. Атанасов сформулировал, а в 1939 году опубликовал окончательный вариант своей концепции современной вычислительной машины. Полностью проект Атанасова не был завершен, однако через три десятка лет в результате судебного разбирательства профессора признали родоначальником электронной вычислительной техники.

Летом 1940 года другим выдающимся ученым Норбертом Винером (1894–1964) были сформулированы требования к созданию быстродействующей вычислительной машины, которые были полностью учтены дальнейшими разработчиками ЭВМ.



Основатель кибернетики
Норберт Винер

Начиная с 1943 года группа специалистов под руководством Говарда Эйкена, Дж. Моучли и П. Эккерта в США начала конструировать вычислительную машину на основе электронных ламп, а не на электромагнитных реле. Эта машина была названа *ENIAC (Electronic Numeral Integrator And Computer)* и работала она в тысячу раз быстрее, чем «Марк-1». ENIAC содержал 18 тысяч вакуумных ламп, занимал площадь 9×15 метров, весил 30 тонн и потреблял мощность 150 киловатт. ENIAC имел и существенный недоста-

ток – управление им осуществлялось с помощью коммутационной панели. У него отсутствовала память, и, для того чтобы задать программу, приходилось в течение нескольких часов или даже дней подсоединять нужным образом провода. Худшим из всех недостатков была ужасающая ненадежность компьютера, так как за день работы успевало выйти из строя около десятка вакуумных ламп.

Чтобы упростить процесс задания программ, Моучли и Эккерт стали конструировать новую машину, которая могла бы хранить программу в своей памяти. В 1945 году к работе был привлечен знаменитый математик Джон фон Нейман, который подготовил доклад об этой машине (рис. 12.6). В этом докладе фон Нейман, основываясь на идеях Н. Винера, сформулировал общие принципы функционирования универсальных вычислительных устройств, т.е. компьютеров. Это первая действующая машина, построенная на вакуумных лампах, официально была введена в эксплуатацию 15 февраля 1946 года. Эту машину пы-

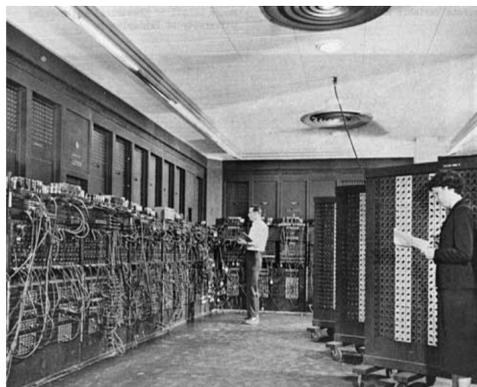


Рис. 12.6. Джон фон Нейман и внешний вид ENIAC
(*Electronic Numeral Integrator And Computer*)

тались использовать для решения некоторых задач, подготовленных фон Нейманом и связанных с проектом атомной бомбы. Затем она была перевезена на Абердинский полигон, где работала до 1955 года. ENIAC стал первым представителем первого поколения компьютеров.

12.3. Понятие об информации

Из курса физики мы знаем, что состояния абсолютного покоя не существует, и физические объекты находятся в состоянии непрерывного движения и изменения, которое сопровождается обменом энергией и ее переходом из одной формы в другую. Все виды *энергообмена* сопровождаются появлением сигналов, то есть все сигналы имеют в своей основе материальную энергетическую природу. При взаимодействии сигналов с физическими телами в последних возникают определенные изменения свойств – это явление называется *регистрацией сигналов*. Такие изменения можно наблюдать, измерять или фиксировать иными способами – при этом возникают и регистрируются новые сигналы, то есть образуются данные. *Данные – это зарегистрированные сигналы.*

Данные несут в себе *информацию* о событиях, произошедших в материальном мире, поскольку они являются регистрацией сигналов, возникших в результате этих событий. Однако данные не тождественны информации. Наблюдая излучения далеких звезд, человек получает определенный поток данных, но станут ли эти данные информацией, зависит еще от очень многих обстоятельств.

Наблюдая за состязаниями бегунов, мы с помощью механического секундомера регистрируем начальное и конечное положение стрелки прибора. В итоге мы измеряем величину ее перемещения за время забега – это регистрация данных. Однако информацию о времени преодоления дистанции мы пока не получаем. Для того чтобы данные о перемещении стрелки дали информацию о времени забега, необходимо наличие *метода* пересчета одной физической величины в другую. Надо знать цену деления шкалы секундомера (погрешность измерения). Если вместо механического секундомера используется электронный, суть дела не меняется.

Несмотря на то, что с понятием информации мы сталкиваемся ежедневно, строгого и общепризнанного ее определения до сих пор не существует, поэтому вместо определения обычно используют *понятие об информации*. Понятия, в отличие от определений, не даются однозначно, а вводятся на примерах, причем каждая научная дисциплина делает это по-своему, выделяя в качестве основных компонентов те, которые наилучшим образом соответствуют ее предмету и задачам.

При этом типична ситуация, когда понятие об информации, введенное в рамках одной научной дисциплины, может опровергаться конкретными примерами и фактами, полученными в рамках другой науки. Например, представление об информации как о совокупности данных, повышающих уровень знаний об объективной реальности окружающего мира, характерное для естественных наук, может быть опровергнуто в рамках социальных наук.

Для информатики как технической науки информация не может основываться на таких понятиях, как *знание*, и не может опираться только на объективность фактов и свидетельств. Средства вычислительной техники обладают способностью обрабатывать информацию автоматически, без участия человека, и ни о каком знании или незнании здесь речь идти не может. Эти средства могут работать с искусственной, абстрактной и даже с ложной информацией, не имеющей объективного отражения ни в природе, ни в обществе. Поэтому *информацию можно рассматривать как продукт взаимодействия данных и адекватных им методов.*

Свойства информации

Информация не является статичным объектом – она имеет *динамический характер*, т.е. непрерывно меняется и существует только в момент взаимодействия данных и методов. Все прочее время она пребывает в состоянии данных. Таким образом, информация существует только в момент протекания *информационного процесса.*

Одни и те же данные могут в момент потребления поставлять разную информацию в зависимости от *степени адекватности* взаимодействующих с ними методов. Например, для человека, не владеющего китайским языком, письмо, полученное из Пекина, дает только ту информацию, которую можно получить методом наблюдения (количество страниц, цвет и сорт бумаги, наличие незнакомых символов и т.п.). Все это информация, но это не вся информация, заключенная в письме. Использование более адекватных методов (владение китайским языком) даст иную информацию.

Данные, полученные в результате регистрации объективно существовавших сигналов, являются *объективными*, в то же время методы являются *субъективными*, так как в основе искусственных методов лежат алгоритмы (упорядоченные последова-

тельности команд), составленные и подготовленные людьми (субъектами). Таким образом, *информация возникает и существует в момент диалектического взаимодействия объективных данных и субъективных методов*. Такой дуализм известен своими проявлениями во многих науках. Например, в физике это двойственная корпускулярно-волновая природа света. Подход к информации как к объекту особой природы, возникающему в результате диалектического взаимодействия объективных данных с субъективными методами, позволяет во многих случаях снять противоречия, возникающие в обоснованиях ряда научных теорий и гипотез.

С точки зрения информатики наиболее важными представляются следующие свойства информации: объективность, полнота, достоверность, адекватность, доступность и актуальность.

Объективность и субъективность информации. Понятие объективности информации является относительным. Это понятно, если учесть, что методы являются субъективными. Более объективной принято считать ту информацию, в которую методы вносят меньший субъективный элемент. Так, например, принято считать, что в результате наблюдения фотоснимка природного объекта или явления образуется более объективная информация, чем в результате наблюдения рисунка того же объекта, выполненного человеком. В ходе информационного процесса степень объективности информации всегда понижается. Это свойство учитывают, например, в правовых дисциплинах, где по-разному обрабатываются показания лиц, непосредственно наблюдавших события или получивших информацию косвенным путем (посредством умозаключений или со слов третьих лиц).

Полнота информации. Полнота информации во многом характеризует *качество информации* и определяет *достаточность* данных для принятия решений или для создания новых данных на основе имеющихся. Чем полнее данные, тем шире диапазон методов, которые можно использовать, тем проще подобрать метод, вносящий минимум погрешностей в ход информационного процесса.

Достоверность информации. Данные возникают в момент регистрации сигналов, но не все сигналы являются «полезными» – всегда присутствует какой-то уровень посторонних сигналов, в результате чего полезные данные сопровождаются определенным уровнем «информационного шума». Если полезный сигнал зарегистрирован более четко, чем посторонние сигналы, достоверность информации может быть более высокой. При увеличении уровня шумов достоверность информации снижается. В этом случае для передачи того же количества информации требуется использовать либо больше данных, либо более сложные методы.

Адекватность информации – это степень соответствия реальному объективному состоянию дела. Неадекватная информация может образовываться при создании новой информации на основе неполных или недостоверных данных. Однако и полные, и достоверные данные могут приводить к созданию неадекватной информации в случае применения к ним неадекватных методов.

Доступность информации – мера возможности получить ту или иную информацию. На степень доступности информации влияют одновременно как доступность данных, так и доступность адекватных методов для их интерпретации. Отсутствие доступа к данным или отсутствие адекватных методов обработки данных приводят к одинаковому результату: информация оказывается недоступной. Отсутствие адекватных методов для работы с данными во многих случаях приводит к применению неадекватных методов, в результате чего образуется неполная, неадекватная или недостоверная информация.

Актуальность информации – это степень соответствия информации текущему моменту времени. Нередко с актуальностью, как и с полнотой, связывают коммерческую ценность информации. Поскольку информационные процессы растянуты во времени, то достоверная и адекватная, но устаревшая информация может приводить к ошибочным решениям.

Носители информации

Носителями информации являются данные, представляющие собой зарегистрированные сигналы. При этом физический метод регистрации может быть любым: механическое перемещение физических тел, изменение их формы или параметров качества поверхности, изменение электрических, магнитных, оптических характеристик, изменение состояния электронной системы и многое другое. В соответствии с методом регистрации данные могут храниться и транспортироваться на носителях различных видов. Самым распространенным носителем данных, хотя и не самым экономичным, по-видимому, является бумага. На бумаге данные регистрируются путем изменения оптических характеристик ее поверхности. Изменение оптических свойств (изменение коэффициента отражения поверхности в определенном диапазоне длин волн) используется также в устройствах, осуществляющих запись лазерным лучом на пластмассовых носителях с отражающим покрытием (*CD-ROM*). В качестве носителей, использующих изменение магнитных свойств, можно назвать магнитные ленты и диски. Регистрация данных путем изменения химического состава поверхностных веществ носителя широко используется в фотографии.

Свойства информации весьма тесно связаны со свойствами ее носителей. Любой носитель можно характеризовать параметром *разрешающей способности* (количеством данных, записанных в принятой для носителя единице измерения) и *динамическим диапазоном* (логарифмическим отношением интенсивности амплитуд максимального и минимального регистрируемого сигналов). От этих свойств носителя нередко зависят такие свойства информации, как полнота, доступность и достоверность. Например, мы можем рассчитывать на то, что в базе данных, размещаемой на компакт-диске, проще обеспечить полноту информации, чем в аналогичной по назначению базе данных, размещенной на гибком магнитном диске, поскольку в первом случае плотность

записи данных на единице длины дорожки намного выше. Для обычного потребителя доступность информации в книге заметно выше, чем той же информации на компакт-диске, поскольку не все потребители обладают необходимым оборудованием.

В ходе информационного процесса данные преобразуются из одного вида в другой с помощью методов. Обработка данных включает в себя множество различных операций. По мере развития научно-технического прогресса и общего усложнения связей в человеческом обществе трудозатраты на обработку данных неуклонно возрастают. Прежде всего это связано с постоянным усложнением условий управления производством и обществом. Второй фактор, также вызывающий общее увеличение объемов обрабатываемых данных, тоже связан с научно-техническим прогрессом, а именно с быстрыми темпами появления и внедрения новых носителей данных, средств их хранения и доставки.

В структуре возможных операций с данными можно выделить следующие этапы:

- *сбор данных* – накопление информации с целью обеспечения достаточной полноты для принятия решений;

- *формализация данных* – приведение данных, поступающих из разных источников, к одинаковой форме, чтобы сделать их сопоставимыми между собой, то есть повысить их уровень доступности;

- *фильтрация данных* – отсеивание «лишних» данных, в которых нет необходимости для принятия решений; при этом должен уменьшаться уровень «шума», а достоверность и адекватность данных должны возрастать;

- *сортировка данных* – упорядочение данных по заданному признаку с целью удобства использования; повышает доступность информации;

- *архивация данных* – организация хранения данных в удобной и легкодоступной форме; служит для снижения экономических затрат по хранению данных и повышает общую надежность информационного процесса;

– *защита данных* – комплекс мер, направленных на предотвращение утраты, воспроизведения и модификации данных;

– *транспортировка данных* – прием и передача (доставка и поставка) данных между удаленными участниками информационного процесса; при этом источник данных в информатике принято называть *сервером*, а потребителя – *клиентом*;

– *преобразование данных* – перевод данных из одной формы в другую или из одной структуры в другую. Преобразование данных часто связано с изменением типа носителя: например, книги можно хранить в обычной бумажной форме, но можно использовать для этого и электронную форму, и микрофотопленку. Необходимость в многократном преобразовании данных возникает также при их транспортировке.

Совокупность технических средств, требуемых для передачи сообщения от источника к получателю, называют *системой связи*. В функциональных схемах и их реализациях такие узлы, как *кодер* и *модулятор*, объединяют в передающем устройстве; аналогично *демодулятор* и *декодер* объединяются в едином устройстве – приемнике. Модуляция (от лат. *modulatio* – мерность, размеренность) это преобразование по заданному закону стационарных параметров, характеризующих поток данных. Например, для гармонических колебаний могут быть модулированы по определенному закону амплитуда, частота и фаза колебаний. Соответственно примеры амплитудной, частотной и фазовой модуляции представлены на рис. 12.7. Типичная функциональная схема, включающая основные узлы системы связи, представлена на рис. 12.8. Указанная здесь линия связи, во многих случаях отождествляемая с каналом передачи, предназначена для передачи сигналов с минимально возможной потерей их интенсивности. В линии связи локализована неизбежно присутствующая в системе связи помеха, приводящая к случайному непредсказуемому искажению формы передаваемого сигнала.

Приведенный здесь список типовых операций с данными далеко не полон. Миллионы людей во всем мире занимаются

созданием, обработкой, преобразованием и транспортировкой данных, поэтому *работа с информацией может иметь огромную трудоемкость и ее надо автоматизировать.*

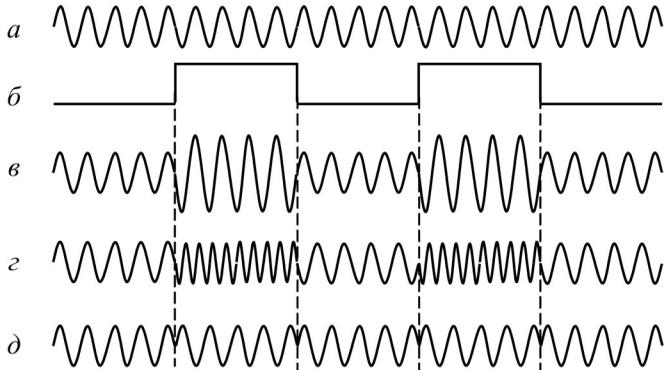


Рис. 12.7. Схемы модуляции: *а* – гармонические колебания несущей частоты; *б* – модулирующий сигнал; *в* – амплитудно-модулированное колебание; *г* – частотно-модулированное колебание; *д* – фазово-модулированное колебание



Рис. 12.8. Функциональная схема передачи информации

12.4. Измерение количества информации. Энтропия

Для автоматизации работы с данными, относящимися к различным типам, очень важно унифицировать их форму представления – для этого обычно используется прием *кодирования*, то есть выражение данных одного типа через данные другого типа. Естественные человеческие *языки* – это не что иное, как системы коди-

рования понятий для выражения мыслей посредством речи. История знает интересные, хотя и безуспешные попытки создания «универсальных» языков и азбук.

Однако эта проблема универсального средства кодирования достаточно успешно реализуется в отдельных отраслях техники, науки и культуры. В качестве примеров можно привести систему записи математических выражений, телеграфную азбуку, морскую флажковую азбуку, систему Брайля для слепых и многое другое. Своя система существует и в вычислительной технике – она называется *двоичным кодированием* и основана на представлении данных последовательностью всего двух знаков: 0 и 1. Эти знаки называются *двоичными цифрами*, по английски – *binary digit* или, сокращенно, *bit* (*бит*).

Бит (*двоичный разряд*) является наименьшей единицей представления данных, принятой в информатике и вычислительной технике. Совокупность двоичных разрядов, выражающих числовые или иные данные, образует битовый рисунок. Практика показывает, что с битовым представлением удобнее работать, если этот рисунок имеет регулярную форму. В настоящее время в качестве таких форм используются *группы из восьми битов*, которые *называются байтами*.

Понятие о байте как группе взаимосвязанных битов появилось вместе с первыми образцами электронной вычислительной техники. Долгое время оно было *машинно зависимым*, то есть для разных вычислительных машин длина байта была разной. Только в конце 60-х годов понятие байта стало универсальным и *машинно независимым*. Во многих случаях целесообразно использовать не восьмиразрядное кодирование, а 16-разрядное, 24-разрядное, 32-разрядное и более. Группа из 16 взаимосвязанных бит (двух взаимосвязанных байтов) в информатике называется *словом*. Соответственно, группы из четырех взаимосвязанных байтов (32 разряда) называются *удвоенным словом*, а группы из восьми байтов (64 разряда) – *учетверенным словом*.

Байт является в информатике наименьшей единицей измерения. Поскольку одним байтом, как правило, кодируется один символ текстовой информации, для текстовых документов размер в байтах соответствует лексическому объему в символах. Более крупная единица измерения – килобайт (Кбайт). Условно можно считать, что 1 Кбайт примерно равен 1000 байт. Условность связана с тем, что для вычислительной техники, работающей с двоичными числами, более удобно представление чисел в виде степени двойки и потому на самом деле 1 Кбайт равен 2^{10} байт (1024 байт). Однако всюду, где это не принципиально, с инженерной погрешностью (до 3 %) «забывают» о «лишних» байтах.

В килобайтах измеряют сравнительно небольшие объемы данных. Условно можно считать, что одна страница неформатированного машинописного текста составляет около 2 Кбайт.

Более крупные единицы измерения данных образуются добавлением префиксов *мега-*, *гига-* *тера-*:

$$1 \text{ Мбайт} = 1024 \text{ Кбайт} \approx 10^{20} \text{ байт};$$

$$1 \text{ Гбайт} = 1024 \text{ Мбайт} \approx 10^{30} \text{ байт};$$

$$1 \text{ Тбайт} = 1024 \text{ Гбайт} \approx 10^{40} \text{ байт}.$$

В более крупных единицах пока нет практической необходимости.

При хранении данных решаются две проблемы: как сохранить данные в наиболее компактном виде и как обеспечить к ним удобный и быстрый доступ. Для обеспечения доступа необходимо, чтобы данные имели *упорядоченную структуру*.

Поскольку адресные данные тоже имеют размер и тоже подлежат хранению, хранить данные в виде мелких единиц, таких как байты, неудобно. Их неудобно хранить и в более крупных единицах (килобайтах, мегабайтах и т.п.), поскольку неполное заполнение одной единицы хранения приводит к неэффективности хранения.

В качестве единицы хранения данных принят объект переменной длины, называемый *файлом*. *Файл* – это последовательность произвольного числа байтов, обладающая уникальным собственным именем. Обычно в отдельном файле хранят данные, относящиеся к одному типу. В этом случае тип данных определяет *тип файла*.

Проще всего представить себе файл в виде безразмерного канцелярского досье, в которое можно по желанию добавлять содержимое или извлекать его оттуда. Поскольку в определении файла нет ограничений на размер, можно представить себе файл, имеющий 0 байтов (*пустой файл*), и файл, имеющий любое число байтов.

В определении файла особое внимание уделяется имени. Оно фактически несет в себе адресные данные, без которых данные, хранящиеся в файле, не станут информацией из-за отсутствия метода доступа к ним. Кроме функций, связанных с адресацией, имя файла может хранить и сведения о типе данных, заключенных в нем. Для автоматических средств работы с данными это важно, поскольку по имени файла они могут автоматически определить адекватный метод извлечения информации из файла.

Количество информации оценивают количественной мерой H , получившей название *энтропии*. Понятие энтропии (от греч. *entropē* – обращение) распространилось на ряд областей знания. Энтропия в термодинамике характеризует вероятность теплового состояния вещества, его способность совершать работу, в математике – степень неопределенности ситуации или задачи. В теории информации понятие энтропии как способности источника отдавать информацию было введено американским инженером К. Шенноном в работе «Математическая теория связи», опубликованной в 1948 году.

Степень неопределенности состояний объекта зависит от числа n возможных состояний (S_1, S_2, \dots, S_n) с вероятностями $P(S_1), P(S_2), \dots, P(S_n)$. Например, при бросании монеты их может быть только два, при бросании кубика – шесть. Степень неопределенности состояния объекта возрастает с увеличением n . Если

состояние объекта не изменяется, $P(S_1) = 1$, а вероятности остальных состояний равны нулю, то объект вовсе не обладает неопределенностью – *энтропия такого объекта равна нулю*.

Таким образом, в теории информации энтропия как степень неопределенности состояния объекта, имеющего n возможных состояний с вероятностями $P(S_1), P(S_2), \dots, P(S_n)$, определяется выражением

$$H(S) = -\sum_{i=1}^n P(S_i) \log_a P(S_i). \quad (12.1)$$

Логарифм в формуле (12.1) может быть взят при любом основании, изменение основания приводит только к появлению множителя перед знаком суммы, т.е. к изменению единицы измерения. Наименование этой единицы зависит от выбора основания логарифма при вычислении энтропии: при выборе двоичного основания ($a = 2$) получаем уже знакомую единицу измерения информации – *бит*, при других основаниях $a = 10$ – *дит*, при использовании натуральных логарифмов – *нит*.

При использовании двоичного основания формула (12.1) принимает вид

$$H(S) = -\sum_{i=1}^n P(S_i) \log_2 P(S_i). \quad (12.2)$$

Целесообразность использования двоичных логарифмов легко понять, вычисляя энтропию объекта, имеющего два равновероятных состояния. В этом случае $P(S_1) = P(S_2) = 0,5$. По формуле (12.2) находим:

$$\begin{aligned} H(S) &= -P(S_1) \log_2 P(S_1) - P(S_2) \log_2 P(S_2) = \\ &= -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1. \end{aligned}$$

Таким образом, за единицу измерения энтропии при выборе двоичных логарифмов принимается *степень неопределен-*

ности объекта, имеющего два возможных равновероятных состояния. Эта единица измерения и называется двоичной единицей или битом.

Рассмотрим основные свойства энтропии.

1. Энтропия объекта есть величина вещественная и неотрицательная, так как $0 \leq P(S_i) \leq 1$, то $H(S) \geq 0$.

2. Энтропия объекта равна нулю в том крайнем случае, когда одно из возможных состояний S_n имеет вероятность $P(S_n) = 1$, т.е.

$$H(S) = 0 \text{ при } P(S_n) = 1. \quad (12.3)$$

В этом случае информация о состоянии объекта достоверно известна заранее, поэтому в объекте нет никакой неопределенности.

3. Энтропия объекта с одинаковыми вероятностями состояний максимальна и равна логарифму числа состояний. Действительно, пусть $P(S_i) = 1/n$, тогда по формуле для энтропии объекта (12.2) находим

$$\begin{aligned} H(S) &= -\sum_{i=1}^n P(S_i) \log_2 P(S_i) = \\ &= -\sum_{i=1}^n \frac{1}{n} \log_2 \frac{1}{n} = \sum_{i=1}^n \frac{1}{n} \log_2 n = \log_2 n. \end{aligned} \quad (12.4)$$

В качестве примера рассмотрим бинарный объект с двумя возможными несовместимыми состояниями. Вероятность первого состояния равна P , тогда вероятность второго состояния равна $1 - P$. Определим энтропию этого объекта для следующих случаев:

а) состояния объектов равновероятны, т.е. $P_1 = P_2 = 0,5$; $P_1 + P_2 = 1$. Из формулы (12.2) получаем

$$H(S) = -(0,5 \log_2 0,5 + 0,5 \log_2 0,5) = -[0,5(-1) + 0,5(-1)] = 1 \text{ бит};$$

б) неравновероятные состояния. Пусть $P_1 = 0,9$; $P_2 = 0,1$; $P_1 + P_2 = 1$. Тогда

$$H(S) = -(0,9 \log_2 0,9 + 0,1 \log_2 0,1) = \\ = -[0,9(-0,1520) + 0,1(-3,3)] = 0,46 \text{ бит};$$

в) детерминированные состояния. $P_1 = 1$; $P_2 = 0$; $P_1 + P_2 = 1$:

$$H(S) = -(1 \log_2 1 + 0 \log_2 0) = 0 \text{ бит.}$$

Изменение энтропии бинарного объекта в зависимости от вероятности $P(S_i)$ показано на рис. 12.9, из которого видно, что максимум $H(S) = 1$ достигается при $P_1 = P_2 = 0,5$, т.е. когда состояния равновероятны. При вероятностях $P = 0$ или $P = 1$, что соответствует полной невозможности или полной достоверности одного из состояний, энтропия равна нулю.

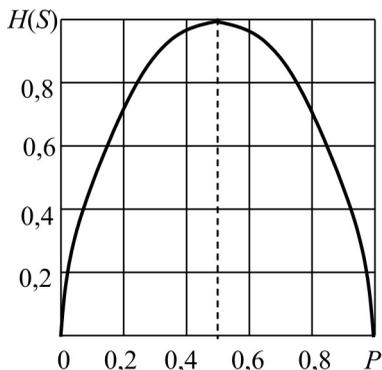


Рис. 12.9. Зависимость энтропии бинарного объекта от вероятности первого состояния

Из этого примера также видно, что количество информации, получаемой при диагностировании, определяется разностью неопределенностей (энтропий) объекта до и после диагностирования.

Если начальная энтропия объекта равна $H(S)$, а после диагностирования она составляет $H^*(S)$, то информация J , полученная в результате диагностирования, равна

$$J = H(S) - H^*(S). \quad (12.5)$$

Поясним количество информации, получаемой при диагностировании, на следующем примере. Объект находится в одном из двух равновероятных состояний – работоспособном или неработоспособном, т.е. $P_1 = P_2 = 0,5$; $\sum_{i=1}^2 P_i = 1$. Допустим также, что работоспособность объекта оценивается по двум параметрам – температуре и давлению. Известно, что при поступлении сообщения от датчика температуры об изменении температуры менее чем на 40°C объект находится в работоспособном состоянии с вероятностью $P = 0,5$. При поступлении сообщения от датчика давления о том, что оно больше $0,15$ МПа, можно с вероятностью $P = 1$ гарантировать работоспособное состояние объекта. Какое из этих сообщений несет больше информации?

В приведенном примере начальная энтропия объекта

$$H(S) = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}\right) = 1 \text{ бит.}$$

После получения сообщения от датчика температуры вероятности состояний стали $P_1^* = P_2^* = 0,5$, и энтропия объекта

$$H^*(S) = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}\right) = 1 \text{ бит.}$$

Информация, полученная в результате диагностирования, $J = H(S) - H^*(S) = 1 - 1 = 0$ бит. После получения сообщения от датчика давления вероятности состояний стали $P_1^* = 1$, $P_2^* = 0$,

и энтропия объекта $H^*(S) = -(1 \log_2 1 + 0 \log_2 0) = 0$ бит, т.е. информация, полученная в результате диагностирования, $J = H(S) - H^*(S) = 1 - 0 = 1$ бит.

Очевидно, второе сообщение (от датчика давления) несет больше информации, так как оно полностью устраняет неопределенность состояния объекта.

12.5. Архитектура компьютера

Общие сведения, касающиеся внутренней организации вычислительных систем, необходимы не только системным программистам, но и обычным пользователям ЭВМ. Эти сведения позволят определить или наметить принципы организации обработки информации так, чтобы эффективно использовать все возможности вычислительной системы. Они также позволяют определить производительность, которую можно получить на классе задач, интересующих пользователя, и тем самым дать ответ относительно пригодности той или иной вычислительной системы для решения поставленных задач.

Несмотря на то, что производительность вычислительных машин год от года растет в геометрической прогрессии, появляются новые задачи, для решения которых существующих возможностей отдельных процессоров недостаточно, и проблема разработки эффективных программ не теряет своей актуальности.

Современные мультипроцессорные и многомашинные вычислительные системы предоставляют возможности параллельной и распределенной обработки данных. Использование параллелизма требует поиска методов распараллеливания алгоритмов и, более того, поиска новых алгоритмов для решения задач. Тем самым знание структуры вычислительных систем может стимулировать чисто теоретические поиски новаций в вычислительной математике и в сфере обработки данных.

Удобное для пользователя взаимодействие с вычислительной системой требует больших накладных расходов, и эти

расходы приводят к увеличению времени решения целевой задачи. Системным программистам, разрабатывающим программное обеспечение, приходится все время решать задачу нахождения приемлемого компромисса между созданием удобств пользователю и эффективным использованием аппаратных средств.

Очевидно, что этого рода профессионалам совершенно необходимо знать принципы функционирования той аппаратной платформы, для которой они ведут свои разработки.

Однако и обычным прикладным программистам также необходимо иметь сведения об архитектуре той ЭВМ, для которой они готовят задачи. Разумеется, для них сведения об архитектуре не должны быть перегружены техническими деталями, но должны быть достаточными для правильной организации вычислений. Это позволит более эффективно использовать вычислительную технику, сократить время решения задач и в конечном счете будет содействовать ускорению научно-технического прогресса.

Под *архитектурой ЭВМ* понимают внутреннюю организацию вычислительной системы. Определение понятия архитектуры вычислительных машин может быть сформулировано так: это то, что должен знать о вычислительной машине программист. При таком определении может показаться, что программисту не так уж много надо знать.

Современные системы программирования скрывают от пользователя многие важные черты внутренней организации машины и не требуют от него их знания. Но без знания некоторых черт внутренней организации машины трудно разработать программу, оптимально использующую аппаратные возможности.

В понятие архитектуры обычно включают сведения:

- о составе и взаимодействии основных блоков вычислительной системы;
- о системе команд и их репертуаре;
- о регистровой структуре центральных процессоров;
- об организации памяти, иерархии запоминающих устройств и их взаимосвязи;
- об организации системы прерываний;

- об организации обмена данными с внешними устройствами;
- о топологии связей отдельных устройств и модулей. Уровень детализации перечисленных выше сведений зависит от уровня пользователя. Системному программисту необходима более детальная информация о внутреннем устройстве машины в сравнении с разработчиком прикладных программ.

Самая общая схема отдельно взятой вычислительной машины от калькулятора до суперкомпьютера содержит центральный процессор, запоминающее устройство, устройства ввода-вывода (рис. 12.10).

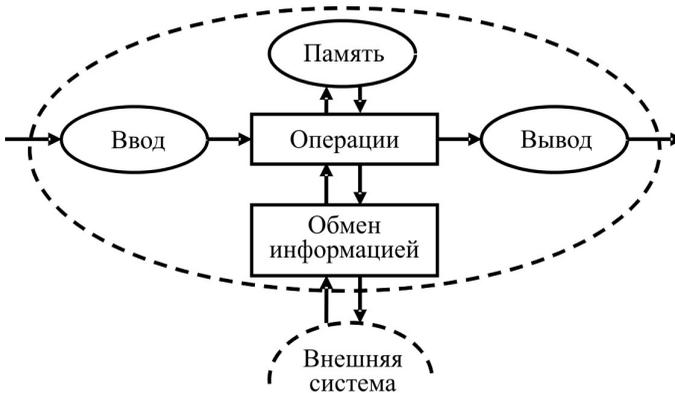


Рис. 12.10. Концептуальная схема компьютера

Центральный процессор получает команду из памяти или от внешних устройств, выполняет действия, указанные в команде, полученный результат помещает в память или выводит на внешние устройства и переходит к выбору и выполнению следующей команды.

В разных типах компьютеров выполнение этих основных функций организовано по-разному. Архитектура ЭВМ отвечает на вопрос об организации выполнения этих функций.

Вычислительные машины принято делить на несколько архитектурных типов. Классификация архитектур вычислительных

систем характеризует способы обработки потока команд и потока данных, а также взаимоотношение этих двух главных процессов. Под терминами «*поток команд*» и «*поток данных*» понимают последовательность команд в процессе их выполнения и последовательность обрабатываемых данных.

К типу ОКОД (Одиночный поток Команд, Одиночный поток Данных) относились почти все ЭВМ первого поколения. Этот тип архитектуры ЭВМ предполагает *строго последовательное выполнение команд над строго линейной общей памятью для команд и данных.*

По типу ОКМД (Одиночный поток Команд, Множественный поток Данных) проектируются мультипроцессорные вычислительные системы, состоящие из многих процессорных элементов, каждый из которых имеет свою локальную память. Все эти процессоры выполняют, как правило, синхронно одну и ту же команду, выбираемую из одного потока команд, над данными, выбираемыми каждым процессором из своей локальной памяти.

По типу МКМД с множественными потоками команд и данных функционируют мультипроцессорные асинхронно работающие вычислительные комплексы. Каждый процессор имеет собственную локальную память для данных и программ, при этом все процессоры связаны между собой линиями передачи данных. Этот тип архитектуры, пожалуй, наиболее представительный по числу различных вычислительных систем, которые могут быть к нему отнесены. Однако под МКМД-системами чаще всего подразумевают установки, предназначенные для решения одной большой задачи, разбитой на более-менее независимые части.

Тип МКОД многопроцессорной архитектуры объединяет несколько процессоров с общей памятью. При этом каждый процессор может работать по своей собственной программе, пользуясь общим полем данных (общей оперативной памятью).

С целью ознакомления со сложившейся терминологией, принятой в литературе, посвященной описаниям архитектуры вычислительных машин, рассмотрим принципиальную схему однопроцессорной вычислительной машины.

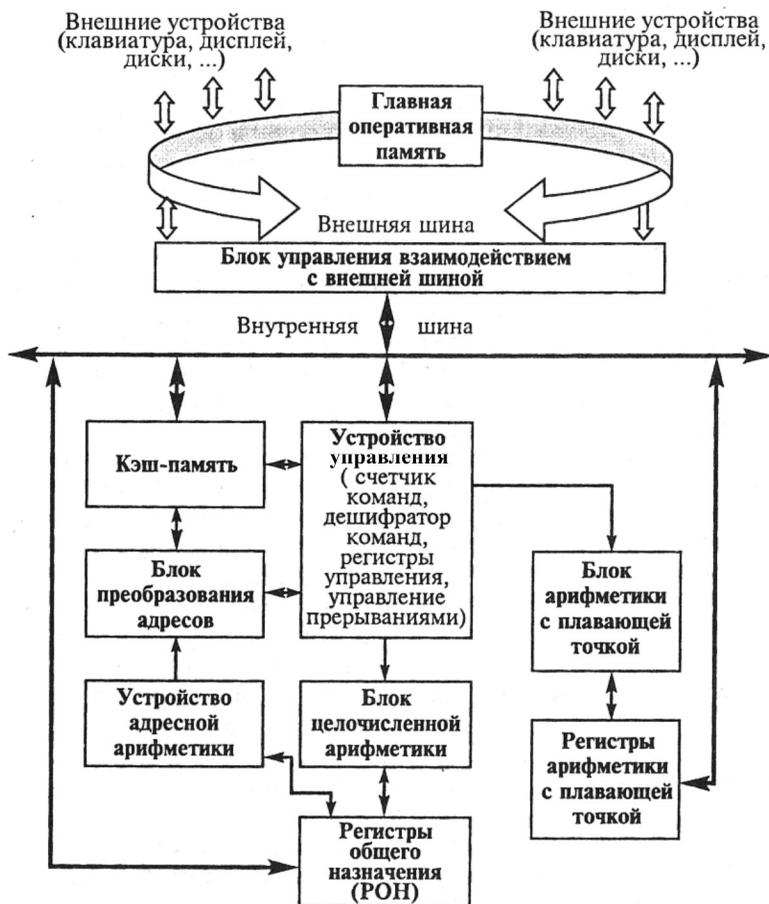


Рис. 12.11. Укрупненная схема микропроцессора

Приведенная на рис. 12.11 схема центральной части отдаленно напоминает структуру современных микропроцессоров, используемых в персональных компьютерах, включающих в свой состав такие узлы, как: оперативная память; кэш-память; регистры; внешние и внутренние шины; устройство управления; арифметические устройства; устройство адресной арифметики; блок преобразования адресов.

В этот набор не включены многие другие узлы и блоки, обеспечивающие работу компьютера с современным обязательным набором внешних устройств – диски, дисплеи, клавиатура, мыши и т.п.

Внешняя шина представляет собой сложное устройство, к которому могут подключаться: устройства процессора, оперативная память, внешние запоминающие устройства и устройства ввода-вывода, устройства взаимодействия с межкомпьютерной сетью. Шина обеспечивает взаимный обмен информацией между абонентами, подключенными к шине. Существует несколько типов шин, различающихся быстродействием, логикой работы, числом и правилами подключения абонентов к шине. В состав шины входят регистры, хранящие принимаемую и передаваемую информацию, и собственная система управления. Шина обеспечивает интерфейс процессора с внешними устройствами, получает от внутренних схем процессора заявки на прием и выдачу информации другим абонентам шины.

Одним из абонентов шины является *главная оперативная память*, из которой выбираются команды и числа. Для того чтобы выбрать из памяти, например, команду, следует по шине запросить готовность памяти к работе, послать в устройство управления памятью адрес байта, с которого начинается эта команда, принять эту команду на выходной регистр памяти и отправить в процессор. Приблизительно такие же операции производятся при выборке чисел и при записи информации из процессора в оперативную память.

В процессорах имеется также система *внутренних шин*, абонентами которой являются блоки самого процессора. По внутренним шинам передаются команды, операнды и адреса, т.е. осуществляется обмен содержательной информацией между внутренними регистрами и блоками процессора. Внутренние шины осуществляют передачу данных намного быстрее, чем внешняя шина, и от пропускной способности внутренних шин во многом зависит быстродействие процессора.

Кроме того, в процессоре существует сеть передачи управляющих сигналов, включающих в работу блоки и схемы процессора. *Устройство управления* – это сердце процессора, которое задает ритм работы всех его устройств и организует их согласованное взаимодействие. В состав устройства управления входит группа управляющих регистров, играющих важную роль в определении правил функционирования процессора.

К управляющим регистрам относится *счетчик команд*, в котором автоматически формируется адрес команды, подлежащей выполнению. Этот регистр назван счетчиком по той причине, что он действительно в основном работает как счетчик, автоматически прибавляя к себе некоторую константу для получения следующего по порядку адреса команды, подлежащей выполнению. Этот процесс изменения содержимого счетчика команд прерывают команды переходов, которые загружают счетчик новым содержимым – адресом перехода, взятым из такой команды или вычисленным другим способом. Полученный тем или иным способом адрес команды передается в оперативную память для выборки команды.

Важную роль играет *регистр управления*, характеризующий состояние процессора. В разрядах этого регистра можно задать указания о выполнении программы в пользовательском режиме или в режиме работы операционной системы, указания о блокировании прерываний, включении или отключении механизмов защиты памяти. В нем хранится информация о характере результата выполненной операции, необходимая для выполнения команд условных переходов и многое другое, в зависимости от типа и сложности процессора.

К устройству управления относится *дешифратор команд*, логически сложное устройство, которое на основе анализа поступившего из оперативной памяти кода команды формирует серию управляющих сигналов и данных, заставляющих работать соответствующие устройства процессора.

В процессе работы дешифратора команд из кода команды выделяется поле кода операции. В зависимости от значения выделенного кода производится дальнейший анализ. Выделяется адресное поле, в котором указываются правила формирования адресов операндов или адреса перехода и выдается соответствующий приказ в устройство адресной арифметики на формирование необходимого адреса.

Сформированный адрес передается в блок преобразования адресов и затем в оперативную память или в другие устройства процессора.

С устройством управления тесно связана электронная схема – *генератор тактовой частоты*, выдающая тактирующие импульсы. Многочисленные электронные схемы устройств процессора срабатывают только в дискретные моменты времени, связанные с появлением на входах этих схем тактирующих импульсов. Говорят о тактовой частоте процессора, имея в виду частоту появления тактовых импульсов. Эта частота измеряется в герцах (числом импульсов в одну секунду). Тактовая частота – важнейший технический параметр, определяющий быстродействие процессора. Различные операции для своего выполнения требуют различного количества тактов. Самые короткие, например логические операции, требуют одного такта, некоторые, например деление, могут потребовать более десяти тактов.

Существует несколько критериев определения производительности процессора, один из которых формулируется как среднее число тактов, необходимых для выполнения одной команды. В машинах середины прошлого столетия среднее число тактов для выполнения одной команды исчислялось десятками. В современных процессорах это число меньше единицы, т.е. за один такт могут выполняться несколько операций.

Первые ЭВМ, появившиеся в мире, работали на частоте в несколько килогерц (КГц), затем частота повысилась до сотен мегагерц (МГц). В настоящее время существуют процессоры,

работающие на частотах, приближающихся к миллиарду тактовых импульсов в секунду, гигагерц (ГГц).

Блок преобразования адресов выполняет работу по преобразованию логического (виртуального) адреса в адрес физической памяти. Компиляторы преобразуют исходную программу в последовательность команд машины (машинный код). В процессе этих преобразований компилятором вычисляются адреса памяти, по которым предполагается разместить переменные, константы и другие объекты программ. Эти адреса предполагаемого размещения называются логическими. Процесс реального размещения информации в физической памяти машины осуществляется непосредственно перед началом ее выполнения и требует преобразования логического адреса в физический адрес.

Этим занимается операционная система и аппаратура процессора. В современных машинах, способных работать в мультипрограммном режиме, возникает необходимость располагать в непересекающихся участках физической памяти несколько не связанных друг с другом программ и данных, которые подверглись независимой трансляции в машинный код. При независимой трансляции несвязанных друг с другом программ их логические адреса будут, как правило, пересекаться. Устройство преобразования адресов, зная расположение программных объектов в физической памяти, заботится о том, чтобы физические адреса данных и команд независимых программ не совпадали между собой.

Главную оперативную память чаще всего относят к центральному процессору как неотъемлемую часть любой вычислительной машины. Вычислительную машину можно определить как устройство, изменяющее состояние памяти, переводящее память из одного состояния в другое. В этом смысле память можно считать главным объектом работы вычислительной машины.

Кэш-память – промежуточное, скрытое от пользователя запоминающее устройство ассоциативного типа, позволяющее ми-

нимизировать время обращения к основной оперативной памяти и тем самым сократить общее время решения задач. Кэш-память более быстродействующее устройство, но значительно меньшее по объему по сравнению с оперативной памятью. Кэш-памятью снабжены почти все современные процессоры.

Блок целочисленной арифметики, или целочисленное *арифметико-логическое устройство* (АЛУ), выполняет арифметические и логические операции над целыми числами в двоичном представлении: сложение, вычитание, умножение, деление с остатком, логические операции, сдвиги.

Блок арифметики с плавающей точкой (АЛУ с плавающей точкой) выполняет арифметические операции с вещественными числами, представленными экспонентой и мантиссой. Диапазон вещественных чисел, представимых в машине, как правило, значительно больше количества представимых в машине целых чисел.

Регистры общего назначения (РОН) служат для промежуточного хранения операндов и результатов вычислений и непосредственно связаны с целочисленным АЛУ. Они также могут быть использованы для вычисления адресной информации. Если операнды находятся на быстрых регистрах общего назначения, то процесс вычислений ускоряется, так как нет необходимости обращаться в основную сравнительно медленную оперативную память.

Регистры арифметики с плавающей точкой служат для промежуточного хранения операндов и результатов вычислений над вещественными числами, представленными с плавающей точкой. Эти регистры обычно по разрядности в 2–3 раза длиннее быстрых регистров общего назначения. Это необходимо для достижения высокой точности выполнения арифметических операций над вещественными числами.

Вычислительные машины нетрадиционной архитектуры

По принципу организации обработки данных вычислительные системы делятся на две резко различающиеся категории. Большинство однопроцессорных и многопроцессорных машин управляется потоками команд. Архитектуры машин и вычислительных систем, управляемых потоками команд, принято называть *традиционными*. Однако существуют проекты и опытные образцы машин, управляемых потоком данных. Машины, управляемые потоком данных, принято относить к *нетрадиционным* архитектурам. Понятие традиционности относительно: то, что сегодня считается нетрадиционным, завтра может стать вполне обыденным.

Рассмотрим принципы функционирования таких *поточковых машин*. В них отсутствует как таковой счетчик команд, определяющий последовательность действий. Цель создания такого рода компьютеров состоит в том, чтобы достичь максимального распараллеливания вычислительного процесса. Теоретически такие машины позволяют максимально использовать параллелизм, заложенный в алгоритме.

Вычислительный алгоритм может быть представлен графом потока данных. Например, алгоритм вычисления выражения: $(a + b)d + (a + d)b + d \cdot b \cdot c$ представляется графом, в узлах которого расположены операции, выполняемые над переменными, поток которых указан на рис. 12.12 стрелками на дугах. В узлах графа, расположенных в столбцах, все операции могут быть выполнены независимо и одновременно.

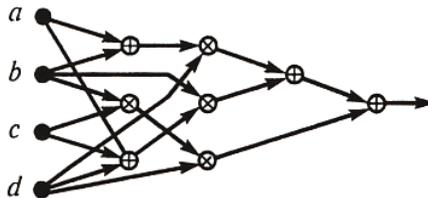


Рис. 12.12. Граф-схема потока данных

Принцип работы машин, управляемых данными (*Data Flow Computers – DFC*), состоит в том, что *всякое действие выполняется в тот момент, когда для него готовы операнды*. Очевидно, что машина, спроектированная точно под данный алгоритм, будет работать с максимальной производительностью, полностью используя внутренний параллелизм алгоритма.

Говоря об уровнях параллелизма, имеют в виду число узлов графа потока данных, в которых указанные действия могут выполняться одновременно.

Когда же идет речь о проектировании универсальной DFC-машины, возникает множество проблем, не поддающихся простому решению, поскольку универсальная потоковая машина, пригодная для выполнения любого алгоритма, теоретически должна содержать неограниченное число функциональных устройств. Сделать такую универсальную потоковую машину практически невозможно, так как число функциональных устройств будет ограничено соображениями технической реализуемости и стоимости. Информация о графе оказывается весьма емкой по объему, и ее необходимо иметь на каждом шаге вычислительного процесса. Этот недостаток потоковых машин, требующий передавать большой дополнительный объем информации от одного узла к другому, сводит на нет возможность достижения максимального параллелизма.

Нейрокомпьютеры

Особое место в вычислительной технике занимают *нейрокомпьютеры*, которые все более и более широко используются для решения задач распознавания образов, классификации, поиска экстремумов, решения задач искусственного интеллекта.

Нейрокомпьютер, или нейроподобная вычислительная система, представляет собой *сеть элементарных процессоров* (ЭП), которые, следуя изначальной биологической терминологии, называются нейронами. Нейроны объединены в однослой-

ную или многослойную структуру. В идеальном случае все нейроны связаны между собой линиями передачи данных по принципу «каждый с каждым».

Нейрон, получив данные от других нейронов, с которыми он связан, выполняет вычисления некоторой функции S , называемой функцией активации. Ее значениями в простейшем случае могут быть 0 либо 1.

Связям, идущим от других нейронов к некоторому i -му нейрону, присваиваются весовые коэффициенты w_{ki} . Эти весовые коэффициенты, умноженные на значения S_k , поступающие от других нейронов, суммируются, и полученная сумма становится аргументом функции активации, вычисляемой принимающим нейроном. В каждый момент времени все нейроны работают одновременно, обеспечивая максимальный параллелизм нейрокомпьютерной системы.

Схематически один i -й ЭП нейросети, или нейрон, можно представить так, как это показано на рис. 12.13. Здесь использованы обозначения: w_{mi} – вес, на который умножаются значения, поступающие на i -й ЭП от m -го ЭП. Величины $X_{in} = S_i(x_i)$ обозначают значение функции S_i , рассылаемое другим нейронам, связанным с данным. Вес назначается при подготовке (настройке) нейросети к работе. Эта настройка происходит по некоторым алгоритмам «обучения» сети.

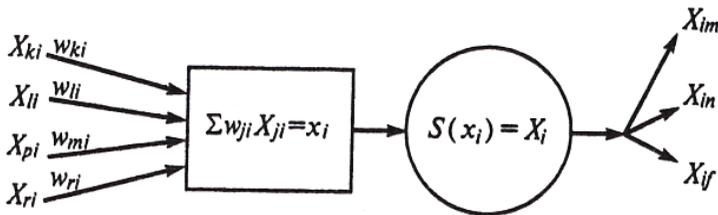


Рис. 12.13. Схема нейрона
(процессорного элемента нейрокомпьютера)

В качестве функции $S_i(x_i) = X_i$ чаще используют бинарную пороговую функцию, принимающую два значения: 0 или 1. Поро-

говая функция, ее называют еще *функцией активации*, переводит нейрон в два состояния – активное или пассивное.

Нейрокомпьютерная сеть обычно работает по тактам (дискретно). За такт происходит вычисление одновременно всех пороговых функций, при этом новые значения этих функций будут участвовать в следующем такте работы нейронной сети. Нейрон, принявший значение 0, становится пассивным. Это равносильно тому, что он перестает передавать информацию другим, связанным с ним нейронам. Во многих нейросетях функции активации полагают одинаковыми для всех нейронов и могут быть приняты линейными, в виде гиперболического тангенса и т.д.

Считается, что сеть пришла в равновесное состояние, когда перестают изменяться значения выходов. Это эквивалентно тому, что некий итерационный процесс сошелся. Распределяя нужным образом веса связей и варьируя правила вычисления функций всех узлов, можно настраивать такую сеть на решение широкого круга задач указанных выше классов.

Нейронные сети можно «обучать», автоматически изменяя вес и значения порогов таким образом, чтобы на обучающих примерах сеть давала правильный ответ на поставленный вопрос. В этом случае можно надеяться на то, что сеть будет давать правильные ответы на аналогичные вопросы, касающиеся других объектов с другими параметрами.

Для нейрокомпьютеров исключаются привычные понятия программирования. Процесс обучения сети заменяет программирование в обычных машинах. Он, как правило, требует больших затрат времени, но легко поддается автоматизации с использованием обычных компьютеров. Обученная нейросеть затем очень быстро находит решения поставленных задач и намного превосходит компьютеры традиционной архитектуры по эффективности работы. Наиболее широко нейрокомпьютеры используются для задач распознавания образов.

В настоящее время интенсивно исследуются возможности применения совершенно новых принципов, на основе которых можно создавать вычислительные системы и машины.

Ведутся работы по применению биохимических реакций для создания так называемых *биокомпьютеров*. Теоретически доказана возможность создания на этой основе универсального вычислителя.

Многообещающей является идея создания *квантовых компьютеров* на основе использования свойств атомов, то есть квантовой машины, основным логическим элементом которой будет отдельный атом. Такая машина должна обладать быстродействием, достаточным для расшифровки самых сложных кодов в считанные секунды, в то время как на самых быстродействующих современных машинах на это потребуются сотни лет. Новые принципы требуют нового взгляда на архитектуру компьютеров начавшегося тысячелетия, которая будет резко отличаться от архитектур современных вычислительных систем, построенных на традиционной элементной базе. Оптические компьютеры, основанные на квантовых эффектах, идеях голографии будут отличаться глубоким параллелизмом, вытекающим из естественного параллелизма физических процессов, лежащих в основе их работы.

Взаимоотношение классов архитектур представлено на рис. 12.14.

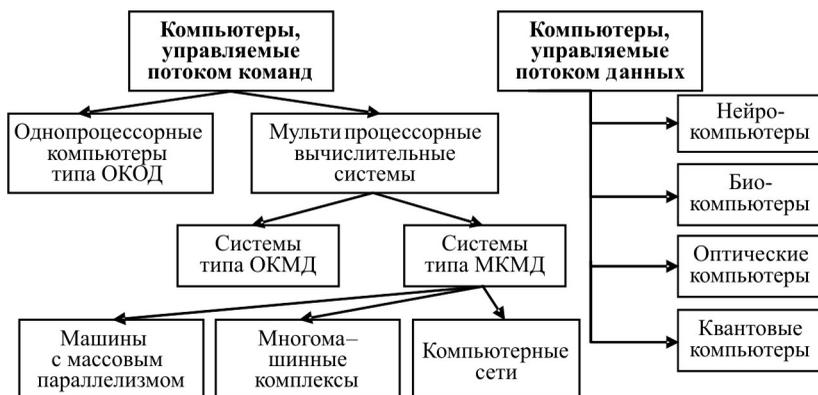


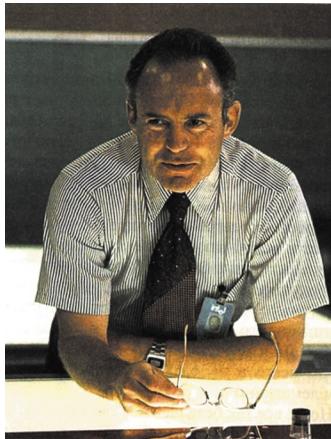
Рис. 12.14. Иерархия архитектур

12.6. Предельные возможности электронной компьютерной техники

В 1965 году вышла знаменитая теперь статья Гордона Мура «Переполнение числа элементов на интегральных схемах» в которой тогдашний директор отдела разработок компании *Fairchild Semiconductors* и будущий сооснователь корпорации *Intel* дал прогноз развития микроэлектроники на ближайшие десять лет. Он предсказал, что количество элементов на кристаллах электронных микросхем будет и далее удваиваться каждый год.

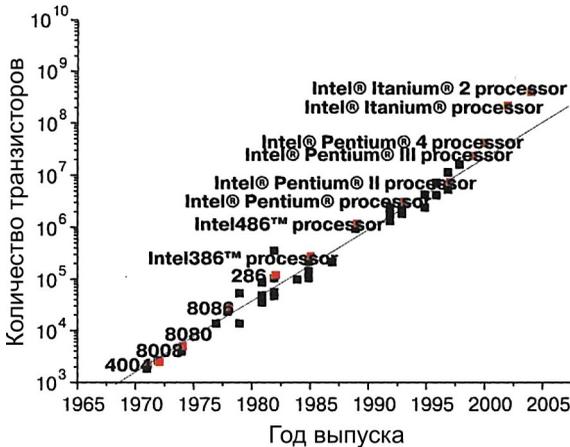
Позднее, в 1975 году, Гордон Мур отметил, что за прошедшее десятилетие количество элементов на кристаллах действительно удваивалось каждый год, однако в будущем, когда сложность чипов возрастёт, удвоение числа транзисторов в микросхемах будет происходить каждые два года. Это новое предсказание также сбылось, и закон Мура продолжает в этом виде (удвоение за два года) действовать и поныне, что можно наглядно видеть по графикам на рис. 12.15. Этот закон подтверждается и для тактовой частоты микропроцессоров, хотя Гордон Мур неоднократно утверждал, что его закон относится только к числу транзисторов на кристалле и отражает общие для многих процессов экспоненциальные закономерности развития.

В связи с бурным развитием микроэлектроники возникают вопросы. Можно ли ожидать появления в 2020 году электронного микропроцессора с тактовой частотой 100 ГГц? Возможно ли уменьшение электрического напряжения на ядре микропроцессора и, соответственно, уменьшение теплоотдачи? Например, производство памяти и процессоров по технологии 65 наномет-

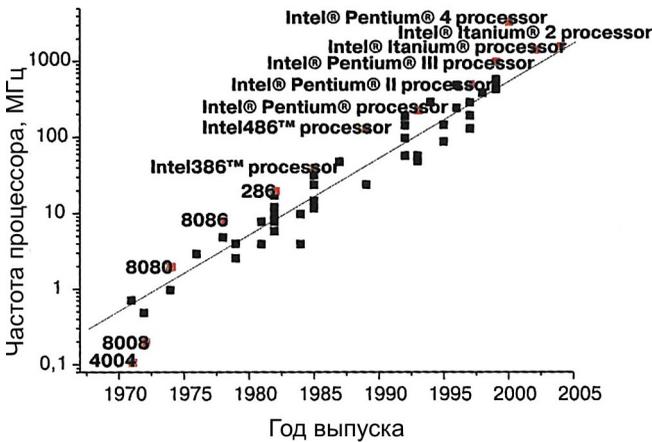


Гордон Мур
(*Gordon Moore*)

ров уже отлажено, на 2009 год намечено внедрение 32-нанометрового, а в 2011 году настанет черёд технологического процесса 22 нм. Но что ожидается после 2011 года? Станет ли возможным создание транзисторов по технологиям 16, 11 и 8 нм с длинами затворов транзисторов 7, 5 и 3 нм соответственно?



a



b

Рис. 12.15. Темпы увеличения плотности транзисторов (*a*) и частоты процессоров Intel (*b*)

Возможности миниатюризации электроники ограничены тремя фундаментальными законами. Первый из них ограничивает минимально возможное напряжение в транзисторе 0,1 В, ниже которого наступает нестабильная работа, вызванная тепловыми шумами.

Второй фундаментальный предел по энергии электронной ячейки может быть получен из квантовой механики, из гейзенберговского соотношения неопределенности, который может интерпретироваться следующим образом: физическое изменение энергии квантовой системы связано со временем переключения Δt следующим неравенством:

$$\Delta E \geq \frac{\hbar}{\Delta t}, \quad (12.6)$$

где $\hbar = 1,055 \cdot 10^{-34}$ Дж·с – постоянная Планка. Соответственно, для мощности переключения можно записать:

$$\Delta P \geq \frac{\hbar}{(\Delta t)^2}. \quad (12.7)$$

Третий фундаментальный предел связан со скоростью распространения электрического импульса по микрочипу, которая не может быть больше скорости света в вакууме. Отсюда следует ограничение на предельную тактовую частоту микропроцессора размером 1×1 см², которое составляет около 22 ГГц.

Ограничивает тактовую частоту и электрическая емкость системы, возрастающая с увеличением числа элементов на микросхеме.

Увеличение тактовой частоты возможно при размещении на чипе нескольких синхронизированных тактовых генераторов, но это приводит к усложнению архитектуры микропроцессора. Современный процессор Pentium IV, содержащий 55 млн транзисторов, работает с тактовой частотой 3 ГГц при потребляемой мощности около 100 Вт.

Потребляемая и соответственно выделяемая процессом энергия определяется не только *активными потерями*, связанными с информационным представлением, но также и с различными *пассивными потерями* в электронных схемах. К таким паразитным эффектам относится диссипация энергии на сопротивлениях переключателей и соединений из-за утечки тока.

Физические механизмы, обуславливающие эти утечки тока, определяются туннельным эффектом, термоэлектрической эмиссией, генерацией носителей в области пространственного заряда и другими физическими эффектами.

Уменьшение характерного размера микросхем приводит только к увеличению пассивных потерь, причем их рост происходит быстрее активных, что наглядно представлено на рис. 12.16. Энергия активных и пассивных потерь полностью диссипируется в электронных логических ячейках и порождает проблемы теплоотвода.

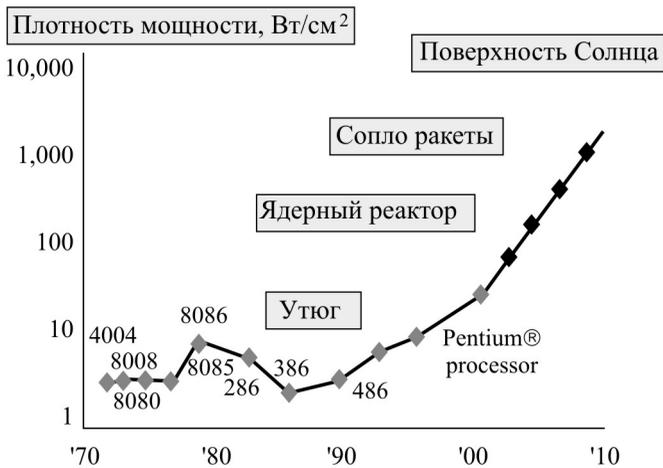


Рис. 12.16. Рост потерь мощности в полупроводниковых процессорах

Таким образом, основные ограничения электронной информационной технологии сводятся к следующим. С ростом чис-

ла элементов на микросхеме увеличивается *электрическая емкость системы*, препятствующая увеличению тактовой частоты; растет число межсоединений и, соответственно, увеличивается *время задержки прохода сигнала* между макроструктурами процессора; возрастают *активные и пассивные потери*, что приводит к нагреву системы и проблеме отвода тепла.

12.7. Оптические системы обработки информации

В основе оптических методов обработки информации лежат явления преобразования пространственно-модулированных оптических сигналов в оптических устройствах и системах на принципах как геометрической, так и волновой оптики. Оптическая обработка информации осуществляется в *оптическом процессоре* – аналоговом либо оптоэлектронном устройстве, определенным образом изменяющем амплитуду и фазу *пространственно-модулированного* оптического сигнала, содержащего информацию об объекте. Системы оптической обработки информации являются составной частью *оптического компьютера*.

Волновая и корпускулярная природа света обуславливает многочисленные *преимущества оптических систем* для задач передачи, хранения и обработки информации:

- частота оптического излучения составляет $10^{12} \dots 10^{16}$ Гц, что позволяет создать 10^4 информационных каналов со спектральной шириной 100 ГГц;

- передача информации происходит действительно со скоростью света $c = 3 \cdot 10^{10}$ см/с;

- большое число световых пучков могут свободно проходить по одной и той же области пространства, пересекаться и не влиять друг на друга;

- использование двумерного (изображения) и трехмерного (голограммы) характера световых полей;

- параллельная передача и обработка информации с одновременной работой на различных длинах волн;

- когерентная обработка оптической информации с использованием фазовых соотношений;
- два состояния поляризации (горизонтальная и вертикальная или круговая, по левому или правому кругу) увеличивают вдвое объем переносимой информации;
- оптическая система ничего не излучает во внешнюю среду, обеспечивая защиту от перехвата информации, и нечувствительна к электромагнитным помехам.

Модуляция света

Модуляция света включает в себя управление параметрами световых потоков: *амплитудой, частотой, фазой, поляризацией*. Простейшим амплитудным модулятором света является механическая заслонка, однако быстдействие и надежность ее невелики. В основе действия современных пространственно-временных модуляторов света лежат различные температурные, электрооптические, магнитооптические, акустооптические и другие эффекты – явления Поккельса, Керра, Фарадея, дифракция Брэггов и др. Модуляция светового потока осуществляется *внешним электрическим сигналом*. Одни оптические модуляторы изменяют яркость света на выходе, другие – направление распространения излучения. В первом случае оптические модуляторы применяются в сверхбыстродействующей оптической связи (несколько Гбит/с и выше) для модуляции излучения полупроводниковых лазеров. Во втором случае модуляторы используются в лазерных печатающих устройствах и устройствах лазерного сканирования.

В качестве примера рассмотрим *фазовую модуляцию света* на основе эффекта Поккельса. В этом эффекте используется зависимость показателя преломления несимметричного кристалла n от величины внешнего электрического поля E :

$$n = n_0 + \frac{1}{2} n_0^3 r E, \quad (12.8)$$

где n_0 – показатель преломления кристалла в отсутствие внешнего поля; r – электрооптический коэффициент, зависящий от свойств и ориентации кристалла, направления поля E и поляризации проходящего света.

Световой пучок, прошедший путь l в кристалле, помещенном в электрическое поле, приобретает фазовый сдвиг:

$$\varphi = 2\pi nl/\lambda = \varphi_0 + \pi n_0^3 r E l / \lambda, \quad (12.9)$$

где λ – длина волны света в вакууме; φ_0 – начальный сдвиг фаз, приобретенный светом при прохождении кристалла в отсутствие поля. Наличие фазового сдвига, вызванного внешним полем, и означает фазовую модуляцию света. Инерционность этого эффекта мала и позволяет изменять фазу света за время $\sim 10^{-12}$ с.

На рис. 12.17 показан один из распространенных вариантов модулятора света, называемый модифицированным интерферометром Маха–Цендера, изготовленный в пленочном исполнении из ниобата лития (LiNbO_3). На поверхности этого материала методом диффузии титана создают необходимую световую конфигурацию. Между световодами напыляют электроды. В структуре сформировано два разделенных световода, показатель преломления которых меняется при изменении направления приложенного электрического поля на обратное. В результате появляются соответствующий сдвиг фаз и интерференция световых пучков. Чтобы

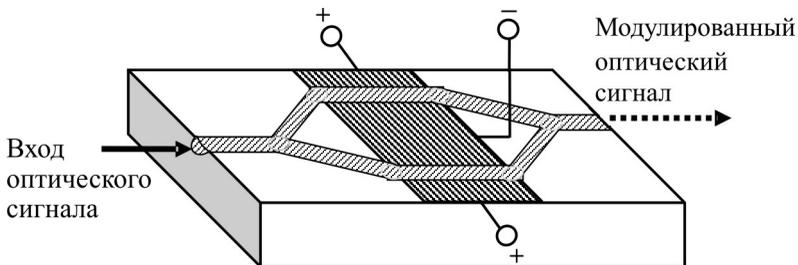


Рис. 12.17. Оптический модулятор интерференционного типа с электрооптическим кристаллом

световод был одномодовым, его ширина не должна превышать несколько мкм. При подаваемом переменном полуволновом напряжении 0,3 В этот модулятор обеспечивает высокую частоту модуляции порядка 18 ГГц.

Оптические устройства хранения информации

По аналогии с обычной письменностью первые устройства хранения информации использовали бумажные или картонные носители – так называемые *перфокарты* и *перфоленты* (рис. 12.18).

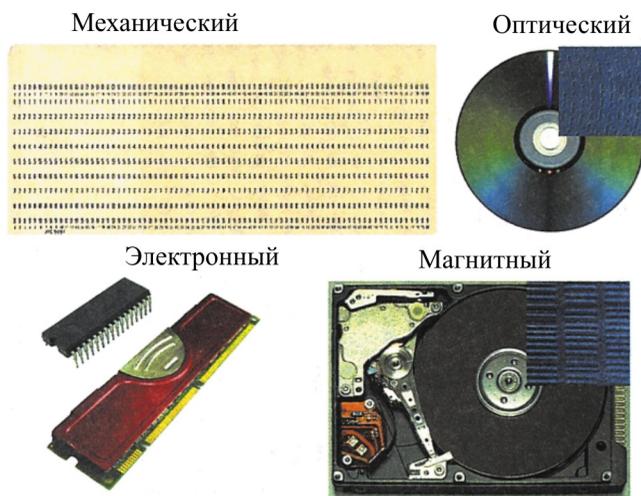


Рис. 12.18. Эволюция устройств записи информации

Первым цифровым носителем информации стал *магнитный дисковый накопитель* (IBM RAMAC, 1956 г.), являвшийся компромиссным решением между магнитной лентой и граммофонной пластинкой.

Основным конкурентом устройств магнитной записи на рынке являются оптические диски. В 1982 году фирмы *Sony* и *Philips* завершили работу над форматом CD-аудио (*Compact Disk*), открыв тем самым эру *цифровых носителей на компакт-дисках*.

Чтение и запись информации в этом случае осуществляется лазером с длиной волны от 780 нм для CD и 650 нм для DVD до 405 нм для новых дисков DVD BR (от англ. *blue ray* – голубой луч). Максимальный объем информации для оптических дисков составляет от 720 Мбайт (CD) до 17 Гбайт (DVD), а также 100 Гбайт для *blue-ray*.

В 1988 году компания *Intel* разработала еще один способ хранения данных на основе микросхем Flash-памяти, запоминающая ячейка которой представляет собой транзистор с двумя изолированными затворами: управляющим и плавающим, способным удерживать электроны, то есть заряд. Низкий заряд на плавающем затворе соответствует логической единице, а высокий – нулю. При чтении эти состояния распознаются путем измерения порогового напряжения транзистора. При стирании с флэшки какого-либо файла на управляющий затвор подается высокое отрицательное напряжение и электроны с плавающего затвора переходят (туннелируют) на исток.

Наконец, совсем недавно создатель сканирующего туннельного микроскопа и сотрудник исследовательского института ИВМ, нобелевский лауреат Герд Биннинг предложил вернуться к механическому принципу хранения информации, разработав технологию *Millipede* (англ. «многоножка») (рис. 12.19). Биннинг обратил внимание на возможность формирования наноразмерных «ямок» с помощью наноиндентации поверхности полимера зондом атомно-силового микроскопа. Наличие или отсутствие такой «ямки» в определенной точке поверхности можно трактовать, как единичное значение бита памяти. Для ускорения работы устройства наноиндентация (запись) и сканирование поверхности полимера (считывание) может производиться не одним *кантилевером*, а целой *матрицей зондов*.

В марте 2005 года на выставке в Ганновере ИВМ представила работающий чип с плотностью хранения информации 153 Гбайт на площади в квадратный дюйм (1 дюйм = 2,54 см). Следующие поколения устройств *Millipede*, как обещают иссле-

дователи IBM, будут иметь в 100 раз большую емкость, что позволит хранить огромное количество информации на крошечных чипах. По прогнозам IBM, это новое устройство хранения данных должно вытеснить с рынка чипы *flash*-памяти.

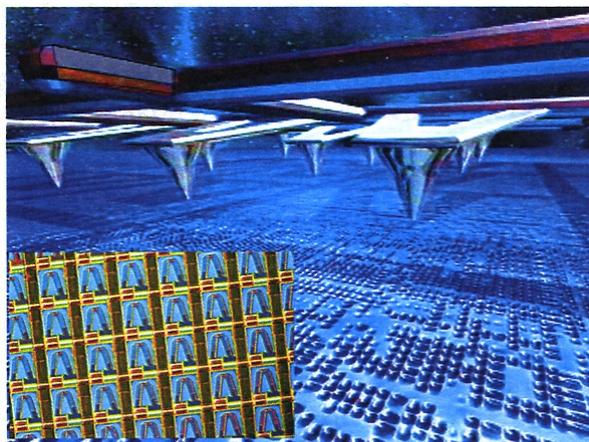


Рис. 12.19. Изображение устройства записи *Millipede* (чтение информации производится «ощупыванием» поверхности большим набором зондов)

Японская компания TDK разработала технологию, которая позволяет уместить на болванке до 200 гигабайт данных. Израильская компания *Mempile* разработала технологию *TeraDisc*, которая позволяет записать на диск, изготовленный из полимера, сходного с плексигласом, до 5 терабайта данных.

Оптические устройства хранения информации (оптическая память) потенциально обладают весьма высокой информационной емкостью или объемной плотностью записи информации благодаря малым размерам элементарной ячейки записи информации, исчисляемым в единицах – долях длины световой волны. Электронная информация в двоичном коде фиксируется на физическом носителе (оптическом диске) вдоль спиральной траектории записи в виде элементарных ячеек – областей высо-

кого либо низкого коэффициентов отражения света – размерами около $0,5 \times 3$ мкм. Период спирали 1,6 мкм, скорость вращения диска до 2000 об/мин, типичный диаметр диска 120 мм. Запись информации производится экспонированием специального материала – фоторезиста. Для считывания информации обычно используются полупроводниковые излучатели инфракрасного диапазона 0,8–0,9 мкм.

На рис. 12.20 представлен общий вид видеодиска, а на рис. 12.21 – структура и организация записывающей поверхности. Однако расстояние между записывающими дорожками составляет всего 1,6 мкм. Дорожки на поверхности расположены по спирали, раскручивающейся изнутри наружу.

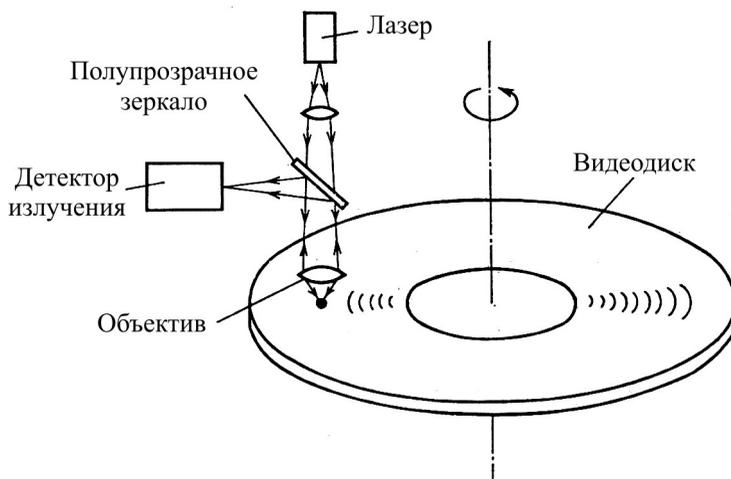


Рис. 12.20. Схема записи на видеодиск

На этом практически исчерпываются все возможности повышения плотности записи информации в однослойном и многослойном рельефно-фазовом виде.

Дальнейший прогресс в разработке дисковых систем оптической памяти разработчики связывают с использованием метода оптической голографии, позволяющего использовать не только

поверхность, но и весь объем диска. Голография (от греч. *hólos* – весь, полный и *gráphō* – пишу) – метод получения объемного изображения объекта путем регистрации и последующего восстановления волновых полей, изобретенный английским физиком венгерского происхождения Д. Габором в 1948 году. Этот метод открывает новые возможности при аналоговой обработке и хранении информации.

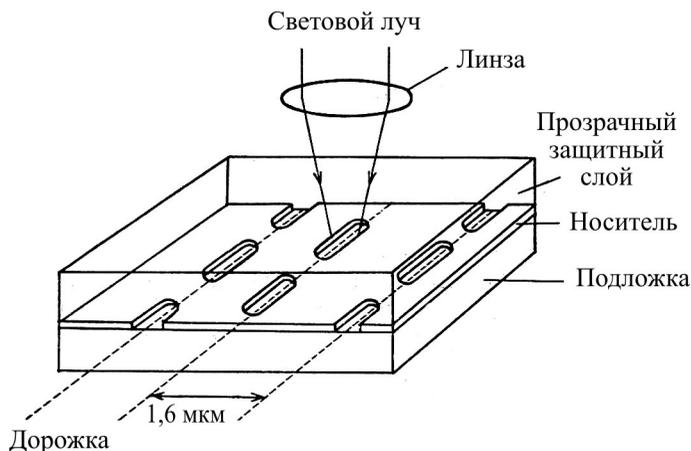


Рис. 12.21. Увеличенное изображение записывающей поверхности видеодиска

На рис. 12.22 изображен принцип записи информации при помощи средств голографии. Источником света служит лазер. Лазерный луч расщепляется на *опорный* и *предметный* лучи, которые используются для освещения объекта. На голограмме получается сложная комбинация световых волн в результате интерференции световых потоков. Полученное изображение на голографической пластинке внешне совершенно не похоже на объект записи. Для воспроизведения исходного объекта, или, как говорится, для регенерации основного объекта, пластинку с голограммой нужно осветить *опорным* лучом, полученным от лазерного источника.

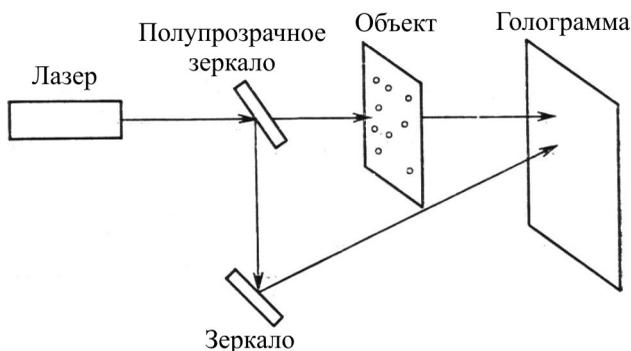


Рис. 12.22. Принцип голографической записи данных

Идея использования метода голографии для создания систем хранения информации была опубликована Питером ван Херденом, сотрудником компании «Поляроид», еще в 1963 году. В качестве физического предела плотности хранения информации в голографических системах им была определена плотность в $1/\lambda^3$ бит/м³.

Основным сдерживающим фактором в развитии голографических систем памяти долгие годы являлось отсутствие регистрирующей среды, пригодной для мультиплексной объемной голографической записи. Другим фактором, сдерживающим использование голографических запоминающих устройств в широких масштабах, является разработка средств, обеспечивающих работу и управление голографических запоминающих устройств с высокой скоростью и разрешающей способностью.

Ситуация изменилась в конце девяностых годов с разработкой объемных фотополимерных сред для объемной голографической записи. Так, американская компания *InPhase Technologies* сообщила о совместной разработке с японской фирмой *Hitachi Maxell Ltd* голографического диска, размером с диск DVD, способного хранить 300 Гбайт информации, что примерно в 60 раз превышает емкость диска DVD, и обладающего скоростью доступа к информации, в 10 раз превышающей скорость, реализуемую в DVD-сис-

темах. Компания предполагает выпускать линейку систем голографической дисковой памяти с емкостями до 1,6 Тбайт при скоростях доступа к информации 120 Мбит в секунду.

По заключению разработчиков, технология голографической записи информации будет принята в качестве основной при создании следующих поколений относительно простых и недорогих систем голографической дисковой памяти. Были анонсированы принципы создания первых голографических HVD (*Holographic Versatile Discs*) дисков емкостью до 4 Тбайт.

Каждый год в согласии с законом Мура средняя плотность записи всех типов устройств возрастает в ~1,5 раза. Какая же из технологий является оптимальной и будет доминировать на рынке через 10 лет? Время покажет...

Оптический компьютер

В современном электронном компьютере можно отметить следующие оптические узлы и элементы:

- устройства ввода информации – оптический сканер, оптическая мышь;
- устройства обмена информацией – инфракрасный порт, оптоволокно;
- устройства вывода информации – лазерный принтер, дисплей, голографический (объемный) дисплей;
- устройства памяти – долговременная память на перезаписываемых оптических дисках, магнитооптические диски, голографические диски.

В 2003 году фирмой *Lenslet* (Израиль) был представлен первый коммерческий оптический процессор *Enlight 256*, способный производить 8 Тера операций с плавающей запятой в секунду.

В отличие от цифрового способа обработки информации оптические технологии допускают аналоговое представление света с непрерывно изменяющейся интенсивностью. В аналоговых оптических компьютерах помимо непрерывно изменяю-

щейся интенсивности света объектом являются все точки непрерывных координат, а не дискретное (точечное) представление всей информации в окружающем пространстве.

Рассмотрим основные методы аналоговых вычислений, производимых в аналоговых оптических компьютерах, с использованием основных законов оптики. На рис. 12.23, *a*, *б* показано, как с помощью светового луча можно выполнять операции сложения и умножения, собирая три луча в одной точке и измеряя их суммарную интенсивность. Фокусировка лучей облегчается линзой. Сложение когерентных световых пучков возможно с высокой точностью с учетом фазовых характеристик, поскольку складываются не интенсивности, а амплитуды падающих волн.

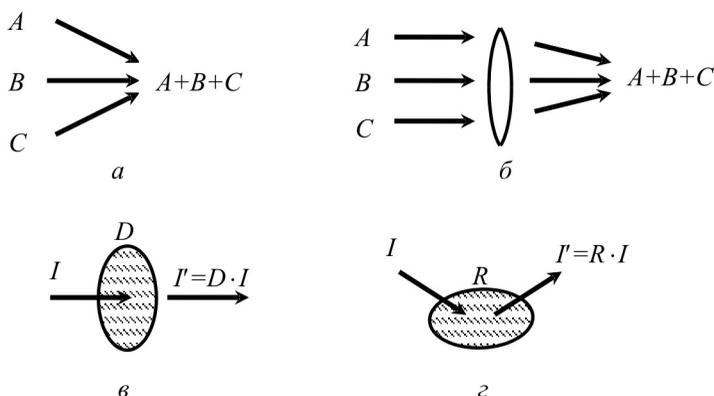


Рис. 12.23. Основные аналоговые оптические операции:
a – сложение; *б* – сложение с помощью линзы;
в – умножение на основе эффекта пропускания света;
г – умножение на основе эффекта отражения света.

Теперь рассмотрим операции умножения и деления. Такие операции возможны с использованием оптических элементов с управляемой прозрачностью D (отношением интенсивности выходного светового луча I' к интенсивности входного I), например с использованием свойств жидких кристаллов, либо

с управляемым коэффициентом отражения R на основе свойств полупроводников.

Основными операциями аналогового оптического компьютера являются только две операции – сложение и умножение, однако одни лишь эти операции не позволяют выполнять сложные вычисления. Отличительным свойством света является способность его равномерного распространения во всех направлениях. Благодаря этой особенности появляется возможность параллельной (одновременной) обработки больших объемов информации, и в этом смысле применение оптического компьютера имеет большое значение, однако для этого недостаточно использовать лишь свойство прямолинейного распространения света. Необходимо воспользоваться другими свойствами света, такими как преломление (рефракция) и дифракция, лежащими в основе работы линз и дифракционных элементов (решеток, голограмм).

Рассмотрим свойства линз, которые играют центральную роль в аналоговых оптических компьютерах. Тонкая линза представляет собой простейший оптический прибор с двумя сферическими поверхностями. Как показано на рис. 12.24, если перед линзой, например с левой стороны (входная плоскость), поместить некоторый предмет, то с противоположной стороны мы получим перевернутое и уменьшенное изображение того же предмета. Аналоговая операция *инвертирования* и *масштабирования* произведена со скоростью света – попробуйте провести такую же операцию с данным оптическим изображением (имеющим размер 6×12 см, обладающим 60000×120000 элементами разрешения, что соответствует минимально 7,2 Гбайт с использованием программы, например *Adobe Photoshop*). Таким образом, когерентные оптические системы могут быть эффективно использованы для решения широкого круга задач, связанных с получением, преобразованием и обработкой визуальной информации.

Поскольку оптические сигналы реализуются в виде реальных физических сигналов с помощью простейшей оптической системы, над ними можно производить различные математиче-

ские операции методами пространственной фильтрации. Оптическая система обработки информации методами пространственной фильтрации состоит из следующих компонентов: источника света, когерентного аналогового процессора, реализующего матричное преобразование информации, устройства ввода информации, пространственного операционного фильтра и детектора выходных сигналов.

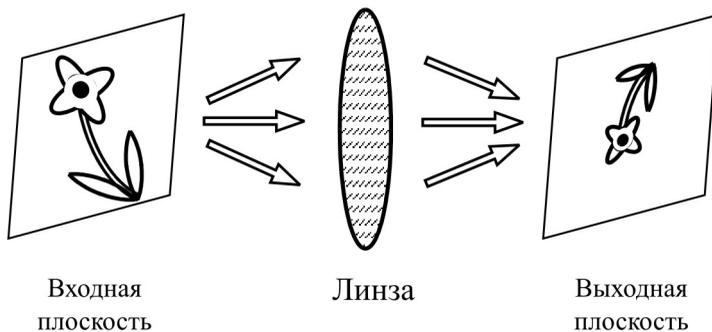


Рис. 12.24. Операции инвертирования и масштабирования, выполняемые линзой при построении изображения

На практике часто приходится иметь дело с одномерными сигналами. Процессор, способный параллельно обрабатывать множество одномерных сигналов, называется *астигматическим*.

Число элементов вектора и матрицы не обязательно может равняться одному, двум или трем, оно может быть любым. Возможность параллельной (одновременной) обработки множества данных можно назвать классической отличительной чертой аналогового оптического компьютера.

Характерной особенностью оптического сигнала как носителя информации является его двумерность. При малой длине волны света ($\lambda < 1$ мкм) размеры участка изображения, передаваемого с помощью волны, не могут быть меньше λ^2 , что позволяет передавать по лучу сечением 1 см^2 до 10^8 бит информации параллельно.

Ввод информации в световой луч осуществляется с помощью *модулятора света*, который изменяет ее, используя, на-

пример, преобразование Фурье. В результате оптическая линза дает фурье-спектр оптического изображения, падающего на эту линзу. Вводя соответствующий фильтр в фокальную плоскость после линзы, можно улучшить качество изображения или даже увидеть изображение невидимого объекта.

На рис. 12.25 представлена схема оптического процессора, предназначенного для *распознавания образов*. Точечный источник света в фокусе линзы L_1 освещает плоской волной набор распознаваемых картинок в плоскости P_1 . Фронт световой волны после P_1 искажается и попадает в плоскость P_2 , являющуюся согласованным фильтром для искомой двумерной картинки. Фильтр обладает свойством компенсации искажений волнового фронта, если падающая на него волна является двумерным фурье-спектром от искомой картинки. В результате волна снова становится квазиплоской и собирается линзой L_3 в светящееся пятно в фокусе P_3 . При несовпадении картинок свет разбрасывается по всей плоскости P_3 .

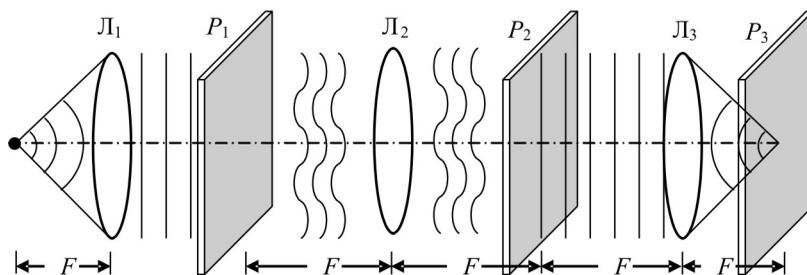


Рис. 12.25. Схема оптического процессора

Такая схема позволяет решать задачи оптической обработки информации с большой скоростью, ограниченной только скоростью ввода информации в плоскости P_1 и P_2 и скоростью вывода информации из плоскости P_3 .

Принцип параллельной обработки сигналов впервые был предложен в 1975 году в Станфордском университете (США) и

лег в основу многих оптических информационных устройств, разработанных впоследствии, в частности в основу первого коммерческого цифрового оптического компьютера *Enlight256*.

Оптический процессор *Enlight256* по принципу действия является аналоговым оптическим вычислительным устройством и аппаратно представляет собой развитую гибридную цифроаналоговую систему. Производительность процессора *Enlight256* составляет $8 \cdot 10^{12}$ операций в секунду: за один такт (8 нс) процессор выполняет каноническую операцию в вычислительной математике умножения 256×256 -байтной матрицы на 256 -байтный вектор.

Ядро процессора *Enlight256* – оптическое, а входная и выходная информация представляется в электронном виде. Ядро состоит из 256 лазеров, пространственного модулятора света, набора линз и приемников излучения, образующих оптическую матрицу VMM (*Vector-Matrix Multiplication*), которая конвертирует электрическую информацию в свет, затем производит необходимые преобразования этой информации, направляя свет через программируемую внутреннюю оптику. Выходное излучение регистрируется приемниками и преобразуется снова в электрический сигнал.

EnLight256 уже сейчас используется для задач, требующих высокой производительности. В частности, один процессор такого типа способен в реальном времени обрабатывать до 15 видеоканалов, может использоваться для распознавания голоса, человеческих лиц, обработки изображений и т.д. *EnLight256* идеально подходит для применения в военных радарх высокого разрешения для обработки данных от массивов антенн.

В 1984 году Б. Дженкинс из университета Южной Калифорнии продемонстрировал первый цифровой оптический компьютер, выполнявший достаточно сложную последовательность команд. Взаимодействие двух лучей осуществлялось элементом, состоящим из жидкого кристалла и фотопроводника. Свет, проходя, влияет на электрическое поле, приложенное к жидкому

кристаллу, отчего меняется прозрачность элемента. Быстродействие определяется инерционностью жидкого кристалла.

В 80-е годы прошлого века интенсивно работали над созданием полностью оптических компьютеров нового поколения. Сердцем такого компьютера должен был стать оптический процессор, использующий элементы, в которых *свет управляет светом*, а логические операции осуществляются в процессе взаимодействия световых волн с веществом. Значительные усилия, направленные на создание оптического компьютера, привели к определенным успехам. Так, в 1990 году в лабораториях американской фирмы «Белл» был создан макет цифрового оптического устройства (рис. 12.26).

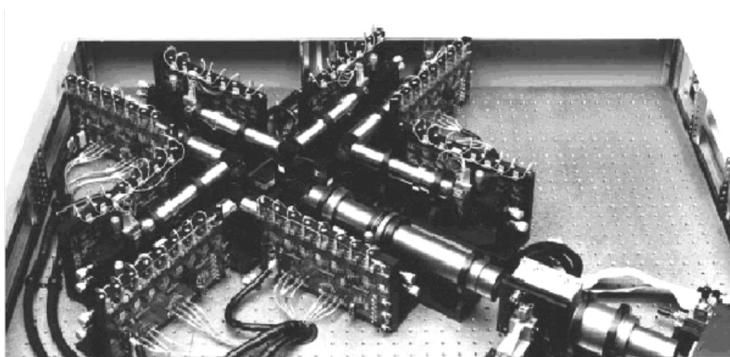


Рис. 12.26. Оптический компьютер DOC-II

С его помощью была продемонстрирована возможность выполнения цифровых и логических операций с высокими параметрами быстродействия и потребления энергии. Основу процессора разработанного оптического компьютера составляли двумерные матрицы бистабильных элементов (размерностью 4×8) на основе квантоворазмерных полупроводниковых структур, обладающих нелинейными электрооптическими свойствами (*self-electro-optic-effect devices* – SEED).

Первоначально остановимся на основных параметрах оптических бистабильных элементов. Система называется биста-

бильной, если она имеет два устойчивых состояния (мультистабильная – более двух).

Бистабильные системы можно классифицировать по способу осуществления обратной связи и механизму нелинейного пропускания. В оптическом бистабильном элементе SEED обратная связь осуществляется за счет приложения электрического поля к квантоворазмерной структуре, причем само поле возникает при прохождении через структуру светового излучения. SEED, разработанный в 1986 году, представлял собой фотодиоды из GaAlAs и структуру сверхрешетки, состоящей из 100 чередующихся слоев GaAs и GaAlAs толщиной по 95 нм, с множественными квантовыми ямами. Нелинейное пропускание элемента SEED связано с уменьшением экситонного поглощения. При больших концентрациях экситонов и свободных носителей происходит просветление в области экситонного резонанса. Это связано, во-первых, с тем, что в присутствии большого числа электронов и дырок кулоновское взаимодействие между электроном и дыркой ослабляется (этот эффект называется *экранированием*). Во-вторых, при большой концентрации экситонов они начинают интенсивно взаимодействовать, разрушая друг друга.

Важным достоинством первого оптического компьютера явилась возможность последовательного объединения его отдельных каскадов благодаря искусственному аналогу эффекта внутреннего усиления. Параметры системы были следующие: разрядность – 32 бита (массив 4×8); логика – бинарная; тактовая частота – 1,1 МГц (определялась быстродействием жидкокристаллической маски); число переключений в секунду – 40 Мб/с. Одним из достижений данного процессора была величина энергии на одно переключение, которая составляла 20 фДж и была на 6 порядков меньше величины энергии переключения в электронных компьютерах того времени.

Второе поколение оптических цифровых компьютеров представлено компьютером DOC-II (digital optical computer), разработанного в научно-исследовательской фирме США Opticom Corporation. В DOC-II использован принцип векторно-матрич-

ного умножения, однако вектор и матрица являются булевскими логическими.

В данном устройстве входной поток данных образовывался излучением линейки 64 независимо модулируемых полупроводниковых лазеров. Свет от каждого лазера линейки отображался на одну строчку матричного пространственного модулятора света с размером 64×128 элементов. Отдельный элемент матрицы представлял собой акустооптическую брэгговскую ячейку на основе полупроводника GaP. Свет, выходящий из рядов пространственного модулятора, попадал на линейку из 128 лавинных фотодиодов. DOC-II имел $64 \times 128 = 8192$ межсоединений и работал на частоте передачи данных $100 \text{ Мб} \cdot \text{с}^{-1}$, что соответствует $0,8192 \times 10^{12}$ переключений в секунду. Энергия на одно переключение составляет 7,15 фДж (~ 30000 фотонов). Для иллюстрации быстрodeйствия представим, что нужно найти какое-то слово в тексте. Типичный современный персональный компьютер *Duron 1,6/256 MB/Win XP SP1* на поиск слова в документе *Win Word*, состоящем из 953 страниц текста, тратит чуть больше трех секунд, в то время как оптический компьютер DOC-II просматривает за одну секунду 80 000 страниц обычного ASCII-текста.

Принципиальным недостатком макетов первых оптических компьютеров являлась неинтегрируемость их отдельных компонентов. Исходя из этого основной задачей следующего этапа работ по оптическому компьютеру было создание его интегрального варианта.

В конце 90-х годов прошлого века велись работы по созданию интегрального модуля оптического компьютера с логической матрично-тензорной основой, названного НРОС (*High Performance Optoelectronic Communication*). В устройстве планировалось использовать входную матрицу VCSEL лазеров, соединенную планарными волноводами и обычной оптикой с матрицами переключения, на основе дифракционных оптических элементов, и выходную систему, состоящую из матрицы лавинных фотодиодов, совмещенной с матрицей вертикально-излу-

чающих диодов. Опытные образцы показали производительность $4,096 \text{ Тб}\cdot\text{с}^{-1}$, а оценки свидетельствуют, что данная система способна развить скорость 10^{15} операций в секунду с энергией менее 1 фДж на одно переключение. В настоящее время фирма Opticom Corporation разработала новый интегральный оптический элемент, состоящий из матрицы лазеров и фотодетекторов, соединенных волноводом, и планирует использовать данные устройства как для обработки информации, так и для создания сверхбыстрых переключателей в сверхплотных волоконных линиях связи.

Применение фотоники в разработке оптических компьютеров стремительно расширяется и, подобно тому, как это произошло с электроникой, с ее прежними компьютерами, в будущем мы, вероятно, будем иметь дело не с электронами, а с фотонами. Соответствующие проектно-конструкторские работы ныне успешно продвигаются. Оптические компьютеры смогут обрабатывать более значительные объемы информации, чем электронные. Впрочем, оптические компьютеры вряд ли смогут в ближайшее время полностью вытеснить электронные, более того, предполагается дополнять электронные устройства соответствующими оптическими блоками с целью повышения их мощности.

Вопросы для самоконтроля

1. История информатики до компьютеров. Основные вехи создания компьютерной техники. Поколения компьютерной техники.

2. Как вы можете объяснить бытовой термин «переизбыток информации»? Что имеется в виду: излишняя полнота данных; излишняя сложность методов; неадекватность поступающих данных и методов, имеющихся в наличии?

3. Как вы понимаете термин «средство массовой информации»? Что это? Средство массовой поставки данных? Средство, обеспечивающее массовое распространение методов? Средство, обеспечивающее процесс информирования путем поставки

данных гражданам, обладающим адекватными методами их потребления?

4. Как вы понимаете следующие термины: аппаратно-программный интерфейс, программный интерфейс, аппаратный интерфейс? Как бы вы назвали специальность людей, разрабатывающих аппаратные интерфейсы? Как называется специальность людей, разрабатывающих программные интерфейсы?

5. Как вы понимаете динамический характер информации? Что происходит с ней по окончании информационного процесса?

6. Можем ли мы утверждать, что данные, полученные в результате информационного процесса, адекватны исходным? Почему? От каких свойств исходных данных и методов зависит адекватность результирующих данных?

7. Что характеризуют свойства информации: объективность, полнота, достоверность, адекватность, доступность и актуальность?

8. Перечислите этапы возможных операций с данными.

9. В каких единицах измеряется информация? Энтропия как количественная мера информации. Свойства энтропии.

10. При каких условиях энтропия источника дискретных сообщений с фиксированным числом состояний принимает максимальное значение?

11. Что понимается под архитектурой компьютера? Какие типы архитектур вам известны?

12. Концептуальная схема компьютера, основные элементы.

13. Принципиальные схемы компьютеров нетрадиционной архитектуры, иерархия архитектур.

14. В чем состоят термодинамический и квантовый пределы электронной ячейки? Прогноз Г. Мура.

15. В чем состоят преимущества оптических систем для задач передачи, хранения и обработки информации?

16. Типы оптических процессоров. Аналоговый оптический процессор. Аналоговые операции.

17. Что такое модуляция света? Какие виды модуляции вам известны?

18. Способы записи и хранения информации, принцип записи на компакт-диск.

19. В чем состоит преимущество записи информации при помощи средств голографии?

20. Параметры первого оптического цифрового процессора и их анализ.

13. ОСНОВЫ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ НЕРАВНОВЕСНЫХ ТЕПЛОФИЗИЧЕСКИХ ПРОЦЕССОВ В ФОТОНИКЕ

13.1. Роль тепло- и массообмена в фотонике

При генерировании, усилении, модуляции и переносе оптических сигналов происходит преобразование энергии и ее диссипация.

Так, лазерные технологии обработки материалов характеризуются высокой напряженностью электрического поля с частотой 10^{15} Гц, амплитудой $\sim 10^5$ – 10^6 В/см и высокой плотностью мощности, достигающей 10^{19} Вт/см² при скоростях нагрева более 10^{15} град/с. Лазерные воздействия применяются в следующих технологиях:

1. Испарение (и абляция в расплавленной фазе) – наиболее популярный процесс, который лежит в основе большинства промышленных технологий любых материалов в микроэлектронике, микромеханике и микрооптике;

2. Нагревание до температуры размягчения (или плавления) с последующим деформированием в вязкотекучей фазе (дополненное тем или иным механическим воздействием – вытяжкой, вращением и т.п.) применяется для изготовления ближнепольных оптических зондов, медицинского оптического инструментария и т.п.

3. Направленное локальное нагревание, вызывающее появление контролируемого поля напряжений, приводящего

к управляемому деформированию листовых материалов (*laser forming*), причем не только с целью формообразования, но и прецизионной сборки и юстировки микромеханических компонентов.

4. Послойный синтез трехмерных объектов методом послойного наращивания, в том числе стереолитография, селективное лазерное спекание и послойная сборка из листовых материалов (*laminated object manufacturing*).

На поверхности твердых тел при этом возникают различные физические процессы (табл. 13.1).

Таблица 13.1

Физические процессы, возникающие на поверхности
твердых тел при лазерном воздействии

Наименование процессов	Содержание процессов
Эмиссионные процессы	Десорбция газа; термоэлектронная и термоионная эмиссия; эмиссия нейтральных атомов; тепловое излучение
Структурные процессы	Рекристаллизация; структурные изменения в Fe-C сплавах (закалка сталей); размягчение и структурные изменения в стекле; аморфизация и кристаллизация стеклокерамик; аморфизация тонких металлических пленок; взаимная диффузия нагретых слоев (микрометаллургия); отжиг дефектов (в полупроводниках)
Поверхностные химические реакции	Локальное окисление металлов и полупроводников; восстановление окислов; термическое разложение металлоорганических соединений; полимеризация (деструкция) полимеров
Теплофизические эффекты	Тепловое расширение (включая импульсное); появление термонапряжений; генерация ударных волн в твердом теле и в воздухе; генерация ультразвука (дефектоскопия); оптический пробой в прозрачных диэлектриках; фазовые переходы (плавление, испарение); воспламенение и горение; детонация активных и взрыв пассивных сред

Процессы диссипации приводят к тепловыделениям и сопровождаются неоднородным распределением температуры.

К локализации тепловыделений приводит и стремление к уменьшению характерного размера микросхем. Энергия активных и пассивных потерь полностью диссипируется в электронных логических ячейках и порождает проблемы теплоотвода.

13.2. Виды теплообмена.

Законы молекулярного тепло- и массообмена

Теплообменом или теплопередачей называют учение о самопроизвольных необратимых процессах распространения теплоты в пространстве.

Различают три способа (механизма) распространения теплоты в пространстве:

- *теплопроводность* (кондукция, диффузия);
- *конвекция*;
- *тепловое излучение* (радиация).

Явление теплопроводности состоит в переносе теплоты микроскопическими структурными частицами вещества (молекулами, атомами, электронами) в процессе их теплового движения в телах с неоднородным распределением температуры.

Механизмы теплопроводности:

- *в газах* – вследствие обмена энергией при упругом соударении молекул, имеющих различную скорость теплового движения;
- *в жидкостях, полупроводниках и твердых диэлектриках* – путем непосредственной передачи теплового движения молекул и атомов соседним частицам вещества, передачей упругих волн колебаний кристаллической решетки;
- *в металлах* – главным образом движением свободных электронов, а также передачей колебаний кристаллической решетки.

Таким образом, в основе теплопроводности лежит *молекулярный (кондукционный, диффузионный)* механизм переноса тепла.

Явление конвекции наблюдается в текучих теплоносителях (жидкостях, газах) и состоит в переносе теплоты крупными (макроскопическими) частицами вещества (молями). Конвекция сопровождается теплопроводностью.

Виды конвекции:

– *свободная* – в неоднородном поле массовых или поверхностных сил, например, тепловая, термокапиллярная конвекция;

– *вынужденная* – под действием внешних вынуждающих сил (перепада давлений, электромагнитных сил, вибраций и пр.).

Теплообмен излучением (радиация) представляет собой перенос теплоты посредством электромагнитного поля. При этом происходит двойное превращение энергии: внутренняя энергия излучающего тела переходит в энергию электромагнитного поля, которая вновь переходит во внутреннюю энергию поглощающего тела. Особенность теплообмена излучением: телом могут обмениваться тела, разделенные вакуумом.

В природе все три вида переноса теплоты осуществляются одновременно. Их совокупность называют *сложным теплообменом*.

Виды сложного теплообмена:

– *радиационно-кондуктивный* теплообмен представляет совокупность теплообмена излучением и теплопроводностью;

– *радиационно-конвективный* теплообмен включает все три способа переноса теплоты;

– *теплоотдача* – это теплообмен между текучей средой и твердой поверхностью тела;

– *теплопередача* – это теплообмен между двумя жидкими или газообразными средами, разделенными твердой стенкой.

Процесс переноса тепла в сложном теплообмене всегда сопровождается переносом массы, поэтому говорят о явлениях *тепломассообмена*.

Процесс теплопроводности подчиняется *закону Фурье*, в соответствии с которым *плотность теплового потока* (количество теплоты dQ , проходящей через площадку dS за время dt) пропорциональна температурному градиенту

$$\vec{q}_T = -k \frac{\partial T}{\partial n} \vec{n} = -k \nabla T = -k \text{ grad } T. \quad (13.1)$$

Коэффициент пропорциональности k называется *коэффициентом теплопроводности*.

$$k = \frac{|q_T|}{|\partial T / \partial n|}, \quad \left[\frac{\text{Вт м}}{\text{м}^2 \text{ К}} \Rightarrow \frac{\text{Вт}}{\text{м} \cdot \text{К}} \right]. \quad (13.2)$$

Коэффициент теплопроводности является важнейшим теплофизическим свойством веществ и характеризует плотность теплового потока при единичном температурном градиенте. Знак «минус» отражает противоположность направлений векторов плотности теплового потока и температурного градиента, т.е. плотность теплового потока возрастает в соответствии со вторым законом термодинамики в направлении уменьшения температуры.

При конвективном теплообмене перенос теплоты неразрывно связан с переносом массы. Если текучая среда плотностью ρ [кг/м³] движется в направлении оси x со скоростью u_x [м/с], то ее массовая скорость характеризует массу среды, проходящей в единицу времени через единичную площадку:

$$\rho u_x \left[\frac{\text{кг м}}{\text{м}^3 \text{ с}} \Rightarrow \frac{\text{кг}}{\text{м}^2 \text{ с}} \right]. \quad (13.3)$$

Теплосодержание среды может быть выражено через массовую теплоемкость c [Дж/(кг · К)] и температуру T :

$$cT \left[\frac{\text{Дж} \cdot \text{К}}{\text{кг} \cdot \text{К}} \Rightarrow \frac{\text{Дж}}{\text{кг}} \right]. \quad (13.4)$$

Плотность теплового потока, определяемая конвекцией, равна произведению массовой скорости на теплосодержание,

$$q_k = \rho u_x c T \left[\frac{\text{кг}}{\text{м}^2 \text{с}} \frac{\text{Дж}}{\text{кг}} \Rightarrow \frac{\text{Вт}}{\text{м}^2} \right]. \quad (13.5)$$

Конвекция всегда сопровождается теплопроводностью, поэтому общая плотность теплового потока при конвективном теплообмене

$$\vec{q} = \vec{q}_t + \vec{q}_k = -k\nabla T + \rho \vec{u} c T. \quad (13.6)$$

Нагретая поверхность является источником теплового излучения, плотность теплового потока которого определяется законом Стефана–Больцмана, в соответствии с которым *плотность потока поверхностного излучения тела пропорциональна его абсолютной температуре в четвертой степени.*

$$q_n = \varepsilon \sigma T^4, \quad (13.7)$$

где $\sigma = 5,67 \cdot 10^{-8} \text{ Вт}/(\text{м}^2 \cdot \text{К}^4)$ – постоянная Стефана–Больцмана; ε – степень черноты излучающей поверхности ($0 < \varepsilon < 1$).

13.3. Перенос тепла теплопроводностью

В закрытой физической системе, в которой не происходит фазовых переходов, отношение между температурой и количеством тепла выражается соотношением

$$dQ = mcdT, \quad (13.8)$$

где m – масса; c – удельная массовая теплоемкость.

Однако в случае реальных взаимодействий лазерного излучения с веществом необходимо учитывать потери тепла, которые происходят вследствие *теплопроводности, конвективного теплообмена и теплового излучения.*

Температура является основной физической величиной, характеризующей все тепловые взаимодействия света с материалом. Основной задачей теории теплопроводности является опре-

деление и изучение пространственно-временного изменения температуры, $T = f(x, y, z, t)$; x, y, z – пространственные прямоугольные координаты, t – время.

Совокупность значений температур для всех точек пространства в данный момент времени t называется *температурным полем*. Если температура является функцией только от координат, то поле является *стационарным*. Если же температура также зависит от времени, поле будет *нестационарным*.

Дифференциальное уравнение теплопроводности связывает пространственное распределение температуры T с изменением ее во времени t и записывается следующим образом:

$$\frac{\partial T}{\partial t} = a \nabla^2 T + \frac{q_v}{\rho c}, \quad (13.9)$$

где ρ – плотность, кг/м^3 ; c – удельная массовая теплоемкость, $\text{Дж}/(\text{кг}\cdot\text{К})$; $a = \lambda/(\rho c)$ – коэффициент температуропроводности, $\text{м}^2/\text{с}$; λ – коэффициент теплопроводности, $\text{Вт}/(\text{м}\cdot\text{К})$; q_v – объемная плотность источников тепла, $\text{Вт}/\text{м}^3$, $\nabla^2 T = \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}$ –

оператор Лапласа.

Уравнение является частным случаем первого закона термодинамики и показывает изменение энергии вещества в элементарном объеме. Это изменение определяется количеством теплоты, накопленной за счет теплопроводности, и количеством теплоты, выделившейся в элементарном объеме за счет внутренних источников тепла.

В частном случае одномерного нестационарного температурного поля и отсутствия объемных источников тепла уравнение (13.9) принимает вид

$$\frac{\partial T}{\partial t} = a \frac{\partial^2 T}{\partial x^2}. \quad (13.10)$$

Дифференциальное уравнение теплопроводности имеет бесконечное множество решений. Чтобы найти единственное решение, характеризующее конкретный процесс, необходимо задать краевые условия.

Краевые условия включают в себя начальное (временное) и граничные (пространственные) условия.

Начальное краевое условие необходимо для нестационарного процесса и характеризует распределение температуры в начальный момент времени: $T(x, y, z, 0) = f(x, y, z)$, часто его принимают однородным:

$$T(t = 0) = T_0. \quad (13.11)$$

Граничные краевые условия характеризуют форму тела и условия его теплообмена с окружающей средой. Различают *четыре вида* граничных краевых условий.

При *граничных условиях 1-го рода* на поверхности тела для каждого момента времени задается распределение температуры $T_n = f(x_n, y_n, z_n, t)$, в частном случае температура поверхности может поддерживаться постоянной во времени, такая граница называется *изотермической*:

$$T_n = \text{const}. \quad (13.12)$$

При *граничных условиях 2-го рода* на поверхности тела для каждого момента времени задается плотность теплового потока $q_n = f(x_n, y_n, z_n, t)$. В частном случае плотность теплового потока может поддерживаться постоянной во времени, например при нагревании металла в высокотемпературных печах:

$$q_n = \text{const}. \quad (13.13)$$

Частным случаем граничного условия 2-го рода является *адиабатная граница*, теплообмен на которой отсутствует ($q_n = 0$), например ось симметрии тела.

При *граничных условиях 3-го рода* на поверхности тела для каждого момента времени задается температура окружающей среды и закон конвективного теплообмена между поверхностью тела и окружающей средой:

$$q_n = \alpha (T_n - T_c), \quad (13.14)$$

где T_n , T_c – температуры соответственно поверхности тела и окружающей среды; α – коэффициент теплоотдачи, Вт/(м²·К), характеризующий плотность теплового потока при единичной разности температур между поверхностью тела и окружающей средой. В частном случае при излучении нагретой поверхности в открытое пространство по закону Стефана–Больцмана коэффициент теплоотдачи имеет вид

$$\alpha = \sigma \varepsilon (T_n^2 + T_c^2)(T_n + T_c). \quad (13.15)$$

Граничные условия 4-го рода это условия теплообмена на границе контакта двух тел. В частном случае идеального контакта на границе эти условия отражают равенство плотностей тепловых потоков в направлении нормали к границе:

$$-k_1 \frac{\partial T_1}{\partial n} = -k_2 \frac{\partial T_2}{\partial n}. \quad (13.16)$$

Дифференциальное уравнение теплопроводности вместе с краевыми условиями образуют *краевую задачу теплопроводности*, имеющую единственное решение.

В качестве примера рассмотрим одномерную стационарную задачу теплопроводности плоского слоя толщиной δ , не ограниченного в направлении осей y , z и не содержащего внутренних источников тепла ($q_V = 0$). Его поверхности $x = 0$ и $x = \delta$ поддерживаются изотермическими: $T(x = 0) = T_1$ и $T(x = \delta) = T_2$, т.е. заданы граничные условия первого рода. Температурное поле в этом случае зависит только от одной координаты, и математическая формулировка краевой задачи теплопроводности имеет вид

$$\frac{d^2T}{dx^2} = 0, \quad T(x=0) = T_1, \quad T(x=\delta) = T_2. \quad (13.17)$$

Общее решение уравнения теплопроводности получается после двойного интегрирования:

$$\frac{dT}{dx} = C_1 \Rightarrow dT = C_1 dx \Rightarrow \int dT = \int C_1 dx \Rightarrow T = C_1 x + C_2.$$

Постоянные интегрирования C_1 и C_2 находятся подстановкой граничных условий в общее решение: $T_1 = C_1 \cdot 0 + C_2$; $T_2 = C_1 \cdot \delta + C_2$ и имеют вид

$$C_1 = -\frac{T_1 - T_2}{\delta}; \quad C_2 = T_1.$$

В результате получается решение задачи

$$T = T_1 - \frac{T_1 - T_2}{\delta} x, \quad (13.18)$$

дающее линейное распределение температуры по толщине слоя.

Плотность теплового потока определяется в соответствии с законом Фурье

$$q = -k \frac{dT}{dx} = k \frac{T_1 - T_2}{\delta} = \frac{T_1 - T_2}{\delta/k} \quad (13.19)$$

и является постоянной. Отношение δ/k называется *тепловым сопротивлением* плоского слоя.

13.4. Основы вычислительного эксперимента в теплофизике

Аналитические методы оказываются практически непригодными для нахождения двух- и трехмерных температурных полей в областях сложной конфигурации. От этих недостатков свободны численные методы, в которых дифференциальные операторы заменяются алгебраическими, получающиеся матричные

уравнения решаются на компьютерах с нахождением температурного поля в узловых точках конечно-разностной сетки.

Основную идею численных методов рассмотрим на примере одномерной нестационарной задачи теплопроводности. Она состоит в замене непрерывных производных по времени и координатам, входящих в уравнения теплопроводности и в краевые условия, их приближенными значениями в отдельных точках (узлах) конечно-разностной сетки.

В общем случае расположение узлов сетки в исследуемой области может быть произвольным. На практике часто применяют сетку, равномерно покрывающую расчетную область. Такая сетка с постоянными расстояниями между ближайшими узлами (шагами сетки) называется *регулярной*. Фрагмент такой сетки показан на рис. 13.1, а ее узлы определяются координатами

$$\begin{aligned} x_i &= (i-1)h_x; \quad i=1, 2, 3, \dots, N+1; \quad h_x = H_x/N, \\ t_k &= (k-1)h_t; \quad k=1, 2, 3, \dots; \quad h_t, \end{aligned} \quad (13.20)$$

где N – число разбиений по толщине слоя H_x ; h_x, h_t – соответственно шаги пространственной (по x) и временной (по t) сеток; i, k – номера узловых точек в направлении координат x, t .

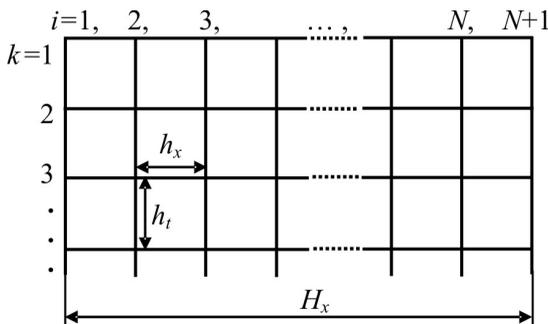


Рис. 13.1. Фрагмент регулярной сетки

С точностью до ошибок аппроксимации входящая в уравнение теплопроводности (13.10) первая производная от темпе-

ратуры по времени может быть найдена в i -й точке сетки, а вторая производная от температуры по координате на k -м слое по времени по конечно-разностным формулам

$$\frac{\partial T}{\partial t} \approx \frac{T_k - T_{k-1}}{h_t}, \quad \frac{\partial^2 T}{\partial x^2} \approx \frac{T_{i-1} - 2T_i + T_{i+1}}{h_x^2}. \quad (13.21)$$

Аппроксимацию уравнения (13.10) можно представить схематически, рассмотрев фрагмент сетки (шаблон) с минимальным количеством узловых точек (рис. 13.2). Существующие схемы аппроксимации делятся на *явные*, когда все производные по координате в уравнении переноса записываются на «старом» $(k-1)$ -м временном слое с известным распределением температуры, и *неявные*, когда все производные по координате в этом уравнении записываются на «новом» k -м временном слое с неизвестным распределением температуры.

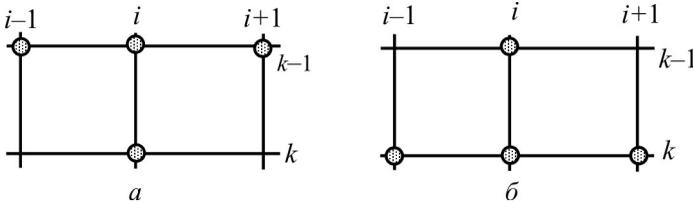


Рис. 13.2. Сеточные шаблоны явной (а) и неявной (б) схем аппроксимации уравнения теплопроводности

Явная схема аппроксимации уравнения (13.10) дает соотношение

$$\frac{T_{i,k} - T_{i,k-1}}{h_t} = a \frac{T_{i+1,k-1} - 2T_{i,k-1} + T_{i-1,k-1}}{h_x^2}, \quad (13.22)$$

из которого получается явная формула для неизвестной температуры

$$T_{i,k} = T_{i,k-1} \left(1 - \frac{2ah_t}{h_x^2} \right) + \frac{ah_t}{h_x^2} (T_{i+1,k-1} + T_{i-1,k-1}). \quad (13.23)$$

где $[H]$ – матрица коэффициентов; $\{T\}$ – вектор-столбец неизвестных температур в узловых точках; $\{F\}$ – неизвестный вектор-столбец, характеризующий краевые условия и распределение температуры на предыдущем временном слое. Видно, что матрица $[H]$ обладает рядом специальных свойств, которые необходимо использовать при решении системы. Она имеет высокий порядок, зависящий от густоты сетки, является редко заполненной с размещением ненулевых элементов по диагонали в три ряда. Такие матрицы называются ленточными трехдиагональными. Важным свойством является симметрия матрицы относительно ее диагонали.

Рассмотрим решение системы уравнений (13.25) методом прогонки, являющимся модификацией метода исключения Гаусса и учитывающим свойства матрицы H .

Решение системы в узловой точке ищется в виде линейной функции. В частности, для $(i-1)$ -й точки эта функция имеет вид

$$T_{i-1} = \beta_i T_i + z_i, \quad (13.28)$$

где β_i , z_i – неизвестные пока вспомогательные коэффициенты. Подставим (13.28) в (13.25):

$$A(\beta_i T_i + z_i) + B T_i + C T_{i+1} = F_i, \quad (13.29)$$

откуда находим

$$T_i = -\frac{C}{A\beta_i + B} T_{i+1} - \frac{Az_i - F_i}{A\beta_i + B}. \quad (13.30)$$

Полученное соотношение имеет ту же форму, что и функция (13.28), только для i -й точки

$$T_i = \beta_{i+1} T_{i+1} + z_{i+1}, \quad (13.31)$$

откуда заключаем, что

$$\beta_{i+1} = -\frac{C}{A\beta_i + B}; \quad z_{i+1} = -\frac{Az_i - F_i}{A\beta_i + B}. \quad (13.32)$$

Полученные коэффициенты называются прогоночными коэффициентами, а формулы (13.31–13.32) дают процедуру решения.

Сначала при $i = 2, 3, \dots, N$ считаются прогоночные коэффициенты (13.32), при этом начальные значения прогоночных коэффициентов β_2, z_2 определяются из граничных условий на левой границе. Эта операция называется прямой прогонкой. После определения всех β_i, z_i в обратном направлении ($i = N, N-1, \dots, 2$) с учетом значения температуры T_{N+1} , найденного из граничного условия на правой границе, по формуле (13.31) последовательно находятся неизвестные значения T_i в узловых точках сетки.

Рассмотрим реализацию метода прогонки для задачи о стационарном распределении температуры в плоском слое с известным решением (13.18). В качестве теста для проверки алгоритма рассмотрим пример при граничных условиях первого рода: $T(x = 0) = T_n, T(x = \delta) = T_n$. Решение задачи методом сеток дает систему уравнений с граничными условиями

$$\left. \begin{aligned} T_{i-1} - 2T_i + T_{i+1} &= 0, \\ i &= 2, 3, \dots, N, \\ T_1 &= T_n, \quad T_{N+1} = T_n. \end{aligned} \right\} \quad (13.33)$$

Алгоритм решения этой системы имеет следующий вид:

$$\left. \begin{aligned} \beta_2 &= 0, \quad z_2 = T_n, \\ \beta_{i+1} &= -\frac{C}{A\beta_i + B}, \quad z_{i+1} = -\frac{Az_i - F_i}{A\beta_i + B}, \\ i &= 2, 3, \dots, N, \quad T_{N+1} = T_n, \\ T_i &= \beta_{i+1}T_{i+1} + z_{i+1}, \quad i = N, N-1, \dots, 1. \end{aligned} \right\} \quad (13.34)$$

В частности, для числа разбиений $N = 4$ при граничных условиях $T_{\text{л}} = 100$, $T_{\text{п}} = 200$ запишем эту систему в векторно-матричной форме:

$$\begin{bmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix} \cdot \begin{Bmatrix} T_2 \\ T_3 \\ T_4 \end{Bmatrix} = \begin{Bmatrix} -100 \\ 0 \\ -200 \end{Bmatrix},$$

алгоритм прогонки (13.34) реализуется для этой системы при $A = C = 1$, $B = -2$ следующим образом:

$$\beta_2 = 0; \quad z_2 = T_{\text{л}} = 100;$$

$$\beta_3 = -\frac{c}{a\beta_2 + b} = -\frac{1}{1 \cdot 0 - 2} = \frac{1}{2}; \quad z_3 = -\frac{az_2 - f_2}{a\beta_2 + b} = -\frac{1 \cdot 100 - 0}{1 \cdot 0 - 2} = 50;$$

$$\beta_4 = -\frac{c}{a\beta_3 + b} = -\frac{1}{1 \cdot 1/2 - 2} = \frac{2}{3}; \quad z_4 = -\frac{az_3 - f_3}{a\beta_3 + b} = -\frac{1 \cdot 50 - 0}{-3/2} = \frac{100}{3};$$

$$\beta_5 = -\frac{c}{a\beta_4 + b} = -\frac{1}{1 \cdot 2/3 - 2} = \frac{3}{4};$$

$$z_5 = -\frac{az_4 - f_4}{a\beta_4 + b} = -\frac{1 \cdot 100/3 - 0}{1 \cdot 2/3 - 2} = 25;$$

$$T_5 = T_{\text{п}} = 200; \quad T_4 = \beta_5 T_5 + z_5 = \frac{3}{4} \cdot 200 + 25 = 175;$$

$$T_3 = \beta_4 T_4 + z_4 = \frac{2}{3} \cdot 175 + \frac{100}{3} = 150;$$

$$T_2 = \beta_3 T_3 + z_3 = \frac{1}{2} \cdot 150 + 50 = 125; \quad T_1 = T_{\text{л}} = 100.$$

Таким образом, численным решением получили искомое линейное распределение температуры, что подтверждает правильность работы алгоритма прогонки.

Мы рассмотрели наиболее простые схемы аппроксимации уравнения теплопроводности. Существуют и другие более сложные схемы, позволяющие уменьшить *ошибки аппроксимации*, вызванные заменой производных в уравнении теплопроводности приближенными значениями. Ошибки аппроксимации можно оценить, находя решение на последовательности сгущающихся сеток.

При выполнении арифметических операций на компьютере числа представляются в экспоненциальной форме с ограниченным числом разрядов и возникают *ошибки округления*. Ошибки округления можно уменьшить, изменяя метод решения матричных уравнений, последовательность арифметических операций и увеличивая число разрядов для записи чисел в компьютере (например, применяя двойную точность).

Ошибки аппроксимации, округления и другие образуют спектр, оценка которого для реальных задач является далеко не простой. *Проблема аппроксимации* – одна из основных в вычислительном эксперименте.

В процессе решения на компьютере спектр ошибок проявляется в виде возмущений. Кроме того, возмущения вносятся крайними условиями. Суммарные возмущения в процессе вычислительного эксперимента могут затухать или возрастать. В первом случае говорят об устойчивом численном алгоритме. Во втором случае появляются осцилляции нарастающей амплитуды, суммарные возмущения увеличиваются до больших значений, и численное решение теряет всякий смысл. Возникает *проблема устойчивости* численного алгоритма.

И, наконец, существует *проблема эффективности*, связанная с разработкой таких алгоритмов и программ, которые обеспечивают решение задачи с минимальными ошибками аппроксимации за наименьшее время.

Вопросы для самоконтроля

1. Запишите первую и вторую производные на регулярной конечно-разностной сетке.
2. Явная и неявная схемы аппроксимации уравнения теплопроводности, их преимущества и недостатки.
3. Как оценить погрешность в вычислительном эксперименте?
4. Метод прогонки решения матричных уравнений и его реализация на компьютере.
5. В чем состоят проблемы вычислительного эксперимента: аппроксимации, устойчивости и эффективности?

ЗАКЛЮЧЕНИЕ

Фотоника и оптоинформатика одно из наиболее актуальных для развития цивилизации научно-технических направлений. Уже в ближайшем будущем оно существенно изменит важнейшие характеристики измерительной аппаратуры, значительно ускорит создание эффективных систем управления глобальными экономическими, социальными и экологическими процессами и улучшит качество жизни человека.

Фотон начинает конкурировать с электроном как носитель информации и энергии высокой плотности. Если электроника успешно осваивает «верхний» диапазон наноразмеров (10...100 нм), то фотоника дает принципиально иные подходы для освоения «нижнего» диапазона (1...10 нм). В результате на полупроводниковых материалах удастся создавать наноэлементы для обработки и хранения информации, в которых используются особые квантовые структуры: нульмерные, одномерные и двумерные. Это позволяет сделать качественный скачок в разработке малагабаритных устройств хранения информации, в повышении на порядки быстродействия компьютеров, в разработке новых источников излучения с высокой плотностью энергии и т.д.

Если вспомнить, что около 20 % электрической энергии на Земле расходуется на освещение, значительное энергосбережение уже в ближайшие годы дадут лампы нового поколения – светоиспускающие диоды, которым при одинаковой светоотдаче требуется впятеро меньше энергии, чем лампам накаливания.

В самых различных областях современной оптической техники: в высокоточных оптических датчиках, биосенсорах, кремниевой фотонике, волоконно-оптических линиях связи и во многих других широко используются нелинейные оптические эффекты. В связи с разработкой мощных оптических излучателей и высококачественных оптических сред области практического применения

нелинейной оптики непрерывно расширяются, при этом величина пороговой мощности, при которой наблюдаются эффекты нелинейности, имеет тенденцию к снижению. Это позволяет средствами волоконной оптики передавать значительные плотности энергии на большие расстояния с малыми потерями.

Понимание и решение проблем фотоники и оптоинформатики немислимо без знания законов нелинейной оптики, без фундаментального физико-математического образования. Успешное решение проблем разработки оптической техники зависит от уровня и состояния технологии производства оптических материалов и оптико-электронных компонентов этой техники. Применение математического моделирования, вычислительного компьютерного эксперимента позволяет значительно ускорить поиск оптимальной технологии производства, повысить точность и качество проектируемых приборов.

Одним из достоинств специальности «Фотоника и оптоинформатика» является ее чрезвычайно высокая наукоемкость. Эта специальность основана на новейших достижениях физики, химии и биологии. Элементы нанофотонных и наноэлектронных систем, технология их изготовления, используемая контрольная и диагностическая аппаратура проектируются и функционируют непосредственно на основе фундаментальных законов природы, управляющих атомными процессами в материальных объектах. Обучение на данной специальности позволяет студенту освоить законы фундаментальных наук настолько глубоко, что любые новации в современной технике будут им легко осмысливаться и осознанно использоваться в последующей профессиональной деятельности. Это открывает выпускнику специальности «Фотоника и оптоинформатика» перспективы эффективного участия в научно-техническом прогрессе не только своей, но и смежных областей.

СПИСОК ЛИТЕРАТУРЫ

1. Айхлер Ю., Айхлер Г.И. Лазеры. Исполнение, управление, применение. – М.: Техносфера, 2008. – 440 с.
2. Андреева О.В. Прикладная голография: учеб. пособие / С. Петерб. ин-т точной механики и оптики. – СПб., 2008. – 184 с.
3. Арбузов В.И. Основы радиационного оптического материаловедения: учеб. пособие / С. Петерб. ин-т точной механики и оптики. – СПб., 2008. – 284 с.
4. Астапенко В.А. Резонансные процессы в электромагнитном поле: учеб. пособие. – М.: Изд-во Моск. физ.-техн. ин-та, 2007. – 171 с.
5. Бейли Дэвид, Райт Эдвин. Волоконная оптика: теория и практика: пер. с англ. – М.: Кудиц-пресс, 2008. – 320 с.
6. Беликов А.В., Скрипник А.В. Лазерные биомедицинские технологии: учеб. пособие / С.-Петербург. ин-т точной механики и оптики. – СПб., 2008. – Ч. 1. – 116 с.
7. Бертолотти М. История лазера. – Долгопрудный: Интеллект, 2011. – 336 с.
8. Беспалов В.Г. Основы оптоинформатики. Ч. 1. Информационные технологии – от электронного к оптическому компьютеру / С.-Петербург. ин-т точной механики и оптики. – СПб., 2006. – 52 с.
9. Богатырева В.В., Дмитриев А.Л. Оптические методы обработки информации: учеб. пособие / С.-Петербург. ин-т точной механики и оптики. – СПб., 2009. – 74 с.
10. Борисенко В.Е., Воробьева А.И., Уткина Е.А. Нанoeлектроника. – М.: БИНОМ. Лаборатория знаний, 2009. – 223 с.
11. Бочкарев С.В., Цаплин А.И. Диагностика и надежность автоматизированных систем: учеб. пособие. – Пермь: Изд-во Перм. гос. техн. ун-та, 2008. – 485 с.

12. Булгакова С.А., Дмитриев А.Л. Нелинейно-оптические устройства обработки информации: учеб. пособие / С.-Петерб. ин-т точной механики и оптики. – СПб., 2009. – 56 с.
13. Бусурин В.И., Носов Ю.Р. Волоконно-оптические датчики: Физические основы, вопросы расчета и применения. – М.: Энергоатомиздат, 1990. – 256 с.
14. Васильев В.Н., Павлов А.В. Оптические технологии искусственного интеллекта: учеб. пособие / С.-Петерб. ин-т точной механики и оптики. – СПб., 2005. – 99 с.
15. Вейко В.П., Петров А.А. Опорный конспект лекций по курсу «Лазерные технологии». Введение в лазерные технологии / С.-Петерб. ин-т точной механики и оптики. – СПб., 2009. – 143 с.
16. Оптические кабели связи: учеб. пособие для техникумов / Верник С.М., Гитин В.Я., Иванов В.С. [и др.]. – М.: Радио и связь, 1988. – 144 с.
17. Волоконно-оптические датчики / под ред Т. Окоси. – М.: Энергоатомиздат, 1990. – 256 с.
18. Волоконно-оптические датчики. Вводный курс для инженеров и научных работников; под ред. Э. Удда. – М.: Техносфера, 2008. – 520 с.
19. Ворзобова Н.Д., Денисюк И.Ю. Оптические методы формирования микроэлементов информационных систем: учеб. пособие / С.-Петерб. ин-т точной механики и оптики. – СПб., 2008. – 82 с.
20. Гусев А.И. Наноматериалы, наноструктуры, нанотехнологии. – 2-е изд., испр. – М.: Физматлит, 2009. – 416 с.
21. Дианов Е.М. Волоконная оптика: от систем связи к «нервным системам» // Вестн. Рос. акад. наук, 1007. – Т. 77, № 8. – С. 714–718.
22. Дианов Е.М., Плотниченко В.Г. Инфракрасные волоконные световоды. – М.: Знание, 1991. – 64 с.
23. Дмитриев А.Л. Оптические методы обработки информации: учеб. пособие / С.-Петерб. ин-т точной механики и оптики. – СПб., 2005. – 46 с.

24. Дмитриев А.Л. Оптические системы передачи информации: учеб. пособие / С.-Петербург. ин-т точной механики и оптики. – СПб., 2007. – 96 с.
25. Дмитриев В.Г., Тарасов Л.В. Прикладная нелинейная оптика. – 2-е изд. – М.: Физматлит, 2004. – 512 с.
26. Ермаков О.Н. Прикладная оптоэлектроника. – М.: Техносфера, 2004. – 416 с.
27. Ефимов А.М. Оптические свойства материалов и механизмы их формирования / С.-Петербург. ин-т точной механики и оптики. – СПб., 2008. – 103 с.
28. Апенко М.И., Запрыгаева Л.А., Свешникова И.С. Задачник по прикладной оптике: учеб. пособие. – 2-е изд. – М.: Высш. шк., 2003. – 591 с.
29. Заказнов Н.П., Кирюшин С.И., Кузичев В.И. Теория оптических систем: учеб. пособие. – 4-е изд. – СПб.: Лань, 2008. – 448 с.
30. Звелто О. Принципы лазеров / пер. под науч. ред. Т.А. Шмаонова. – 4-е изд. – СПб.: Лань, 2008. – 720 с.
31. Золотарев В.М. Методы исследования материалов фотоники: элементы теории и техники: учеб. пособие / С.-Петербург. ин-т точной механики и оптики. – СПб., 2008 – 275 с.
32. Иванов А.Б. Волоконная оптика: компоненты, системы передачи, измерения. – М.: Сайрус системс, 1999. – 671 с.
33. Игнатов А.Н. Оптоэлектроника и нанофотоника: учеб. пособие. – СПб.: Лань, 2011. – 544 с.
34. Информатика. Базовый курс. – 2-е изд. / под ред. С.В. Симоновича. – СПб.: Питер, 2007. – 640 с.
35. Королев Л.Н., Миков А.И. Информатика. Введение в компьютерные науки: учебник. – М.: Высш. шк., 2003. – 341 с.
36. Ионина Н.В., Орлов В.В., Павлов А.В. Оптические технологии искусственного интеллекта: эксперим. практикум: учеб. пособие / С.-Петербург. ин-т точной механики и оптики. – СПб., 2008. – 62 с.
37. Калитеевский Н.И. Волновая оптика: учеб. пособие. – 4-е изд. – СПб.: Лань, 2006. – 480 с.

38. Каманина Н. В. Электрооптические системы на основе жидких кристаллов и фуллеренов – перспективные материалы нанoeлектроники. Свойства и области применения: учеб. пособие / С.-Петербург. ин-т точной механики и оптики. – СПб., 2008. – 137 с.
39. Ларкин А.И., Юу Ф.Т.С. Когерентная фотоника. – М.: БИНОМ. Лаборатория знаний, 2007. – 319 с.
40. Корешев С.Н. Основы голографии и голограммной оптики / С.-Петербург. ин-т точной механики и оптики. – СПб., 2009. – 97 с.
41. Котоусов А.С. Теория информации: учеб. пособие для вузов. – М.: Радио и связь, 2003. – 80 с.
42. Либенсон М.Н., Яковлев Е.Б., Шандыбина Г.Д. Взаимодействие лазерного излучения с веществом (силовая оптика): конспект лекций. Ч. I. Поглощение лазерного излучения в веществе / под общ. ред. В.П. Вейко / С.-Петербург. ин-т точной механики и оптики. – СПб., 2008. – 141 с.
43. Листвин А.В., Листвин В.Н., Швырков Д.В. Оптические волокна для линий связи. – М.: ЛЕСАРт, 2003. – 288 с.
44. Листвин А.В., Листвин В.Н. Рефлектометрия оптических волокон. – М.: ЛЕСАРт, 2005. – 208 с.
45. Лозовский В.Н., Константинова Г.С., Лозовский С.В. Нанотехнология в электронике. Введение в специальность: учеб. пособие. – 2-е изд., испр. – СПб.: Лань, 2008. – 336 с.
46. Нанотехнологии. Азбука для всех / под ред. Ю.Д. Третьякова. – 2-е изд., испр. и доп. – М.: Физматлит, 2009. – 368 с.
47. Уильямс Л., Адамс У. Нанотехнологии без тайн; пер. с англ. Ю.Г. Гордиенко. – М.: Эксмо, 2009. – 368 с.
48. Балабанов В.И. Нанотехнологии. Наука будущего. – М.: Эксмо, 2009. – 256 с.
49. Борисенко В.Е., Воробьева А.И., Уткина Е.А. Нанoeлектроника. – М.: БИНОМ. Лаборатория знаний, 2009. – 223 с.
50. Мартинес-Дуарт Дж.М., Мартин-Палма Р.Дж., Агулло-Руеда Ф. Нанотехнологии для микро- и оптоэлектроники. – М.: Техносфера, 2007. – 368 с.

51. Верник С.М., Гитин В.Я., Иванов В.С. Оптические кабели связи: учеб. пособие для техникумов. – М.: Радио и связь, 1988. – 144 с.
52. Основы прикладной нанотехнологии / А.А. Абрамян [и др.]. – М.: МАГИСТР-ПРЕСС, 2007. – 208 с.
53. Пул Ч.-мл., Оуэне Ф. Нанотехнологии. – изд. 4-е, испр. и доп. – М.: Техносфера, 2009. – 336 с.
54. Пушкарева А.Е. Методы математического моделирования в оптике биоткани: учеб. пособие / С.-Петербург. ин-т точной механики и оптики. – СПб., 2008. – 103 с.
55. Розанов Н.Н. Нелинейная оптика. Ч. I. Уравнения распространения излучения и нелинейный отклик среды / С.-Петербург. ин-т точной механики и оптики. – СПб., 2008. – 95 с.
56. Розеншер Э., Винтер Б. Оптоэлектроника. – М.: Техносфера, 2004. – 592 с.
57. Серегин В.В. Прикладная теория и принципы построения гироскопических систем: учеб. пособие / С.-Петербург. ин-т точной механики и оптики. – СПб., 2007. – 78 с.
58. Сильман Г.И. Материаловедение: учеб. пособие для студ. высш. учеб. заведений. – М.: Академия, 2008. – 336 с.
59. Суздаев И.П. Нанотехнология: Физико-химия нанокластеров, наноструктур и наноматериалов. – изд. 2-е, испр. – М.: Либроком, 2009. – 592 с.
60. Теренин А.Н. Фотоника молекул красителей. – Л.: Наука, 1967. – 616 с.
61. Федоров А.В. Физика и технология гетероструктур, оптика квантовых наноструктур: учеб. пособие / С.-Петербург. ин-т точной механики и оптики. – СПб., 2009. – 195 с.
62. Ахманов С.А., Никитин С.Ю. Физическая оптика: учебник. – 2-е изд. – М.: Изд-во МГУ; Наука, 2004. – 656 с.
63. Цаплин А.И., Лихачев М.Е. Методы измерений в волоконной оптике: учеб. пособие / под общ. ред. д-ра техн. наук, проф. А.И. Цаплина. – Пермь: Изд-во Перм. нац. исслед. политехн. ун-та, 2011. – 277 с.

64. Цаплин А.И. Теплофизика в металлургии: учеб. пособие. – Пермь: Изд-во Перм. гос. техн. ун-та, 2008. – 230 с.
65. Эткинз П. Десять великих идей науки. Как устроен наш мир; пер. с англ. В. Герцика. – М.: АСТ: Астрель, 2008. – 384 с.
66. Янг М. Оптика и лазеры, включая волоконную оптику и оптические волноводы: пер с англ. – М.: Мир, 2005. – 541 с.

Учебное издание

Цаплин Алексей Иванович

ФОТОНИКА И ОПТОИНФОРМАТИКА

Введение в специальность

Учебное пособие

Редактор и корректор *И.А. Мангасарова*

Подписано в печать 03.10.2012. Формат 60×90/16.
Усл. печ. л. 25. Тираж 100 экз. Заказ № 37/2012.

Издательство
Пермского национального исследовательского
политехнического университета.
Адрес: 614990, г. Пермь, Комсомольский проспект, 29, к. 113.
Тел. (342) 219-80-33.