

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Физический факультет
Кафедра общей физики физического факультета

А.В. Иванов

**ЛЕКЦИИ ПО ЭЛЕКТРОДИНАМИКЕ, ОПТИКЕ
И КВАНТОВОЙ ФИЗИКЕ**

Курс лекций

Новосибирск
2020

Иванов А.В. Лекции по электродинамике, оптике и квантовой физике. Курс лекций / Новосибир. гос. ун-т. Новосибирск, 2018. – 116 с.

Курс лекций представляет собой взаимоувязанные лекции по трем разделам физики – электродинамике, оптике и квантовой физике. Курс состоит из трех частей, каждый из которых посвящен рассмотрению соответствующего раздела. Основное внимание в курсе сосредоточено на анализе волновых явлений и процессов, возникающих в этих областях физики. Этот курс так же знакомит студентов с процессами, происходящими в электрических цепях как с сосредоточенными, так и с распределенными параметрами, дает студентам ряд практически важных навыков для их расчета.

Курс лекций предназначен для студентов второго курса факультета информационных технологий НГУ. Он может быть полезен также студентам физического и геолого-геофизического факультетов НГУ, преподавателям физики.

Курс лекций подготовлен в рамках реализации
Программы развития НИУ-НГУ на 2009-2018 г.г.

@ Новосибирский государственный
университет, 2020

Оглавление

Часть I. Электродинамика	4
1.1 Методы расчета электрических цепей постоянного тока	4
1.2 Переходные процессы.....	8
1.3 Цепи синусоидального переменного тока	12
1.4 Резонанс.....	17
1.5 Фильтры	18
1.6 Широтно-импульсная модуляция.....	21
1.7 Применение закона электромагнитной индукции	23
1.8 Длинные линии.....	24
1.9 Квазистационарное приближение. Скин-эффект.....	28
1.10 Поток энергии электромагнитного поля.....	32
1.11 Электромагнитные волны	33
Часть II. Оптика.....	39
2.1 Отражение и преломление электромагнитных волн	39
2.2 Геометрическая оптика.....	44
2.3 Интерференция	48
2.4 Дифракция.....	58
2.5 Дисперсия света.....	68
2.6 Фазовая и групповая скорости	74
2.7 Оптическая спектроскопия.....	77
Часть III. Квантовая физика	85
3.1 Кризис классической физики.....	85
3.2 Равновесное тепловое излучение	86
3.3 Фотоэффект.....	88
3.4 Модель атома Бора.....	90
3.5 Волны де Бройля. Волновая функция	92
3.6 Уравнение Шредингера	93
3.7 Физический смысл волновой функции	95
3.8 Принцип неопределенности Гейзенберга	96
3.9 Операторы в квантовой механике	97
3.10 Некоторые частные решения уравнения Шредингера	99
3.11 Квантовый линейный осциллятор	104
3.12 Движение в поле центральной силы	106
3.13 Оператор момента импульса	110
3.14 Спин	111
3.15 Заполнение электронных оболочек атомов	114

Часть I.

Электродинамика

1.1 Методы расчета электрических цепей постоянного тока

Электрическая цепь – совокупность устройств и элементов, предназначенных для протекания электрического тока. Процессы в электрических цепях могут быть описаны с помощью таких понятий как **сила тока** и **напряжение**. В состав цепей могут входить **линейные** и **нелинейные** элементы. К линейным элементам относятся идеализированные источники токов и ЭДС, резисторы, конденсаторы, катушки индуктивности, т.е. такие элементы, сила тока и напряжение на которых описываются линейными уравнениями (алгебраическими или дифференциальными). Среди нелинейных элементов наиболее известны полупроводниковые элементы, такие как диоды, транзисторы и т.п.

Для расчета электрических цепей, т.е. для определения напряжений и токов в них, применяют различные методы, некоторые из которых мы сейчас рассмотрим.

Правила Кирхгофа

Наиболее известен метод расчета электрических цепей при помощи *правил Кирхгофа*. **Эти правила задают соотношения между токами и напряжениями**, которые должны выполняться на любом участке электрической цепи. Иногда эти правила называют законами Кирхгофа, однако это не совсем верно – правила Кирхгофа не являются фундаментальными законами природы, а следуют из других, более фундаментальных утверждений. Кроме того, эти правила не следует путать с другими законами Кирхгофа – законом излучения в оптике и законом (уравнением) в химии.

Сформулируем понятия узла, ветви и контура электрической цепи. **Ветвь** – любой двухполюсник, входящий в цепь. **Узел** – точка соединения трех и более ветвей. **Контур** – замкнутый цикл из ветвей, т.е. совокупность последовательно соединенных друг с другом ветвей, однократно проходя по которым можно вернуться в начальную точку.

Первое правило Кирхгофа гласит, что алгебраическая сумма токов в каждом узле равна нулю. Это правило – следствие фундаментального закона сохранения заряда: сколько заряда за единицу времени втекает в узел, столько должно и вытекать из него. Для того, чтобы воспользоваться этим правилом, предварительно необходимо выбрать направления токов в ветвях цепи. Направления можно выбрать произвольно, окончательное решение всё равно даст правильные знаки токов.

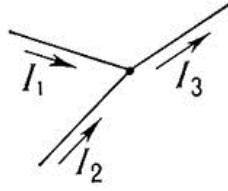


Рисунок 1. Втекающей и вытекающей из узла ток.

Первое правило Кирхгофа позволяет составить алгебраические уравнения относительно неизвестных значений токов в ветвях. В этих уравнениях ток, втекающий в узел, принято учитывать со знаком “плюс”, а вытекающий – со знаком “минус”. В частности, для ситуации, изображенной на рисунке 1, это правило приведет к уравнению

$$I_1 + I_2 - I_3 = 0.$$

Если в цепи N узлов, то независимыми будут $N - 1$ уравнение, N -е уравнение будет следствием из них.

Второе правило Кирхгофа гласит, что для каждого контура алгебраическая сумма падений напряжений на всех ветвях, принадлежащих этому контуру, равна алгебраической сумме ЭДС в этом контуре:

$$\sum_i E_i = \sum_j U_j.$$

Как и в случае первого правила Кирхгофа, не все из этих уравнений являются независимыми, однако для полного расчета электрической цепи необходимо, чтобы каждый из ее участков вошел в получающуюся систему уравнений хотя бы один раз.

При составлении уравнений по 2-му правилу Кирхгофа нужно выбрать направление обхода каждого из контуров. При этом падение напряжения на ветви считают положительным, если направление обхода данной ветви совпадает с ранее выбранным направлением тока ветви, и отрицательным — в противном случае. Аналогичное правило действует и для ЭДС ветвей. Направления обхода в каждом из контуров можно выбирать совершенно произвольно и независимо от выбора направлений в других контурах.

Таким образом, **правила Кирхгофа позволяют свести задачу расчета электрических цепей к задаче решения системы уравнений.** Эту систему необходимо дополнить связью между током и напряжением для каждого из элементов цепи. Так, для электрической цепи, состоящей только из резисторов, падение напряжения на каждой из ветвей цепи пропорционально текущему через нее току:

$$U_j = I_j R_j,$$

где R_j – величина сопротивления ветви. Подставляя эту связь во второе правило, получаем систему алгебраических уравнений, в которой неизвестными являются величины токов в ветвях цепи.

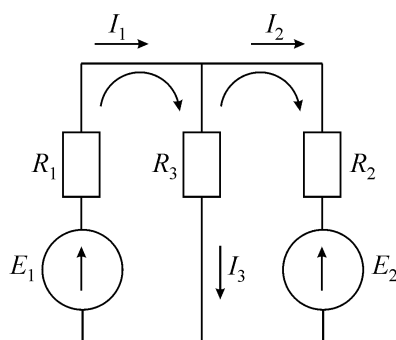


Рисунок 2. Пример схемы для расчета согласно правилам Кирхгофа.

Для примера применения правил Кирхгофа запишем систему уравнений для схемы, изображенной на рисунке 2. На этом рисунке показаны выбранные направления токов и обхода контуров. Согласно вышесказанному, получается следующая система из трех уравнений относительно трех неизвестных токов:

$$\begin{aligned} I_1 - I_2 - I_3 &= 0, \\ E_1 &= I_1 R_1 + I_3 R_3, \\ -E_2 &= I_2 R_2 - I_3 R_3. \end{aligned}$$

Метод контурных токов

Метод контурных токов – модификация метода расчета электрических цепей непосредственным применением правил Кирхгофа. В качестве неизвестных величин в этом методе используются так называемые **контурные токи**. Полагается, что в каждом независимом контуре цепи **течет свой контурный ток**. Если некоторая ветвь цепи принадлежит только одному контуру, ток в ней равен контурному. Если же ветвь принадлежит нескольким контурам, ток в ней равен сумме соответствующих контурных токов (с учётом направления обхода контуров). Рассматриваемые контура должны покрывать собой всю схему (т.е. любая ветвь должна принадлежать хотя бы одному контуру), ток в любой ветви можно выразить через контурные токи, и контурные токи составляют полную систему токов. Таким образом можно ограничиться только вторым правилом Кирхгофа, и уменьшить число уравнений в системе.

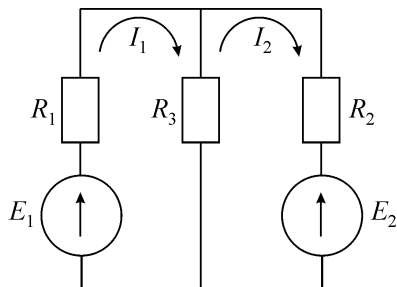


Рисунок 3. Пример схемы для расчета согласно методу контурных токов.

Продемонстрируем применение метода контурных токов на примере схемы, изображенной на рисунке 3. Эта схема полностью аналогична схеме, приведенной на рисунке 2, но в данном случае нам нет необходимости определять ток в каждой ветви, достаточно только определить ток (и его направление) в каждом

контуре. Знаки ЭДС ветвей определяются, как и во втором правиле Кирхгофа. Получается система из двух уравнений

$$\begin{aligned}E_1 &= I_1 R_1 + (I_1 - I_2) R_3, \\ -E_2 &= I_2 R_2 + (I_2 - I_1) R_3.\end{aligned}$$

Метод узловых потенциалов

Еще одной модификацией метода непосредственного применения правил Кирхгофа является **метод узловых потенциалов**. В этом методе система уравнений строится исходя из потенциалов узлов в качестве неизвестных; уравнения при этом составляются исходя из 1-го правила Кирхгофа. Поскольку важна лишь разность потенциалов, то значение потенциала одного из узлов можно положить равным нулю (заземлить).

Какой метод выгоднее использовать – зависит от топологии электрической цепи. Для одних цепей наименьшее количество уравнений дает метод контурных токов, для других – метод узловых потенциалов. Кроме того, существует ряд других методов расчета цепей – метод эквивалентного генератора, принцип суперпозиции и т.д., так же для расчета цепей широко применяются специализированные программные комплексы.

Конденсаторы и катушки индуктивности

Рассмотрим влияние конденсаторов и катушек индуктивности в цепях постоянного тока. В этих цепях ни токи, ни напряжения не зависят от времени. Для конденсатора связь между током, текущим через него, и падением напряжения на нем определяется формулой

$$\frac{dU_C}{dt} = \frac{1}{C} I,$$

где C – величина емкости конденсатора. Отсюда немедленно получаем, что в рассматриваемом случае через конденсатор ток не течет; конденсатор в этом случае можно рассматривать как разрыв цепи.

ЭДС, наводимая катушкой индуктивности, связана с протекающим через нее током:

$$E_L = -L \frac{dI}{dt},$$

где L – величина индуктивности катушки, а формула записана в системе единиц СИ. Отсюда следует, что в цепях постоянного тока катушка индуктивности ЭДС не наводит; ее в этом случае можно рассматривать как обычный провод с очень малым сопротивлением.

Нелинейные элементы

Рассмотрим пример, когда в цепь внесен нелинейный элемент – полупроводниковый диод. Какой ток установится в цепи, если известна **вольтамперная**

характеристика диода – зависимость протекающего через него тока от падения на нем напряжения $I(U_d)$?

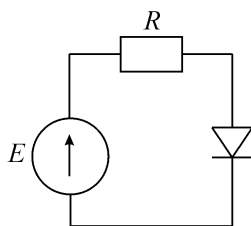


Рисунок 4. Схема с нелинейным элементом.

Запишем для этой цепи 2-е правило Кирхгофа

$$E = IR + U_d.$$

Отсюда ток в цепи

$$I = \frac{E - U_d}{R}.$$

Ток в цепи должен удовлетворять как этому уравнению, так и вольтамперной характеристике $I(U_d)$. Решение получившейся системы нелинейных уравнений (рабочую точку) можно найти графически – оно определяется пересечением графиков этих двух функций (см. рис. 5).

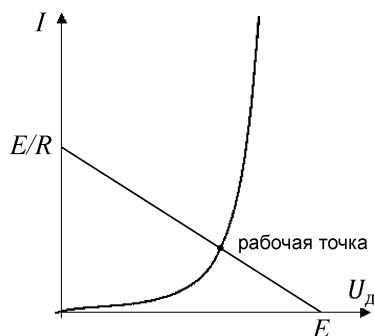


Рисунок 5. Определение рабочей точки в схеме с нелинейным элементом.

1.2 Переходные процессы

Под **переходным процессом** в электрических цепях **понимается** процесс перехода цепи из одного установившегося состояния в другое. Переходные процессы возникают при любых изменениях режима электрической цепи: при подключении и отключении цепи, при изменении нагрузки, при возникновении аварийных режимов (короткое замыкание, обрыв провода и т.д.). Изменения в электрической цепи можно представить в виде тех или иных переключений, называемых в общем случае коммутацией.

Переходные процессы обычно быстро протекающие: длительность их составляет десятые, сотые, а иногда и миллионные доли секунды. Сравнительно редко длительность переходных процессов достигает секунд и десятков секунд. Тем не менее, изучение переходных процессов весьма важно, поскольку в течение этого времени возможно увеличение токов и напряжений на элементах цепи

в десятки раз по сравнению с их значениями в установившихся режимах. Кроме того, переходные процессы лежат в основе работы многих электротехнических устройств, особенно устройств промышленной электроники.

Если бы в цепи не было конденсаторов и катушек индуктивности (реактивных элементов), а присутствовали только резисторы (активные элементы), то тогда токи и напряжения изменялись бы мгновенно. Реактивные же элементы, в отличие от резисторов, способны запасать энергию. Конденсатор запасает энергию в электрическом поле, величина этой энергии пропорциональна квадрату напряжения на нем, катушка индуктивности запасает энергию в магнитном поле, ее величина пропорциональна квадрату тока через катушку. Запасенная энергия не может измениться мгновенно (например, для мгновенного увеличения запасенной энергии требуется внешний источник энергии бесконечно большой мощности), следовательно, переходные процессы в цепях с реактивными элементами занимают определенное время. Отсюда же следуют **правила коммутации**:

1. при переходных процессах напряжения на конденсаторах не могут изменяться мгновенно;
2. при переходных процессах токи, текущие через катушки индуктивности, не могут изменяться мгновенно.

Переходные процессы в цепях с одним реактивным элементом

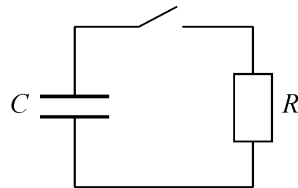


Рисунок 6. Схема разрядки конденсатора через резистор.

Рассмотрим простейший пример – разрядка изначально заряженного до напряжения U_0 конденсатора через резистор с сопротивлением R , соответствующая схема изображена на рисунке 6. Пусть ключ замыкается в момент времени $t = 0$. Запишем для этой цепи 2-е правило Кирхгофа:

$$0 = U_C(t) + U_R(t) = U_C(t) + I(t)R,$$

где U_C – напряжение на конденсаторе. Для дальнейшего рассмотрения этого уравнения продифференцируем его по времени и воспользуемся связью между током, текущим через конденсатор, и напряжением на нем:

$$0 = \frac{dU_C(t)}{dt} + \frac{dI(t)}{dt}R = \frac{1}{C}I(t) + \frac{dI(t)}{dt}R, \quad \frac{dI(t)}{dt} = -\frac{1}{RC}I(t).$$

Задача расчета переходного процесса в цепи путем использования правил Кирхгофа свелась к дифференциальному уравнению; в более общем случае получается система дифференциальных уравнений. Если ввести *постоянную времени цепи* $\tau = RC$, то решение дифференциального уравнения есть

$$I(t) = I_0 \exp(-t/\tau).$$

Напряжение на конденсаторе выразим из исходного уравнения

$$U_C(t) = -I(t)R = -I_0 R \exp(-t/\tau).$$

Чтобы найти неизвестную величину I_0 , воспользуемся первым правилом коммутации. До переходного процесса, при $t < 0$, напряжение на конденсаторе было равно U_0 . Сразу после замыкания ключа, при $t = 0$, напряжение не изменилось:

$$U_C(0) = -I_0 R = U_0 \Rightarrow U_C(t) = U_0 \exp(-t/\tau).$$

Напряжение на конденсаторе при ее разрядке через резистор падает экспоненциально.

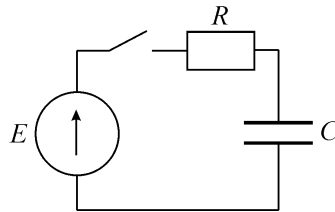


Рисунок 7. Схема зарядки конденсатора через резистор.

Рассмотрим похожий пример – зарядку конденсатора от источника постоянной ЭДС через резистор (см. схему на рис. 7). Изначально конденсатор разряжен, ключ замыкается в момент времени $t = 0$. Согласно 2-му правилу Кирхгофа:

$$E = U_C(t) + U_R(t) = U_C(t) + I(t)R.$$

После дифференцирования по времени получим уравнение, полностью аналогичное рассмотренному в предыдущем примере, изменились только начальные условия. Сразу запишем конечное выражение:

$$U_C(t) = E[1 - \exp(-t/\tau)].$$

Колебательный контур. Добротность

Пусть, в отличие от предыдущего случая, в цепи содержатся два реактивных элемента – конденсатор и катушка индуктивности. Схема такой цепи приведена на рис. 8.

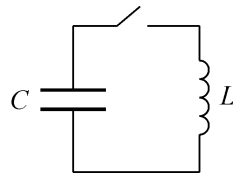


Рисунок 8. Схема разрядки конденсатора через катушку индуктивности.

Изначально конденсатор заряжен до напряжения U_0 , ключ замыкается в момент времени $t = 0$. Записываем уравнение согласно 2-му правилу Кирхгофа:

$$E_L = U_C.$$

Давайте далее будем говорить не об ЭДС, наводимой катушкой индуктивности, а о падении напряжения на ней:

$$U_L = -E_L = L \frac{dI(t)}{dt}.$$

Продифференцировав получающееся уравнение по времени, приходим к однородному линейному дифференциальному уравнению второго порядка

$$\frac{d^2 I(t)}{dt^2} + \frac{1}{LC} I(t) = 0.$$

Такое уравнение описывает гармонические колебания, его общее решение представимо в виде

$$I(t) = I_0 \cos(\omega_0 t + \varphi), \quad \omega_0 = \frac{1}{\sqrt{LC}},$$

параметр ω_0 называют собственной частотой контура. Исходя из второго правила коммутации, в данной схеме ток в начальный момент времени равен нулю, следовательно

$$I(t) = I_0 \sin(\omega_0 t).$$

Найдем, как зависит от времени напряжение на конденсаторе. Исходя из полученного выражения для тока и из начальных условий:

$$U_c(t) = -L \frac{dI}{dt} = -LI_0 \omega_0 \cos(\omega_0 t) = U_0 \cos(\omega_0 t), \quad I_0 = -\sqrt{\frac{C}{L}} U_0.$$

Переходной процесс описывается гармоническими колебаниями тока и напряжения с неубывающей амплитудой. Энергия, изначально запасенная в электрическом поле конденсатора, плавно переходит в энергию магнитного поля катушки индуктивности, и обратно. Соответственно, такая цепь называется колебательным контуром.

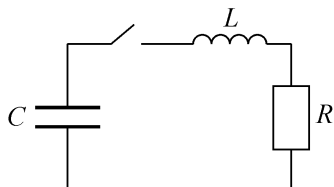


Рисунок 9. RLC цепь.

Рассмотрим, как изменится переходной процесс в цепи, если в нее добавить последовательно включенный резистор (см. рис. 9). Начальные условия зададим совпадающими с предыдущим случаем. Дифференциальное уравнение, описывающее переходной процесс в данной цепи, выглядит следующим образом:

$$U_L + U_R + U_C = 0.$$

После дифференцирования по времени получаем уравнение

$$L \frac{d^2 I}{dt^2} + R \frac{dI}{dt} + \frac{1}{C} I = 0.$$

Решая данные уравнения при начальных условиях

$$U_c(t = 0) = U_0, \quad I_0(t = 0) = 0,$$

получаем ток в этой цепи и напряжение на конденсаторе

$$I(t) = -\frac{U_0}{\omega L} e^{-\gamma t} \sin(\omega t), \quad U_c(t) = \frac{U_0}{\cos(\varphi)} e^{-\gamma t} \cos(\omega t + \varphi),$$

$$\gamma = \frac{1}{2} \frac{R}{L}, \quad \omega = \sqrt{\omega_0^2 - \gamma^2}, \quad \operatorname{tg}(\varphi) = -\frac{\gamma}{\omega}.$$

Зависимость U_C от времени изображена на рис. 10. Видно, что это квазипериодический процесс, представляющий собой затухающие колебания. Энергия электрического поля, изначально запасенная в конденсаторе, перетекает в энергию магнитного поля, запасаемую в катушке индуктивности, и обратно, в то же время рассеиваясь на резисторе. Если γ много меньше ω_0 , как это обычно бывает, можно говорить о гармоническом процессе со слабо уменьшающейся амплитудой; в этом случае частота колебаний $\omega \approx \omega_0$ с точностью до квадратичного члена по γ/ω_0 .

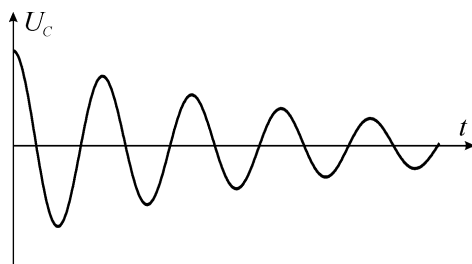


Рисунок 10. Переходный процесс в RLC цепи.

Важнейшим, наряду с частотой колебаний, параметром колебательного контура является **добротность**, определяющая, во сколько раз запасы энергии в системе больше, чем потери энергии за один период колебаний. Общая формула для добротности произвольной колебательной системы имеет вид

$$Q = \omega \frac{W}{P},$$

где W – запасенная в системе энергия, P – средняя мощность потерь. Для рассматриваемого контура запасенную энергию определим в тот момент, когда вся она переходит в энергию магнитного поля катушки (ток становится максимальным):

$$W = \frac{LI^2}{2}.$$

Зная из предыдущих курсов среднюю рассеиваемую на резисторе мощность при протекании через него периодического тока с амплитудой I и учитывая, что реактивные элементы энергию не рассеивают, найдем добротность:

$$P = \frac{RI^2}{2}, \quad Q = \frac{1}{R} \sqrt{\frac{L}{C}} = \frac{\omega_0}{2\gamma}.$$

1.3 Цепи синусоидального переменного тока

Чрезвычайно часто в электрических цепях используются источники переменной ЭДС с гармонической (синусоидальной) зависимостью от времени:

$$E(t) \sim \cos(\omega t + \varphi).$$

Как мы увидим далее, установившиеся напряжения на элементах цепи и токи через них будут иметь такую же зависимость от времени (каждый со своей фазой).

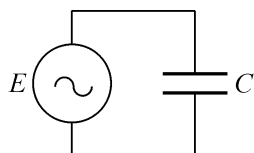


Рисунок 11. Конденсатор, подключенный к источнику синусоидальной ЭДС.

Начнем с рассмотрения простейшей задачи – конденсатора, подключенного к источнику синусоидальной ЭДС (см. рис. 11). Найдем ток, текущий в этой цепи, для чего запишем 2-е правило Кирхгофа:

$$E_0 \cos(\omega t) = U_C.$$

Как и ранее, продифференцируем это выражение по времени и воспользуемся связью между током, текущим через конденсатор, и напряжением на нем:

$$-E_0 \omega \sin(\omega t) = \frac{dU_C(t)}{dt} = \frac{1}{C} I(t), \quad I(t) = -E_0 C \omega \sin(\omega t).$$

Из рисунка 12 видно, что ток через конденсатор достигает максимума на четверть периода раньше, чем напряжение на нем, следовательно, можно сказать, что фаза тока на $\pi/2$ больше.

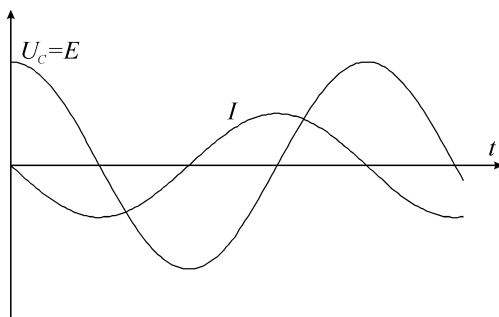


Рисунок 12. Зависимость тока и напряжения на конденсаторе от времени.

Аналогично рассмотрим катушку индуктивности, подключенную к источнику синусоидальной ЭДС. Второе правило Кирхгофа в этом случае примет вид:

$$E_0 \cos(\omega t) = U_L.$$

Используя связь ЭДС, наводимой катушкой индуктивности, с протекающим через нее током, находим:

$$\frac{dI(t)}{dt} = \frac{E_0}{L} \cos(\omega t), \quad I(t) = \frac{E_0}{L\omega} \sin(\omega t),$$

возникающей константой интегрирования можно пренебречь. Из полученного результата видно, что ток через катушку индуктивности достигает максимума на четверть периода позже, чем напряжение на ней, следовательно, можно сказать, что фаза тока на $\pi/2$ меньше.

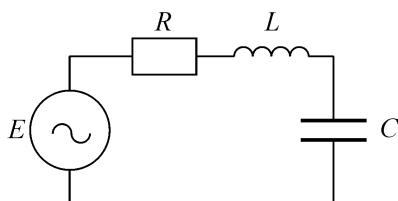


Рисунок 13. Последовательный контур.

Ну а сейчас давайте подключим к источнику синусоидальной ЭДС резистор, конденсатор и катушку индуктивности, собранные вместе в один последовательный контур, как изображено на рисунке 13. Опять запишем 2-е правило Кирхгофа и продифференцируем его по времени:

$$E(t) = E_0 \cos(\omega t) = U_R + U_C + U_L,$$

$$R \frac{dI}{dt} + \frac{1}{C} I + L \frac{d^2 I}{dt^2} = \frac{dE}{dt} = -E_0 \omega \sin(\omega t).$$

Получили линейное дифференциальное уравнение второго порядка с ненулевой правой частью. Как известно, его решение есть сумма общего решения однородного уравнения (с нулевой правой частью) и частного решения неоднородного уравнения. Общее решение зависит от начальных условий, частное – нет.

Как мы видели, решение однородного уравнения описывает переходной процесс, который определяется изначально накопленной в конденсаторе и катушке индуктивности энергией. В силу того, что эта энергия рассеивается (переходит в тепло) на резисторе, это решение является затухающим. По прошествии какого-то времени остается только частное решение, которое определяется балансом мощности, приходящей от внешнего источника ЭДС, и мощности, рассеиваемой резистором. Достаточно очевидно, что оно будет зависеть от времени по синусоидальному закону с той же частотой, что и вынуждающая сила:

$$I(t) = I_0 \cos(\omega t + \varphi).$$

Попробуем найти неизвестные амплитуду I_0 и фазу φ этих вынужденных колебаний. После подстановки этого решения в уравнение получаем:

$$E_0 \omega \sin(\omega t) = I_0 \left[R \omega \sin(\omega t + \varphi) + \left(L \omega^2 - \frac{1}{C} \right) \cos(\omega t + \varphi) \right].$$

Преобразуем это выражение следующим образом:

$$E_0 \sin(\omega t) = I_0 \sqrt{R^2 + \left(L \omega - \frac{1}{C \omega} \right)^2} \left[\frac{R}{\sqrt{R^2 + \left(L \omega - \frac{1}{C \omega} \right)^2}} \sin(\omega t + \varphi) + \frac{\left(L \omega - \frac{1}{C \omega} \right)}{\sqrt{R^2 + \left(L \omega - \frac{1}{C \omega} \right)^2}} \cos(\omega t + \varphi) \right].$$

Можно представить множители, стоящие в скобках, как синус и косинус какого-то угла:

$$\cos \psi = \frac{R}{\sqrt{R^2 + \left(L \omega - \frac{1}{C \omega} \right)^2}}, \quad \sin \psi = \frac{\left(L \omega - \frac{1}{C \omega} \right)}{\sqrt{R^2 + \left(L \omega - \frac{1}{C \omega} \right)^2}}.$$

Действительно, можно проверить, что

$$\cos^2 \psi + \sin^2 \psi = 1.$$

Тогда выражение в скобках есть синус суммы углов

$$[\dots] = \sin(\omega t + \varphi + \psi),$$

Подставим этот результат в ранее полученное выражение:

$$E_0 \sin(\omega t) = I_0 \sqrt{R^2 + \left(L\omega - \frac{1}{C\omega}\right)^2} \sin(\omega t + \varphi + \psi).$$

Так как это выражение должно выполняться для каждого момента времени, окончательно находим амплитуду и фазу вынужденных колебаний тока в последовательном контуре:

$$I_0 = \frac{E_0}{\sqrt{R^2 + \left(L\omega - \frac{1}{C\omega}\right)^2}}, \quad \varphi = -\psi = -\arctan \frac{L\omega - \frac{1}{C\omega}}{R}.$$

Подытожим полученный результат – в случае синусоидальной зависимости ЭДС от времени при помощи правил Кирхгофа получаем дифференциальное уравнение (в общем случае систему дифференциальных уравнений), отыскание частного решения которого даже в простейшем случае оказывается достаточно громоздким. Здесь на помощь приходит *метод комплексных амплитуд*. На первый взгляд этот метод кажется довольно непонятным, однако он задает некий формализм, при помощи которого оказывается очень просто рассчитывать цепи синусоидального тока.

Пусть какая-либо величина гармонически зависит от времени:

$$x(t) = x_0 \cos(\omega t + \varphi).$$

Давайте этой величине сопоставим комплексную величину

$$\hat{x}(t) = x_0 [\cos(\omega t + \varphi) + i \sin(\omega t + \varphi)] = x_0 e^{i(\omega t + \varphi)} = x_0 e^{i\varphi} e^{i\omega t} = \hat{x}_0 e^{i\omega t}.$$

Очевидно, что

$$x(t) = \operatorname{Re}(\hat{x}(t)).$$

Комплексная величина \hat{x}_0 называется *комплексной амплитудой*. Видно, что модуль комплексной амплитуды, как и модуль самой величины $\hat{x}(t)$, равен амплитуде колебаний x_0 .

Далее, пусть нам необходимо совершить над величиной $x(t)$ какую-либо линейную операцию – суммирование, умножение на число, дифференцирование и т.д. Оказывается, что можно это сделать непосредственно, а можно уйти в комплексную плоскость, создав $\hat{x}(t)$, над этой величиной проделать требуемую операцию, а потом, отбросив от получившегося результата мнимую часть, получить окончательный ответ. Отсюда следует, что если $x(t)$ удовлетворяет некоторому линейному уравнению, то этому же уравнению удовлетворяет и ее комплексный аналог. Основное достоинство метода комплексных амплитуд заключается в том, что дифференцирование в нем заменяется умножением. Действительно, найдем по этому методу

$$\frac{dx(t)}{dt} = \operatorname{Re} \left(\frac{d\hat{x}(t)}{dt} \right), \quad \frac{d\hat{x}(t)}{dt} = i\omega x_0 e^{i(\omega t + \varphi)} = i\omega \hat{x}(t).$$

Вернемся обратно к дифференциальному уравнению, заменим в нем все синусоидально зависящие от времени величины (включая ЭДС) на их комплексные аналоги, и будем искать его решение в виде $\hat{I}(t) = I_0 e^{i(\omega t + \varphi)}$:

$$\hat{E}(t) = E_0 e^{i\omega t}, \quad \frac{d\hat{E}}{dt} = i\omega E_0 e^{i\omega t} = i\omega R\hat{I} + \frac{1}{C}\hat{I} - \omega^2 L\hat{I}.$$

Отсюда сразу находим комплексный ток

$$\hat{I}(t) = \frac{E_0}{R + i\left(\omega L - \frac{1}{\omega C}\right)} e^{i\omega t}.$$

Осталось вспомнить формулу Эйлера:

$$a + ib = r \exp(i\varphi), \quad r = \sqrt{a^2 + b^2}, \quad \operatorname{tg}(\varphi) = b/a$$

и преобразовать по этой формуле знаменатель полученного выражения:

$$\hat{I}(t) = \frac{E_0}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} e^{i(\omega t - \psi)}, \quad \operatorname{tg}(\psi) = \left(\omega L - \frac{1}{\omega C}\right)/R,$$

$$I(t) = \operatorname{Re}(\hat{I}(t)) = \frac{E_0}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} \cos(\omega t - \psi).$$

Таким образом, благодаря комплексному формализму, мы практически мгновенно получили частное решение исходного дифференциального уравнения и рассчитали ток в рассматриваемой цепи. Можно расширить этот формализм для расчета цепей любой сложности, введя понятие *импеданса* – отношения комплексного напряжения на элементе к протекающему через него комплексному току. Для линейных элементов это отношение не зависит от времени, амплитуд и фаз токов и напряжений, а определяется только частотой периодического сигнала и параметрами элемента.

Рассмотрим конденсатор емкостью C . Используя связь между текущим через него током и падением напряжения на нем, получаем

$$i\omega \hat{U}_C = \frac{1}{C} \hat{I}_C.$$

Отсюда получаем импеданс конденсатора

$$Z_C = \frac{\hat{U}_C}{\hat{I}_C} = \frac{1}{i\omega C}.$$

Аналогично получается импеданс катушки индуктивности и резистора:

$$Z_L = \frac{\hat{U}_L}{\hat{I}_L} = i\omega L, \quad Z_R = R.$$

Обращаем внимание, что импеданс резистора – чисто действительное число, а импедансы реактивных элементов – чисто мнимые.

Наконец, можно сформулировать правила Кирхгофа для расчета цепей, содержащих реактивные элементы, в установившемся режиме при гармонических

входных сигналах. При этом все токи и напряжения рассматриваются в комплексном виде. Первое правило Кирхгофа при этом будет гласить, что для всех узлов цепи алгебраическая сумма комплексных токов равна нулю:

$$\sum_i \hat{I}_i = 0.$$

Второе правило связывает алгебраическую сумму комплексных ЭДС контура с алгебраической суммой произведения комплексных токов, текущих через ветви контура на импедансы этих ветвей:

$$\sum_i \hat{E}_i = \sum_j \hat{I}_j Z_j.$$

Т.е. все линейные элементы цепи трактуем как резисторы с особыми, комплексными сопротивлениями. При этом правила нахождения полного импеданса цепи, составленной из параллельно или последовательно включенных отдельных элементов будут такие же, как и для резисторов.

Рассмотренные ранее методы контурных токов и узловых потенциалов модифицируются аналогичным образом.

1.4 Резонанс

Давайте еще раз рассмотрим схему, которая содержит резистор, катушку индуктивности и конденсатор, последовательно подключенные к источнику синусоидальной ЭДС (см. рис. 13). Запишем 2-е правило Кирхгофа для таких цепей:

$$E_0 e^{i\omega t} = (Z_R + Z_L + Z_C) \hat{I}, \quad \hat{I}(t) = \frac{E_0}{R + i\omega L + \frac{1}{i\omega C}} e^{i\omega t} = \frac{E_0}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} e^{i(\omega t - \psi)},$$

Полученный результат, естественно, одинаков с ранее полученными решениями другими методами – нахождением решения дифференциального уравнения непосредственной подстановкой и решением методом комплексных амплитуд.

Найдем напряжение на одном из реактивных элементов, скажем, конденсаторе, в зависимости от времени:

$$\hat{U}_C(t) = Z_C \hat{I}(t) = \frac{E_0}{\omega C \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} e^{i\left(\omega t - \psi - \frac{\pi}{2}\right)}.$$

И ток в цепи, и напряжение на конденсаторе изменяются по гармоническому закону; амплитуды этих колебаний $I_0 = |\hat{I}|$ и $U_{C0} = |\hat{U}_C|$ сильно зависят от частоты внешней вынуждающей силы – в данном случае, источника ЭДС. Эти зависимости представлены на рисунке 14; видно, что при частоте $\omega_0 = 1/\sqrt{LC}$ амплитуды колебаний имеют резкий максимум. Это есть явление *резонанса*, имеющее место при периодическом внешнем воздействии на колебательную систему при условии совпадения частоты вынуждающей силы с собственной частотой системы.

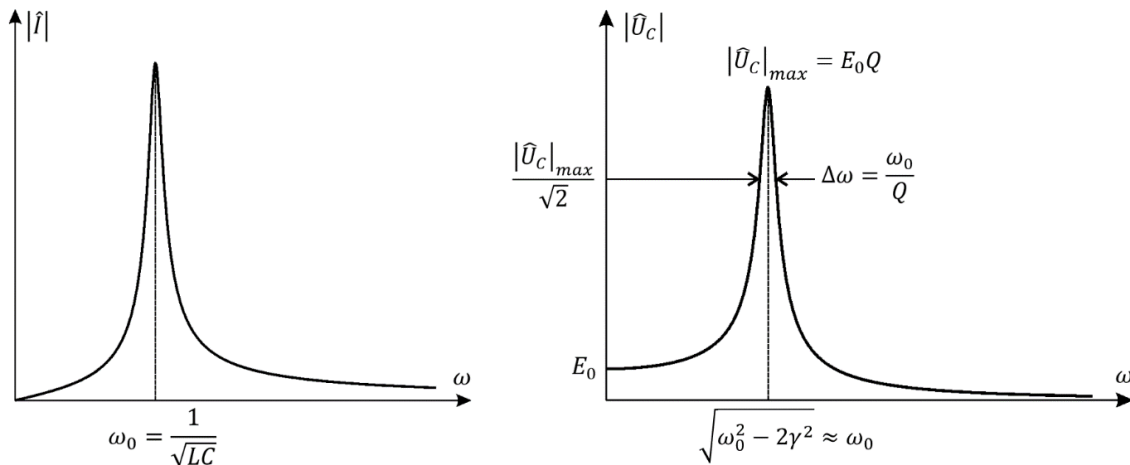


Рисунок 14. Зависимость амплитуды колебаний тока и напряжения на реактивном элементе от частоты внешней ЭДС в последовательном колебательном контуре.

В последовательном контуре на резонансной частоте напряжение на реактивных элементах одинаково и много больше амплитуды ЭДС:

$$U_{C0}|_{\omega=\omega_0} = U_{L0}|_{\omega=\omega_0} = E_0 Q \gg E_0.$$

Такой резонанс называют *резонансом напряжений*. Часто используется цепь с другим расположением элементов – параллельный контур (см. рис. 15). В этом случае уже токи через катушку и конденсатор на резонансной частоте во много раз превышают ток источника; в этом случае говорят о *резонансе токов*.

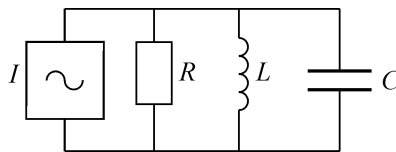


Рисунок 15. Параллельный колебательный контур.

Важно подчеркнуть, что в обоих рассмотренных контурах на резонансной частоте полный импеданс цепи становится чисто действительным, мнимые члены в импедансе взаимно компенсируют друг друга. Это условие является критерием резонанса в более сложных электрических цепях.

1.5 Фильтры

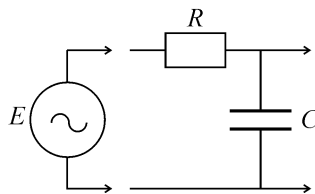


Рисунок 16. RC-цепь.

Давайте рассмотрим источник синусоидальной ЭДС, подключенный к цепи, содержащей резистор и конденсатор (*RC-цепь*). Будем рассматривать источник ЭДС как входящий сигнал, а напряжение на конденсаторе – как выходящий, как изображено на рисунке 16. Расчет этого напряжения проведем, используя 2-ое правило Кирхгофа для цепей синусоидального тока:

$$\hat{I}(t) = \frac{\hat{E}(t)}{Z_R + Z_C} = \frac{E_0}{R + \frac{1}{i\omega C}} e^{i\omega t} = \frac{E_0}{\sqrt{R^2 + \left(\frac{1}{\omega C}\right)^2}} e^{i(\omega t + \psi)}, \quad \text{tg}(\psi) = \frac{1}{\omega CR},$$

$$\hat{U}_C = Z_C \hat{I} = \frac{E_0}{\sqrt{1 + (\omega CR)^2}} e^{i(\omega t + \psi - \frac{\pi}{2})} = U_{C0} e^{i(\omega t + \psi - \frac{\pi}{2})}.$$

Введем величину, показывающую отношение амплитуды выходного сигнала к входному:

$$K(\omega) = \frac{U_{C0}}{E_0} = \frac{1}{\sqrt{1 + (\omega CR)^2}}.$$

Поскольку эта величина, помимо параметров цепи, определяется еще и частотой колебаний, такую схему можно рассматривать как фильтр, передающий с входа на выход сигналы одних частот практически без уменьшения амплитуды, а других частот – сильно ослабленными. Величина $K(\omega)$ называется *коэффициентом передачи фильтра*; в данном случае $K(\omega)$ всегда меньше единицы, ее зависимость от частоты показана на рис. 17. При малых частотах $K(\omega) \approx 1$, при больших частотах этот коэффициент уменьшается до нуля; фильтр с такой зависимостью коэффициента передачи называют *фильтром низких частот (ФНЧ)*.

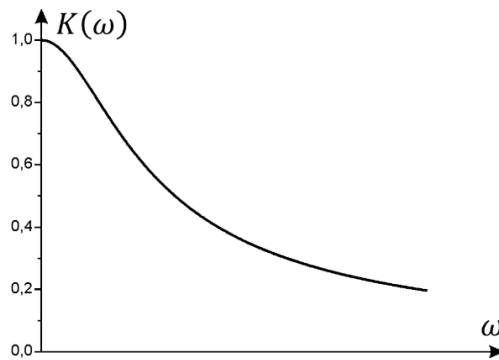


Рисунок 17. Зависимость коэффициента передачи фильтра от частоты.

Для дальнейшего рассмотрения фильтров нам пригодится такая логарифмическая единица как децибел. Предположим, нам нужно сравнить две “энергетические” величины A_1 и A_2 – мощность, энергию, плотность потока мощности и т.п. Можно просто взять их отношение A_1/A_2 , однако это не всегда удобно, поскольку часто эти величины могут сильно – на несколько порядков – отличаться. Было предложено использовать для сравнения десятичный логарифм отношения

$$A_B = \lg\left(\frac{A_1}{A_2}\right),$$

эту величину называли *бел* в честь изобретателя телефона Александра Белла. Однако и она оказалась неудобна – получается слишком малое число (перебор!). На практике оказалось более удобным использовать десятые доли бела – децибелы:

$$A_{дБ} = 10 \lg\left(\frac{A_1}{A_2}\right).$$

Для периодических сигналов энергетические единицы пропорциональны квадрату амплитуды сигнала (в данном случае напряжения), поэтому

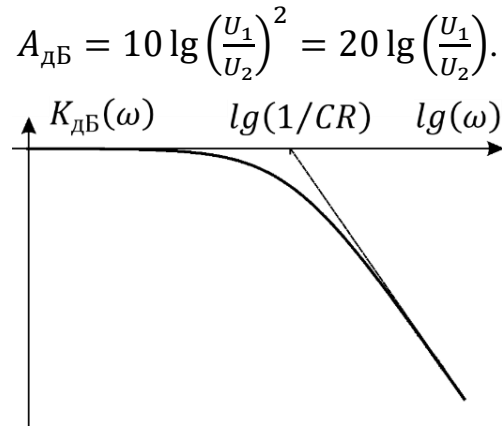


Рисунок 18. Логарифмический масштаб зависимости $K_{\text{дБ}}(\omega)$.

Возвращаясь к рассмотрению фильтров, выразим коэффициент передачи RC фильтра в децибелах:

$$K_{\text{дБ}}(\omega) = 20 \lg \left(\frac{1}{\sqrt{1+(\omega CR)^2}} \right) \approx \begin{cases} 0, & \text{при } \omega CR \ll 1 \\ -20 \lg(\omega CR), & \text{при } \omega CR \gg 1 \end{cases}$$

Благодаря таким приближениям точный график зависимости $K_{\text{дБ}}$ от $\lg(\omega)$ можно приближенно представить двумя прямыми – $K_{\text{дБ}}(\omega) = 0$ при $\omega < 1/CR$ и спадающей прямой $K_{\text{дБ}}(\omega) = 20 [\lg(1/CR) - \lg(\omega)]$ при $\omega > 1/CR$ – см. рис. 18. Соответственно, можно приближенно считать, что частоты из первой области проходят без ослабления, а из второй – ослабляются с ослаблением 20 децибел на декаду.

Поясним последнее утверждение – пусть какой либо частоте ω_1 из второй области соответствует коэффициент передачи $K_{\text{дБ}}(\omega_1)$. Увеличим частоту в 10 раз (на декаду): $\omega_2 = 10\omega_1$. Из формулы для спадающей прямой видно, что коэффициент передачи при этом уменьшится на 20 дБ, независимо от параметров фильтра. Частота $\omega_0 = 1/CR$ называется *частотой среза*, при этой частоте точное значение коэффициента передачи фильтра составляет –3 дБ, что соответствует двукратному уменьшению мощности сигнала.

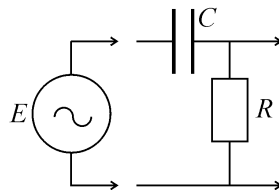


Рисунок 19. Фильтр высоких частот.

Если в исходной схеме RC-цепи выходное напряжение снимать не с конденсатора, а с резистора (как показано на рисунке 19), то так же получим фильтр, который только ослабляет низкие частоты, пропуская высокие частоты практически без изменений – *фильтр высоких частот (ФВЧ)*.

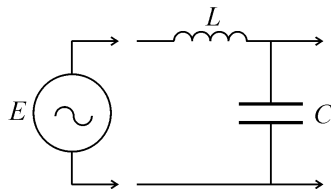


Рисунок 20. LC фильтр.

В силовых электрических цепях активно применяют LC фильтр, изображенный на рисунке 20. Коэффициент передачи такого фильтра

$$K(\omega) = \frac{1}{1 - (\omega/\omega_0)^2}$$

неограниченно возрастает при приближении к резонансной частоте $\omega_0 = 1/\sqrt{LC}$. В реальных схемах к выходу фильтра подключается нагрузка, которая ограничивает этот острый резонансный пик.

В заключение рассмотрим основные области применения фильтров:

1. в системах электропитания для гашения помех и для сглаживания пульсаций напряжения;
2. в звуковой аппаратуре фильтры применяют в усилительных схемах, в эквалайзерах, в акустических системах для разделения частот на низкие/средние/высокие, и т.д.;
3. для генерации и выделения какой-либо частоты применяют узкополосные перестраиваемые фильтры.

1.6 Широтно-импульсная модуляция

Давайте разберем одну задачу, которая, на первый взгляд, отстоит далеко от рассматриваемой темы. Пусть мы должны поддерживать постоянной температуру какого-либо тела, используя только внешний нагреватель, причем мощность этого нагревателя мы изменять не можем, а можем его только включать и выключать. Простейшее решение – включать нагреватель, если температура тела меньше заданной, и выключать, если больше – не подходит, поскольку вследствие инерционности нагревателя колебания температуры тела будут слишком большими. Давайте поступим иначе – будем включать и выключать нагреватель достаточно часто, настолько, чтобы тело не успело сколько-нибудь изменить свою температуру за один такой цикл. Мощность нагрева, поступающая от нагревателя телу, будет определяться тогда соотношением времен включения и выключения, и если эту величину регулировать при помощи обратной связи, то можно добиться высокой стабильности температуры. Такая модель управления подводимой к нагрузке мощностью путём изменения скважности импульсов при постоянной частоте их следования называется *широтно-импульсной модуляцией*. Эта модуляция применяется не только для стабилизации температуры, но и во многих областях техники.

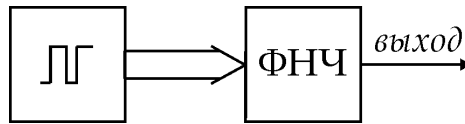


Рисунок 21. Схема ШИМ.

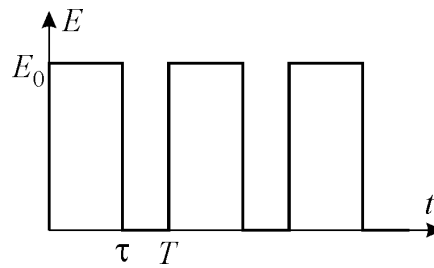


Рисунок 22. Периодический сигнал, формируемый источником.

Рассмотрим источник прямоугольного периодического сигнала, подключенный к фильтру низких частот, как показано на рис. 21. Сигнал имеет период T , время включения τ и амплитуду E_0 (см. рис. 22). Что будет на выходе фильтра, если его частота среза много меньше частоты сигнала? Разложим входной сигнал в ряд Фурье:

$$E(t) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos(n\omega t + \varphi_n),$$

где $\omega = 2\pi/T$ – частота основной гармоники. ФНЧ ослабляет периодический сигнал, что математически выражается в уменьшении амплитуд гармоник выходного сигнала относительно входного, причем чем больше номер n гармоники, тем сильнее уменьшается ее амплитуда. В рассматриваемой схеме фильтр очень сильно ослабляет амплитуды всех гармоник, соответственно, их можно рассматривать как малые колебания около постоянного уровня, определяемого постоянным коэффициентом в разложении Фурье, который пропускается фильтром низких частот без изменения. Вспоминая формулу для его отыскания, сразу определяем этот уровень:

$$U_{\text{вых}} \approx \frac{A_0}{2} = \frac{1}{T} \int_0^T E(t) dt = E_0 \frac{\tau}{T}.$$

Мы видим, что уровень постоянного выходного сигнала можно регулировать путем изменения скважности (отношения T/τ) периодического входного сигнала. Эта схема имеет ряд достоинств, вот только некоторые из них: во-первых, ее можно просто и компактно реализовать на практике; во-вторых, в отличие от прочих схем ее КПД очень велик – ключи, формирующие периодический сигнал, практически не потребляют мощности. На основе этой схемы устроены многие современные источники питания, например, блоки питания компьютеров.

1.7 Применение закона электромагнитной индукции

Закон электромагнитной индукции, открытый английским физиком Майклом Фарадеем, гласит, что для любого замкнутого контура индуцированная электродвижущая сила равна скорости изменения магнитного потока Φ , проходящего через этот контур:

$$E = -\frac{d\Phi}{dt} \text{ (в системе СИ), } E = -\frac{1}{c} \frac{d\Phi}{dt} \text{ (в системе СГС).}$$

Этот закон лежит в основе электротехники, которая без преувеличения является ядром современной цивилизации. Причина этому в том, что электрическая энергия является наиболее удобной из форм энергии, доступных человеку, во многом благодаря закону электромагнитной индукции. Действительно, и получение электрической энергии (электрогенераторы), и ее трансформация, и ее преобразование в механическую энергию (электродвигатели) – все это осуществляется согласно закону Фарадея, причем с высоким КПД.

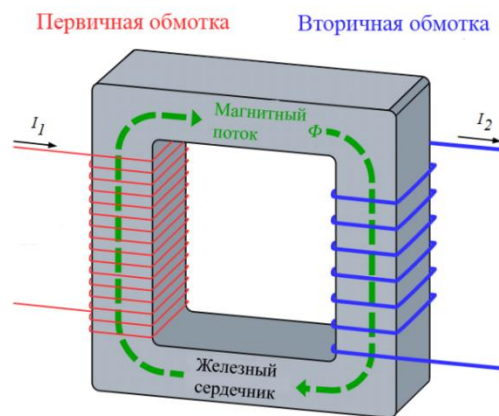


Рисунок 23. Схема трансформатора.

Давайте оставим за рамками работу электродвигателей и генераторов, и рассмотрим работу трансформатора. Обычно он состоит из двух обмоток на общем железном сердечнике, как показано на рисунке 23. Железный сердечник необходим для того, чтобы поток магнитного поля, создаваемый первичной обмоткой, направить сквозь вторичную обмотку. В случае, если нет рассеяния магнитного поля (все силовые линии проходят и через первую, и через вторую обмотки), коэффициент взаимной индукции обмоток максимален и равен

$$M = \sqrt{L_1 L_2},$$

где L_1 и L_2 – коэффициенты самоиндукции обмоток.

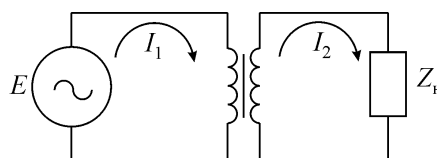


Рисунок 24. Эквивалентная схема трансформатора.

Непосредственно расчет трансформатора будем проводить на основе его эквивалентной схемы, представленной на рисунке 24. Запишем для цепей каждой из обмоток 2-е правило Кирхгофа:

$$\begin{cases} \hat{E} + \hat{E}_{L1} = 0 \\ \hat{E}_{L2} = Z_H \hat{I}_2 \end{cases} \quad \begin{cases} \hat{E}_{L1} = -\frac{d}{dt}(L_1 \hat{I}_1 + M \hat{I}_2) = -i\omega(L_1 \hat{I}_1 + M \hat{I}_2) \\ \hat{E}_{L2} = -\frac{d}{dt}(M \hat{I}_1 + L_2 \hat{I}_2) = -i\omega(M \hat{I}_1 + L_2 \hat{I}_2) \end{cases}.$$

Отсюда из второго уравнения выражаем ток в первой обмотке через ток во второй

$$\hat{I}_1 = -\frac{Z_H + i\omega L_2}{i\omega M} \hat{I}_2$$

и подставим его в первое уравнение:

$$\hat{E} = \left[i\omega M - \frac{L_1}{M} (Z_H + i\omega L_2) \right] \hat{I}_2 = -[L_1 Z_H + i\omega(L_1 L_2 - M^2)] \frac{\hat{I}_2}{M}.$$

Напряжение на нагрузке

$$\hat{U}_H = \hat{I}_2 Z_H = -\frac{\hat{E} M Z_H}{L_1 Z_H + i\omega(L_1 L_2 - M^2)}.$$

Обычно рассеяние магнитного потока в трансформаторе очень мало, коэффициент трансформации является вещественной величиной и равен

$$\frac{\hat{U}_H}{\hat{E}} = -\sqrt{\frac{L_2}{L_1}} = -\frac{N_2}{N_1},$$

где N_2/N_1 – отношение числа витков обмоток. Знак этого коэффициента зависит от выбора начала и конца обмоток. При наличии рассеяния коэффициент взаимной индукции $M < \sqrt{L_1 L_2}$, коэффициент трансформации становится комплексным, причем его модуль уменьшается по сравнению со случаем без рассеяния.

1.8 Длинные линии

Рассмотрим коаксиальный кабель, пространство между внешней оплеткой и центральным проводником (жилой) которого заполнено диэлектриком с диэлектрической проницаемостью ε и магнитной проницаемостью $\mu \approx 1$. Из курса электростатики известно, что такая конструкция представляет собой конденсатор, причем чем длиннее кабель, тем больше его емкость. При рассмотрении таких конденсаторов вводят удельную емкость (емкость на единицу длины). Из курса же магнитостатики известно, что у такого кабеля есть индуктивность, так же пропорциональная его длине; так же вводят удельную индуктивность.

На предыдущих лекциях мы изучали, что емкости и индуктивности реактивных элементов существенно сказываются на переменных во времени процессах, происходящих в электрических цепях. Здесь мы имеем аналогичную ситуацию с той лишь разницей, что реактивные элементы не сосредоточены в отдельных элементах, а распределены вдоль всей длины кабеля.

Дадим определение: *длинная линия* – система из двух и более параллельных проводников, длина которых во много раз превышает длину волны распространяющихся в ней электромагнитных колебаний, а расстояние между проводниками, в свою очередь, значительно меньше этой длины волны. Забегая вперед, отметим, что при нарушении второго условия электромагнитные волны могут распространяться под углом к направлению проводников линии. Такие волны быстро затухают по мере движения вдоль линии, и динамика и скорость затухания в сильной степени определяются их частотой, т.е. их рассмотрение есть другая, значительно более сложная задача.

В дальнейшем будем рассматривать однородные линии, все параметры которых не изменяются вдоль их длины.

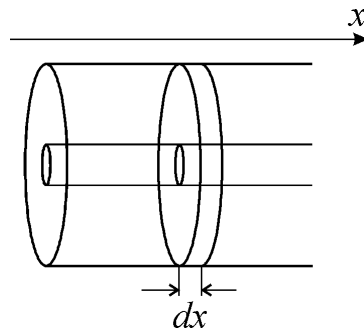


Рисунок 25. Выделение малого отрезка коаксиального кабеля.

Рассмотрим малый отрезок кабеля с удельной емкостью C и удельной индуктивностью L длиной dx , как показано на рисунке 25. У этого отрезка есть емкость Cdx и индуктивность Ldx . Кроме того активные потери при протекании тока по жиле и оплетке можно рассмотреть, если ввести удельное сопротивление линии R . Так же можно рассмотреть случай неидеального (слегка проводящего) диэлектрика, введя удельную проводимость G между жилой и оплеткой. Тогда сопротивление отрезка кабеля равно Rdx , проводимость – Gdx . Получающаяся при этом эквивалентная схема малого отрезка кабеля представлена на рисунке 26. Весь кабель есть последовательное (друг за другом) соединение этих отрезков.

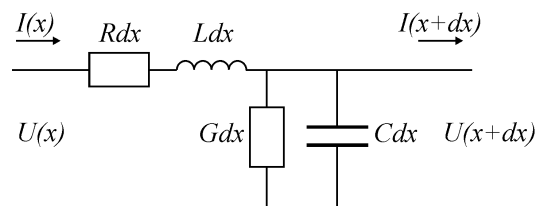


Рисунок 26. Эквивалентная схема малого отрезка кабеля.

Обозначим $I(x, t)$ – ток в центральном проводнике, $U(x, t)$ – напряжение между проводниками. Тогда на выходе из отрезка

$$I(x + dx, t) = I(x, t) - I_G - I_C = I(x, t) - UGdx - \frac{\partial U}{\partial t} Cdx,$$

$$U(x + dx, t) = U(x, t) - U_R - U_L = U(x, t) - IRdx - \frac{\partial I}{\partial t} Ldx.$$

Отсюда легко получаются так называемые *телеграфные уравнения* – система из дифференциальных уравнений в частных производных, описывающая распространение сигналов в длинных линиях:

$$\begin{aligned}\frac{\partial U}{\partial x} &= -RI - L \frac{\partial I}{\partial t}, \\ \frac{\partial I}{\partial x} &= -GU - C \frac{\partial U}{\partial t}.\end{aligned}$$

Далее будем рассматривать линию без потерь: $R = 0$, $G = 0$. Если при этом первое из телеграфных уравнений продифференцировать по x , в правой части поменять последовательность дифференцирования и использовать второе уравнение, то получим:

$$\frac{\partial^2 U}{\partial x^2} = -L \frac{\partial}{\partial x} \left(\frac{\partial I}{\partial t} \right) = -L \frac{\partial}{\partial t} \left(\frac{\partial I}{\partial x} \right) = LC \frac{\partial^2 U}{\partial t^2} = \frac{1}{v^2} \frac{\partial^2 U}{\partial t^2},$$

здесь введено обозначение $v = 1/\sqrt{LC}$. Аналогично получаем

$$\frac{\partial^2 I}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 I}{\partial t^2}.$$

Одним из независимых решений полученного на функцию $U(x, t)$ уравнения является любая функция вида $U_1(t - x/v)$, что можно проверить непосредственно. Это решение представляет собой произвольную волну напряжения, распространяющуюся со скоростью v вдоль оси x в положительном направлении. Вторым независимым решением этого уравнения есть любая функция вида $U_2(t + x/v)$. Поскольку это уравнение второго порядка, то у него больше нет независимых решений, и любое его решение представимо в виде:

$$U(x, t) = U_1(t - x/v) + U_2(t + x/v).$$

Аналогично для тока

$$I(x, t) = I_1(t - x/v) + I_2(t + x/v).$$

Подставим эти решения в первое уравнение системы телеграфных уравнений:

$$-\frac{1}{v} U'_1(t - x/v) + \frac{1}{v} U'_2(t + x/v) = -L(I'_1(t - x/v) + I'_2(t + x/v)).$$

Поскольку волны независимы, то

$$\frac{1}{v} U'_1 = LI'_1, \quad \frac{1}{v} U'_2 = -LI'_2.$$

Пренебрегая ни на что не влияющими постоянными интегрирования, получаем

$$U_1(x, t) = \rho I_1(x, t), \quad U_2(x, t) = -\rho I_2(x, t),$$

здесь введено *волновое сопротивление линии*

$$\rho = Lv = \sqrt{L/C}.$$

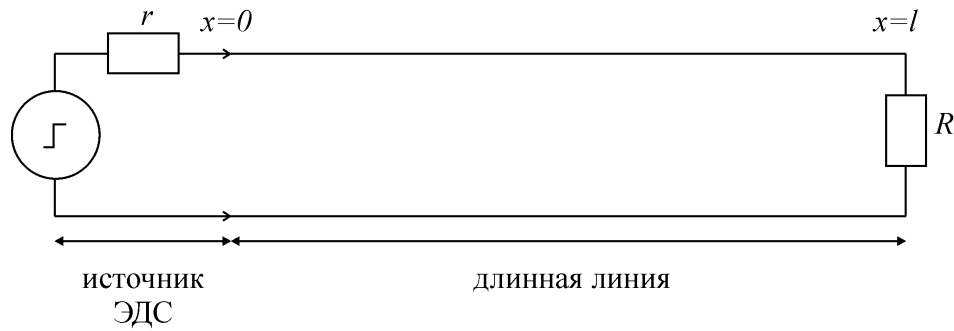


Рисунок 27. Схема длинной линии с нагрузкой.

Рассмотрим изображенную на рисунке 27 схему, состоящую из длинной линии, ко входу которой подключен ступенчатый источник ЭДС с внутренним сопротивлением r , а к выходу – резистор с сопротивлением R .

При $t < l/v$ волна распространяется вдоль линии, не “чувствуя” ее окончания и резистора. В линии накапливается заряд $Q = CU_{\text{вх}}vt$, а из источника в линию идет ток

$$I = \frac{dQ}{dt} = CU_{\text{вх}}v = \frac{U_{\text{вх}}}{\rho} = \text{const.}$$

Линия со стороны источника ЭДС представляется как резистор с сопротивлением ρ , следовательно

$$U_{\text{вх}} = E \frac{\rho}{r + \rho}.$$

В конце линии, на резисторе нагрузки, ток и напряжение связаны по закону Ома:

$$U_{\text{н}} = RI_{\text{н}}.$$

В общем случае, чтобы удовлетворить этим соотношениям, необходимо ввести отраженную волну, которая распространяется в обратном направлении; волну, которая первоначально распространяется от источника, будем называть падающей. Тогда на нагрузке:

$$U_{\text{н}} = U_{\text{п}} + U_{\text{о}}, \quad I_{\text{н}} = I_{\text{п}} - I_{\text{о}}, \quad U_{\text{п}} = I_{\text{п}}\rho, \quad U_{\text{о}} = -I_{\text{о}}\rho.$$

Отсюда можно отношение амплитуды напряжения в отраженной волне к амплитуде напряжения в падающей волне:

$$K_u = \frac{R - \rho}{R + \rho}.$$

Это отношение называют *коэффициентом отражения*.

Рассмотрим различные соотношения волнового сопротивления линии и сопротивления нагрузки:

1. $R = \rho$. Отраженной волны от нагрузки не будет, мощность от источника, первоначально заряжающая линию, начнет рассеиваться на нагрузке.

2. $R = 0$ (короткое замыкание линии) или $R = \infty$ (отсутствие замыкания линии). Отраженная волна при $t = 2l/v$ дойдет до источника ЭДС, но второго отражения, от источника, не произойдет. Источник “увидит” эти значения R , и начнет работать на них. Ток от источника будет $I = E/r$ и $I = 0$ соответственно.
3. Общий случай $R \neq \rho$. По линии будут распространяться последовательно отраженные от источника и нагрузки волны, амплитуда тока в которых будет сходиться к $I = E/R$.

Остается добавить, что не рассмотренные нами потери в линии ведут как к ослаблению распространяющегося сигнала, так и к искажению его формы, поскольку скорость распространения различных составляющих его гармоник становится разной; это явление мы подробно разберем в следующих лекциях.

1.9 Квазистационарное приближение. Скин-эффект

Запишем систему уравнений Максвелла (с системе единиц СГС)

$$\begin{aligned} \operatorname{rot} \mathbf{E} &= -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \quad \operatorname{div} \mathbf{B} = 0, \\ \operatorname{rot} \mathbf{H} &= \frac{4\pi}{c} \mathbf{j} + \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t}, \quad \operatorname{div} \mathbf{D} = 4\pi\rho. \end{aligned}$$

В правой части этих уравнений стоят источники полей. Можно сказать, что электрическое поле возбуждается электрическими зарядами и изменяющимся во времени магнитным полем, а магнитное поле – изменяющимся во времени электрическим полем. Оказывается, что в очень широком круге явлений членом с $\partial \mathbf{D}/\partial t$ можно пренебречь. Такое приближение называется *квазистационарным*.

В квазистационарном приближении производной $\partial \mathbf{B}/\partial t$ отнюдь не пренебрегают. На первый взгляд, здесь есть противоречие. Рассмотрим вопрос детальнее. Пусть характерные размеры системы L , а характерное время изменения полей T . Тогда производные по времени имеют порядок $1/T$, а по координатам – $1/L$. Предположим, что производная $\partial \mathbf{B}/\partial t$ имеет тот же порядок, что $c \operatorname{rot} \mathbf{E}$, то есть

$$c E/L \sim B/T.$$

Тогда отношение члена $\partial \mathbf{D}/\partial t$ к члену $c \operatorname{rot} \mathbf{H}$ будет порядка

$$\frac{D/T}{c H/L} \sim \frac{E L}{B c T} \sim \left(\frac{L}{c T} \right)^2.$$

Вспомним, что электромагнитные волны распространяются со скоростью света c (в среде – чуть меньше, но это не важно). Если размеры системы L малы по сравнению с расстоянием cT , которое пробегает электромагнитная волна за характерное время изменения полей, то вычисленное отношение много меньше

единицы, и действительно можно пренебрегать производной $\partial \mathbf{D} / \partial t$, сохраняя в то же время производную $\partial \mathbf{B} / \partial t$.

Кроме того, член с $\partial \mathbf{D} / \partial t$ должен быть много меньше плотности тока свободных зарядов \mathbf{j} . Запишем это условие в оценке по порядку величины, используя закон Ома в дифференциальной форме:

$$D / T \sim E / T \ll j = \sigma E, \quad T \sigma \gg 1.$$

Для проводников это условие с запасом выполняется для всех интересующих нас частот.

С физической точки зрения квазистационарное приближение соответствует, с одной стороны, отсутствию запаздывания, связанного с конечным временем распространения электромагнитных волн. С другой стороны, квазистационарное приближение соответствует отсутствию излучения ЭМ волн.

Рассмотрим, какие уравнения можно получить из системы уравнений Максвелла в квазистационарном приближении, если рассматривать поля в проводниках. В проводниках плотность тока свободных зарядов полностью определяется токами проводимости $\mathbf{j} = \sigma \mathbf{E}$; сделаем эту замену в соответствующем уравнении и применим операцию ротора к обеим его частям:

$$\text{rot rot } \mathbf{H} = \frac{4\pi}{c} \sigma \text{ rot } \mathbf{E} = -\frac{4\pi}{c^2} \sigma \frac{\partial \mathbf{B}}{\partial t}.$$

Вспомним одну из формул векторного анализа

$$\text{rot rot} = \text{grad div} - \Delta.$$

Здесь Δ – векторный оператор Лапласа, который действует на векторное поле и имеет векторное значение. В прямоугольной системе координат $\Delta \mathbf{A}$ есть вектор, компонентами которого являются скалярные лапласианы компонент поля \mathbf{A} :

$$\Delta \mathbf{A} = (\Delta A_x, \Delta A_y, \Delta A_z).$$

Учитывая, что $\text{div } \mathbf{B} = 0$, для однородной среды ($\mu = \text{const}$) получаем

$$\frac{\partial \mathbf{B}}{\partial t} = \frac{c^2}{4\pi\sigma\mu} \Delta \mathbf{B}.$$

Аналогичное уравнение для поля \mathbf{E} получается при исключении из системы уравнений Максвелла магнитного поля

$$\frac{\partial \mathbf{E}}{\partial t} = \frac{c^2}{4\pi\sigma\mu} \Delta \mathbf{E}$$

при условии равенства нулю плотности свободных зарядов. Подобными уравнениями описываются диффузионные процессы, в частности, теплопроводность (с единственным несущественным отличием, что температура – это скалярное поле).

Рассмотрим пример – длинный проводящий цилиндр, который погружен в стационарное и однородное магнитное поле, направленное вдоль его оси. Пусть

в нулевой момент времени это поле достаточно быстро выключают. Внутри цилиндра магнитное поле также начинает уменьшаться во времени, но возникающие в толще проводника вихревые токи препятствуют его быстрому ослаблению. Потом, постепенно, эти токи затухают, и поле внутри цилиндра затухает вместе с ними. Оказывается, задача уменьшения поля в цилиндре равнозначна задаче остывания равномерно нагретого цилиндра! Диффузия тепла из внутренних областей вызывает изменение температуры со временем как показано на рисунке 28.

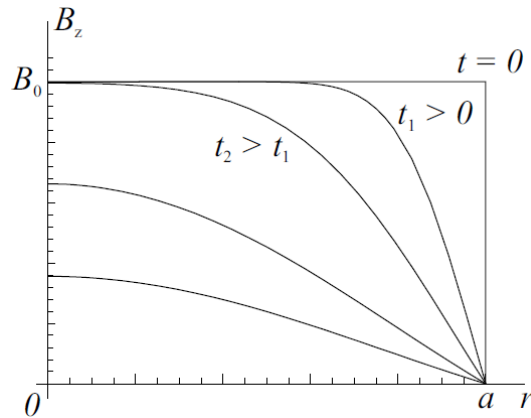


Рисунок 28. Изменение во времени магнитного поля внутри цилиндра.

Важнейшим частным случаем является синусоидальная зависимость полей **E** и **B** от времени. В этом случае каждая интересующая нас физическая величина (в данном случае компоненты полей) будет иметь следующий вид:

$$f(\mathbf{r}, t) = f_0(\mathbf{r}) \cos(\omega t + \varphi(\mathbf{r})).$$

Как и ранее, для упрощения работы с такими величинами введем их представление в комплексном виде:

$$\begin{aligned} \hat{f}(\mathbf{r}, t) &= f_0(\mathbf{r}) [\cos(\omega t + \varphi(\mathbf{r})) + i \sin(\omega t + \varphi(\mathbf{r}))] = f_0(\mathbf{r}) e^{i(\omega t + \varphi(\mathbf{r}))} = \\ &= \hat{f}(\mathbf{r}) e^{i\omega t}. \end{aligned}$$

Как и ранее, величину

$$\hat{f}(\mathbf{r}) = f_0(\mathbf{r}) e^{i\varphi(\mathbf{r})}$$

будем называть *комплексной амплитудой*. Видно, что амплитуда изменения физической величины равна модулю комплексной амплитуды:

$$f_0(\mathbf{r}) = |\hat{f}(\mathbf{r})|.$$

Как мы увидим в дальнейшем, такой поход имеет важные преимущества при сложении полей в точке наблюдения от нескольких источников.

Вернемся к уравнениям для полей. В случае гармонической зависимости от времени поля можно представить в виде

$$\mathbf{E}(\mathbf{r}, t) = \text{Re}(\hat{\mathbf{E}}(\mathbf{r}) e^{i\omega t}), \quad \mathbf{B}(\mathbf{r}, t) = \text{Re}(\hat{\mathbf{B}}(\mathbf{r}) e^{i\omega t}).$$

Выведенные выше уравнения для комплексной амплитуды примут вид

$$\Delta \hat{\mathbf{E}}(\mathbf{r}) = k^2 \hat{\mathbf{E}}(\mathbf{r}),$$

$$k = \sqrt{\frac{i\omega 4\pi\sigma\mu}{c^2}} = \pm \frac{1+i}{\delta}, \quad \delta = \frac{c}{\sqrt{2\pi\sigma\mu\omega}}.$$

Рассмотрим проникновение гармонического поля вглубь проводника. Пусть проводником занято бесконечное полупространство, электрическое поле направлено параллельно плоскости поверхности проводника и во всех точках на его поверхности равно

$$\mathbf{E}(\mathbf{r}, t)|_{z=0} = \mathbf{E}_0 e^{i\omega t}.$$

Направим ось z перпендикулярно плоскости поверхности полупроводника, ее начало совместим с этой поверхностью. Тогда в силу симметрии задачи

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial y} = 0, \quad \Delta = \frac{d^2}{dz^2}.$$

Тогда каждая компонента комплексной амплитуды электрического поля должна в проводнике удовлетворять уравнению

$$\frac{d^2 \hat{E}_i}{dz^2} = k^2 \hat{E}_i.$$

Сразу можно записать решение этого уравнения:

$$\hat{E}_i(z) = E_{i1} e^{+kz} + E_{i2} e^{-kz}.$$

Первый член полученного решения описывает поле, экспоненциально возрастающее вглубь проводника. Очевидно, такое решение не имеет физического смысла, и необходимо принять $E_{i1} = 0$.

Направим другую ось (скажем, x) вдоль вектора \mathbf{E}_0 . Тогда, исходя из граничных условий на поверхности проводника, компоненты поля равны:

$$\hat{E}_y(0) = \hat{E}_z(0) = 0, \quad \hat{E}_x(0) = E_0.$$

Отсюда окончательно получаем решение задачи:

$$\hat{\mathbf{E}}(z) = \mathbf{E}_0 e^{-z/\delta} e^{-iz/\delta},$$

$$\mathbf{E}(z, t) = \mathbf{E}_0 e^{-z/\delta} \cos(\omega t - z/\delta).$$

По мере проникновения в проводник амплитуда поля экспоненциально затухает, а само оно начинает все более отставать по фазе от поля вне проводника. Это физическое явление локализации переменного поля в поверхностном слое называется *скин-эффектом*, а характерная глубина этого слоя δ – *глубиной скин-слоя*. Например, при частоте 50 Гц глубина скин-слоя для меди составит 9 мм, а для алюминия – 12 мм.

Подобный эффект наблюдается и при рассмотрении задач теплопроводности, что неудивительно, поскольку эти явления описываются одинаковыми уравнениями. Например, вспомним, что колебания температуры в погребе гораздо меньше, чем на поверхности, а время наступления максимальной и минимальной температуры – смещено.

1.10 Поток энергии электромагнитного поля

На предыдущих курсах были получены выражения для плотности энергии электрического и магнитного поля:

$$w_E = \frac{\mathbf{E} \cdot \mathbf{D}}{8\pi} = \frac{\varepsilon E^2}{8\pi}, \quad w_M = \frac{\mathbf{B} \cdot \mathbf{H}}{8\pi} = \frac{\mu H^2}{8\pi}.$$

В случае электромагнитного поля, как видно из приведенных ранее уравнений Максвелла, изменяющееся электрическое поле порождает в окружающем пространстве изменяющееся магнитное поле, оно – опять порождает электрическое и т.д. Поля распространяются в пространстве, неся с собой энергию, соответственно, можно говорить о плотности потока энергии электромагнитного поля.

Рассмотрим работу, совершаемую полем над движущимися зарядами. Мощность этой работы, отнесенная к единице объема, равна

$$P = \mathbf{E} \cdot \mathbf{j}.$$

Плотность тока можно выразить из одного из уравнений Максвелла:

$$\mathbf{j} = \frac{c}{4\pi} \left(\text{rot } \mathbf{H} - \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} \right).$$

Подставляя эту плотность тока в выражение для мощности и отнимая нулевой член, получающийся из другого уравнения Максвелла

$$\frac{c}{4\pi} \mathbf{H} \cdot \left(\text{rot } \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} \right) = 0,$$

получаем при перегруппировке членов

$$P = \frac{c}{4\pi} (\mathbf{E} \cdot \text{rot } \mathbf{H} - \mathbf{H} \cdot \text{rot } \mathbf{E}) - \frac{1}{4\pi} \left(\mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} \right).$$

Вспоминая одну из формул векторного анализа

$$\text{div}(\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot \text{rot } \mathbf{A} - \mathbf{A} \cdot \text{rot } \mathbf{B}$$

и вводя обозначения

$$w = \frac{\mathbf{E} \cdot \mathbf{D}}{8\pi} + \frac{\mathbf{B} \cdot \mathbf{H}}{8\pi}, \quad \mathbf{S} = \frac{c}{4\pi} \mathbf{E} \times \mathbf{H},$$

окончательно приходим к формуле

$$\frac{\partial w}{\partial t} + \text{div } \mathbf{S} = -P.$$

Полученную формулу можно интерпретировать как локальный закон сохранения энергии электромагнитного поля. Действительно, запишем его в интегральной форме, для какого-либо объема V , окруженного замкнутой поверхностью Σ . Используя теорему Остроградского-Гаусса, получаем

$$\frac{\partial}{\partial t} \iiint w \, dV = - \iiint P \, dV - \oint \mathbf{S} \cdot d\mathbf{\Sigma}.$$

Величина w есть плотность энергии электромагнитного поля, интеграл от нее — энергия поля в объеме V , следовательно — левый член выражения есть скорость изменения энергии поля в объеме V . Первый член в правой части есть изменение энергии поля вследствие произведенной им работы над зарядами, в единицу времени. Второй член, очевидно должен быть равен количеству энергии поля, выходящей наружу через окружающую объем поверхность Σ , в единицу времени. Отсюда окончательно получаем, что \mathbf{S} есть плотность потока энергии электромагнитного поля; это выражение называют *вектором Пойнтинга*.

1.11 Электромагнитные волны

Перейдем к подробному изучению электромагнитных волн. Общее рассмотрение является весьма сложной проблемой, поэтому ограничимся частным случаем однородной непроводящей среды без зарядов и токов (или вакуумом):

- а. $\rho(\mathbf{r}) = 0, \mathbf{j}(\mathbf{r}) = 0$;
- б. $\varepsilon, \mu = \text{const.}$

Уравнения Максвелла и материальные уравнения среды в этом случае

$$\text{rot } \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \quad \text{div } \mathbf{B} = 0,$$

$$\text{rot } \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t}, \quad \text{div } \mathbf{D} = 0,$$

$$\mathbf{D} = \varepsilon \mathbf{E}, \quad \mathbf{B} = \mu \mathbf{H}.$$

Возьмем частную производную по времени от 1-го уравнения 2-ой пары, и, поменяв порядок дифференцирования (не забывайте, что ротор есть дифференциальная операция, действующая на пространственные координаты) и используя 1-го уравнение 1-ой пары, получим:

$$\frac{\partial}{\partial t} \text{rot } \mathbf{H} = \text{rot} \left(\frac{\partial \mathbf{H}}{\partial t} \right) = -\frac{c}{\mu} \text{rot rot } \mathbf{E} = \frac{\varepsilon}{c} \frac{\partial^2 \mathbf{E}}{\partial t^2}.$$

Вспоминая проведенную выше формулу ротора от ротора и 2-го уравнения 2-ой пары, окончательно получаем

$$\Delta \mathbf{E} - \frac{\varepsilon \mu}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0.$$

Аналогичное уравнение описывает и магнитное поле:

$$\Delta \mathbf{H} - \frac{\varepsilon \mu}{c^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} = 0.$$

Такие уравнения называют *волновыми уравнениями*. Задолго до создания теории Максвелла, подобное уравнение было получено и исследовано в механике и акустике. Впервые же это было сделано учеными Д'Аламбером и Эйлером при изучении упругих колебаний струны. На этом курсе мы также получали и решали скалярное одномерное волновое уравнение при изучении распространения волн в длинных линиях.

Если ввести оператор Д'Аламбера

$$\square = \Delta - \frac{1}{v^2} \frac{\partial^2}{\partial t^2},$$

то волновые уравнения для полей представимы в виде

$$\square \mathbf{E} = 0, \quad \square \mathbf{H} = 0, \quad v = \frac{c}{\sqrt{\epsilon\mu}}.$$

Плоские волны

Наиболее просто решение волнового уравнения выглядит в случае, если все величины зависят только от одной декартовой координаты, например, от z , а при изменении других координат не изменяются. Тогда

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial y} = 0, \quad \Delta = \frac{d^2}{dz^2},$$

а любая координата векторов \mathbf{E} и \mathbf{H} подчиняется уравнению

$$\frac{\partial^2 f}{\partial z^2} = \frac{1}{v^2} \frac{\partial^2 f}{\partial t^2}.$$

Как уже получали ранее, общим решением этого уравнения является функция

$$f(x, t) = f_1(z - vt) + f_2(z + vt),$$

где f_1 и f_2 – произвольные функции, f_1 описывает волну, двигающуюся вдоль оси z в положительном направлении, а f_2 – в отрицательном направлении.

Далее будем рассматривать только волны, распространяющиеся в положительном направлении. Для плоского случая запишем 1-е уравнения обеих пар системы уравнений Максвелла:

$$\begin{aligned} \text{rot} \mathbf{H} &= \begin{vmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \\ 0 & 0 & \frac{\partial}{\partial z} \\ H_x & H_y & H_z \end{vmatrix} = -\mathbf{e}_x \frac{\partial H_y}{\partial z} + \mathbf{e}_y \frac{\partial H_x}{\partial z} = \frac{\epsilon}{c} \left(\mathbf{e}_x \frac{\partial E_x}{\partial t} + \mathbf{e}_y \frac{\partial E_y}{\partial t} + \mathbf{e}_z \frac{\partial E_z}{\partial t} \right), \\ \text{rot} \mathbf{E} &= \begin{vmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \\ 0 & 0 & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix} = -\mathbf{e}_x \frac{\partial E_y}{\partial z} + \mathbf{e}_y \frac{\partial E_x}{\partial z} = -\frac{\mu}{c} \left(\mathbf{e}_x \frac{\partial H_x}{\partial t} + \mathbf{e}_y \frac{\partial H_y}{\partial t} + \mathbf{e}_z \frac{\partial H_z}{\partial t} \right). \end{aligned}$$

Множители при одинаковых ортах должны быть равны, что дает систему дифференциальных уравнений в частных производных. Запишем решение этой системы, причем все константы интегрирования, как тривиальные и не влияющие на распространение электромагнитных колебаний, положим равными нулю:

$$E_x = E_1(z - vt), \quad H_x = -\sqrt{\frac{\epsilon}{\mu}} E_2(z - vt),$$

$$E_y = E_2(z - vt), \quad H_y = \sqrt{\frac{\epsilon}{\mu}} E_1(z - vt),$$

$$E_z = 0, \quad H_z = 0.$$

Можно сделать следующие выводы:

1. В каждый момент времени в каждой точке пространства вектора **E** и **H** плоской электромагнитной волны перпендикулярны друг другу и перпендикулярны направлению распространения волны. Последнее обстоятельство позволяет рассматривать такую волну как поперечную;
2. Компоненты поля по разным поперечным направлениям различны и не зависят друг от друга. Эта характерная особенность поперечных волн называется *поляризацией*;
3. В каждый момент времени в каждой точке пространства плотности энергии электрического и магнитного полей одинаковы, поскольку $\varepsilon E^2/8\pi = \mu H^2/8\pi$.

Сферические волны

Если волна изотропно расширяется от точечного источника или сходится к нему, то в сферической системе координат с источником в центре компоненты полей, в силу симметрии задачи, не будут зависеть от угловых координат. Решение волнового уравнения следует искать в виде

$$f = f(r, t),$$

а само волновое уравнение в силу

$$\frac{\partial}{\partial \theta} = \frac{\partial}{\partial \varphi} = 0$$

преобразуется к виду

$$\frac{1}{r} \frac{\partial^2}{\partial r^2} (rf) = \frac{1}{v^2} \frac{\partial^2 f}{\partial t^2}.$$

Перенеся r в правую часть и занеся его под частную производную по времени, получаем уравнение, аналогичное рассмотренному для плоской волны:

$$\frac{\partial^2}{\partial r^2} (rf) = \frac{1}{v^2} \frac{\partial^2}{\partial t^2} (rf).$$

Его решение

$$f(r, t) = \frac{f_1(r - vt)}{r} + \frac{f_2(r + vt)}{r}.$$

Плоские монохроматические волны. Поляризация

Важнейшим частным случаем плоских волн является их гармоническая зависимость от времени:

$$E_{x,y} \sim \cos(k(z - vt) + \varphi_{x,y}) = \cos(\mathbf{k} \cdot \mathbf{r} - \omega t + \varphi_{x,y}).$$

Введенный вектор **k** называют *волновым вектором*, он направлен вдоль направления распространения волны. Модуль волнового вектора $k = |\mathbf{k}|$ называют *вол-*

новым числом, оно связано с частотой ω и пространственной длиной волны колебаний λ (расстоянием между ближайшими точками, колеблющимися в одинаковой фазе) следующими соотношениями:

$$\omega = kv, \quad k = \frac{2\pi}{\lambda}.$$

Пусть

$$E_x(z, t) = E_{x0} \cos(kz - \omega t),$$

$$E_y(z, t) = E_{y0} \cos(kz - \omega t + \varphi).$$

Из выражений для E_x и E_y можно исключить зависимость от времени и координаты; при этом получится следующее выражение:

$$\left(\frac{E_x}{E_{x0}}\right)^2 + \left(\frac{E_y}{E_{y0}}\right)^2 - 2 \frac{E_x}{E_{x0}} \frac{E_y}{E_{y0}} \cos \varphi = \sin^2 \varphi.$$

Это выражение в общем случае задает на плоскости координат E_x и E_y эллипс. Давайте зафиксируем какую-либо точку пространства с координатой z и будем измерять значение электрического поля в этой точке в разные моменты времени. На плоскости $E_x E_y$ вектор \mathbf{E} будет направлен из начала координат в какую-либо точку на этом эллипсе (см. рис. 29), при изменении времени конец вектора \mathbf{E} будет двигаться вдоль эллипса. Направление движения зависит от значения фазы φ : при φ от 0 до π движение происходит против часовой стрелки, при φ от π до 2π – по часовой стрелке. То же происходит и при фиксированном моменте времени при изменении точки наблюдения вдоль z .

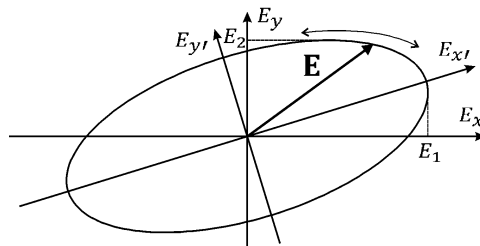


Рисунок 29. Изменение во времени вектора электрического поля монохроматической волны при фиксированном положении точки наблюдения.

Чтобы не рассматривать различные значения φ , сразу заметим, что уравнение любого эллипса можно преобразовать в канонический вид (не содержащий перекрестных членов) поворотом координат. Действительно, в нашем случае строго определено только направление оси z – вдоль направления распространения колебаний. Повернем оси x и y таким образом, чтобы компоненты полей удовлетворяли уравнению

$$\left(\frac{E_{x'}}{E_{x'0}}\right)^2 + \left(\frac{E_{y'}}{E_{y'0}}\right)^2 = 1,$$

амплитуды $E_{x'0}$ и $E_{y'0}$ будут зависеть от E_{x0} , E_{y0} и φ ; точное выражение этой зависимости можно найти в соответствующем курсе математики. Существенно

важно, что есть общий случай $E_{x'0} \neq E_{y'0}$; при этом говорят об *эллиптической поляризации* электромагнитной волны. Кроме того, выделяют два частных случая – *линейную поляризацию*, при которой одна из амплитуд $E_{x'0}$ или $E_{y'0}$ равна нулю и колебания происходят только вдоль одного поперечного направления (*плоскости поляризации*), и *круговую поляризацию*, когда $E_{x'0} = E_{y'0}$. Отметим, что для линейной поляризации разность фаз φ должна быть равна нулю или кратна π . Круговая поляризация имеет место при условии $\varphi = \pi/2 + n\pi$, $E_{x'0} = E_{y'0}$.

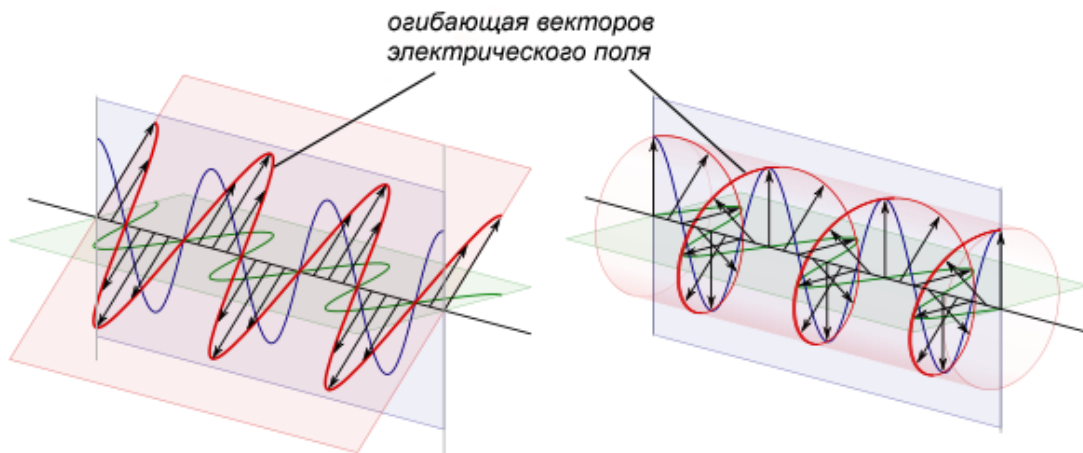


Рисунок 30. Линейная и круговая поляризация электромагнитной волны.

Сравнение линейной и круговой поляризаций электромагнитной волны показано на рисунке 30. Стрелками показаны вектора электрического поля в разных точках пространства вдоль линии распространения волны, красная линия – огибающая этих векторов. Кроме того, на рисунке показаны огибающие колебаний поля по двум поперечным направлениям. Видно, что для линейной поляризации фазы этих колебаний совпадают, для круговой – отстоят на $\pi/2$.

В дальнейшем нам пригодится тот факт, что любую эллиптически поляризованную волну можно представить как сумму двух линейно поляризованных волн с перпендикулярными плоскостями поляризации, или как сумму двух волн с круговой поляризацией, но с противоположными направлениями вращения вектора \mathbf{E} (левая и правая поляризации).

Неполяризованный свет

Как мы увидели, плоские строго монохроматические волны всегда поляризованы. С другой стороны, свет, создаваемый большинством источников (солнечный свет, свет от ламп накаливания и т.п.) поляризованным не является. Если удалось бы отследить пространственное поведение вектора \mathbf{E} в световом пучке, испущенном таким источником, то обнаружилось бы, что вместо вращения по эллипсу он хаотически меняет свою ориентацию. Это изменение происходит за время, много большее периода колебаний, но много меньшее, чем время разрешения любого регистрирующего прибора. Такой свет называется *неполяризованным*.

Причина отсутствия поляризации заключается в том, что обычный источник состоит из большого числа независимых друг от друга элементарных излучателей (атомов). Каждый отдельный атом испускает свет со своей поляризацией, но в результате сложения всех колебаний общий характер движения вектора \mathbf{E} приобретает хаотический характер. Более подробно эффекты, связанные с независимостью элементарных излучателей естественных источников света, обсуждаются в последующих главах.

Отметим, что если все атомы испускают свет с одинаковой поляризацией, излучение источника и в целом будет полностью поляризованным. Такая ситуация типична для лазеров, в которых атомы взаимодействуют друг с другом через поле излучения, или для света, пропущенного через специальные устройства – поляризаторы.

Стоячие волны

Рассмотрим суперпозицию двух монохроматических волн одинаковой частоты, распространяющихся навстречу друг другу. Будем считать, что они линейно поляризованы, их плоскости поляризации совпадают (пусть колебания происходят вдоль оси x), и они имеют одинаковую амплитуду.

$$E_{1x}(z, t) = E_0 \cos(kz - \omega t),$$

$$E_{2x}(z, t) = E_0 \cos(kz + \omega t + \varphi).$$

В силу принципа суперпозиции, итоговое поле есть сумма этих полей:

$$E_x = E_{1x} + E_{2x} = 2E_0 \cos\left(kz + \frac{\varphi}{2}\right) \cos\left(\omega t + \frac{\varphi}{2}\right).$$

Такая волна не является бегущей, в ней нет члена $kz \pm \omega t$. Напряженность во всех точках пространства изменяется с одинаковой частотой и в той же фазе; амплитуда колебаний меняется вдоль оси z по гармоническому закону. Такая волна называется *стоячей*. Мгновенные снимки бегущей и стоячей волн совпадают, но их динамика во времени различается кардинально.

Отметим, что при сложении двух волн, бегущих в одном и том же направлении, получается опять бегущая волна.

Часть II. Оптика

2.1 Отражение и преломление электромагнитных волн

Выше мы рассмотрели распространение плоской монохроматической волны в однородной среде. Напомним, что такую волну можно представить в виде двух линейно поляризованных волн:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_1 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t + \varphi_1) + \mathbf{E}_2 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t + \varphi_2),$$

где вектора \mathbf{E}_1 и \mathbf{E}_2 перпендикулярны друг другу и вектору \mathbf{k} , причем направление вектора \mathbf{E}_1 (и, вместе с ним, вектора \mathbf{E}_2) можно выбрать в плоскости, перпендикулярной \mathbf{k} , произвольно. Вектор \mathbf{H} каждой из волн всегда и везде перпендикулярен вектору \mathbf{E} своей волны.

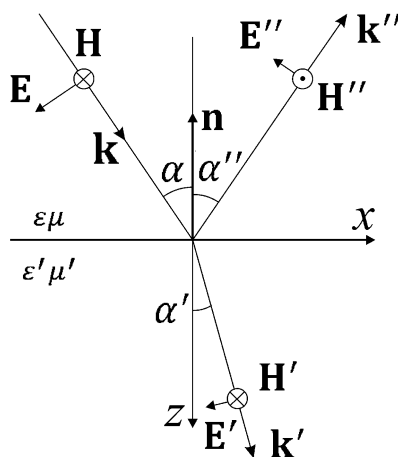


Рисунок 31. Падение электромагнитной волны на плоскую границу двух сред.

Рассмотрим поведение волны при ее падении на плоскую границу раздела двух оптически прозрачных сред с различными значениями ε и μ . Вы уже знаете, что часть волны при этом отразится от границы раздела, часть – пройдет из первой среды во вторую; эти волны изображены на рисунке 31. Давайте рассмотрим этот вопрос подробнее. Во-первых, выделим *плоскость падения* – плоскость, образованную векторами \mathbf{k} и \mathbf{n} , где \mathbf{n} есть единичный вектор нормали к границе раздела сред. Во-вторых, представим падающую волну как сумму двух линейно поляризованных волн; вектор \mathbf{E} одной из этих волн лежит в плоскости падения, вектор \mathbf{E} другой волны перпендикулярен этой плоскости. Для последующих рассуждений важно, что эти две волны можно рассматривать независимо друг от друга. Для первой волны вектор \mathbf{H} перпендикулярен плоскости падения, такую волну называют ТМ-волной (Transverse Magnetic), именно этот случай представлен на рисунке 31. Вторую волну называют ТЕ-волной (Transverse Electric). Так же поляризацию ТМ-волны иногда называют р-поляризацией, а поляризацию ТЕ-волны – s- поляризацией.

На предыдущих курсах при рассмотрении границы раздела двух диэлектриков были получены граничные условия, связывающие величины электрического и магнитного полей во одной среде (непосредственно перед границей) с соответствующими величинами во второй среде (непосредственно за границей). В рассматриваемом случае свободной электромагнитной волны плотность поверхностного заряда σ и плотность поверхностного тока \mathbf{i} равны нулю в каждой точке границы, и граничные условия приобретают следующий вид:

$$D_{1n} = D_{2n}, \quad \mathbf{E}_{1\tau} = \mathbf{E}_{2\tau}, \quad B_{1n} = B_{2n}, \quad \mathbf{H}_{1\tau} = \mathbf{H}_{2\tau}.$$

Будем обозначать величины, относящиеся к волне, прошедшей во вторую среду, с одним штрихом, а величины, относящиеся к отраженной волне – с двумя штрихами. В рассматриваемом случае ТМ-волны из граничного условия для вектора \mathbf{H} следует (см. рисунок 31)

$$H \cos(\mathbf{k} \cdot \mathbf{r} - \omega t) - H'' \cos(\mathbf{k}'' \cdot \mathbf{r} - \omega'' t) = H' \cos(\mathbf{k}' \cdot \mathbf{r} - \omega' t).$$

Это условие должно выполняться в каждый момент времени, что возможно лишь при выполнении условия

$$\omega = \omega' = \omega''.$$

Получаем очень важный вывод – частоты всех волн равны. Вспомним выражение для модуля волнового вектора: $k = \omega/v = \omega\sqrt{\epsilon\mu}/c$, отсюда

$$k'' = k,$$

$$k'/n' = k/n \Rightarrow \lambda'n' = \lambda n.$$

Здесь мы ввели величину $n = \sqrt{\epsilon\mu}$ – *показатель преломления среды*. Мы видим, что в оптически более плотной среде длина волны уменьшается.

Граничное условие должно выполняться в каждой точке границы раздела, в рассматриваемом случае – для каждого x и y при фиксированном z . Если скалярные произведения представить с следующим виде

$$\mathbf{k} \cdot \mathbf{r} = k_x x + k_y y + k_z z, \quad \mathbf{k}' \cdot \mathbf{r} = k'_x x + k'_y y + k'_z z, \quad \mathbf{k}'' \cdot \mathbf{r} = k''_x x + k''_y y + k''_z z,$$

то становится очевидным, что это возможно только при выполнении условий

$$k_y = k'_y = k''_y = 0,$$

$$k \sin(\alpha) = k' \sin(\alpha') = k'' \sin(\alpha'').$$

Второй важный вывод – все три волновых вектора лежат в плоскости падения. С учетом сказанного выше о соотношении волновых чисел получаем

$$\alpha = \alpha'', \quad n \sin(\alpha) = n' \sin(\alpha').$$

Угол отражения равен углу падения, угол преломления связан углом с падения зависимостью, которая носит название *закон Снеллиуса*.

Полностью аналогичные выводы для ТЕ-волны получаются из рассмотрения граничных условий для поля \mathbf{E} . Следовательно, полученные соотношения справедливы для падающей волны любой поляризации.

Формулы Френеля

Приведенные выше соотношения не дают ответа на вопрос о величине амплитуд отраженной и преломленной волн. Рассмотрим ТМ-волну (см. рис. 31). Из граничных условий для векторов **E** и **H** следует

$$E \cos(\alpha) + E'' \cos(\alpha'') = E' \cos(\alpha'), \quad H - H'' = H'.$$

Вспоминая соотношение между величиной электрического и магнитного поля в плоской волне, закон Сенллияуса, учитывая равенство углов падения и отражения, а так же то обстоятельство, что в большинстве сред $\mu \approx 1$, получаем:

$$E + E'' = \frac{\cos(\alpha')}{\cos(\alpha)} E', \quad E - E'' = \frac{\sin(\alpha)}{\sin(\alpha')} E'.$$

Решая эту систему алгебраических уравнений относительно неизвестных E' и E'' , находим:

$$E' = \frac{2 \sin(\alpha') \cos(\alpha)}{\sin(\alpha + \alpha') \cos(\alpha - \alpha')} E, \quad E'' = -\frac{\operatorname{tg}(\alpha - \alpha')}{\operatorname{tg}(\alpha + \alpha')} E.$$

Аналогично для ТЕ-волны используя те же граничные условия находим:

$$E' = \frac{2 \sin(\alpha') \cos(\alpha)}{\sin(\alpha + \alpha')} E, \quad E'' = -\frac{\sin(\alpha - \alpha')}{\sin(\alpha + \alpha')} E.$$

Полученные соотношения между амплитудами падающей, отраженной и преломленной волн называют *формулами Френеля*. Как мы видим, эти соотношения разные для разных поляризаций падающей волны. Из этих формул можно получить величину отношения мощности, уносимой отраженной волной, к мощности, переносимой к границе раздела сред падающей волной; это отношение называется *коэффициентом отражения*. Так как мощность пропорциональна квадрату амплитуды, получаем

$$R_{TM} = \frac{S'' \cos(\alpha'')}{S \cos(\alpha)} = \frac{S''}{S} = \frac{E''^2}{E^2} = \frac{\operatorname{tg}^2(\alpha - \alpha')}{\operatorname{tg}^2(\alpha + \alpha')}, \quad R_{TE} = \frac{\sin^2(\alpha - \alpha')}{\sin^2(\alpha + \alpha')}.$$

Аналогичным образом можно определить *коэффициент прохождения* T ; можно убедиться, что $T + R = 1$ для каждой из поляризаций, как и должно быть согласно закону сохранения энергии.

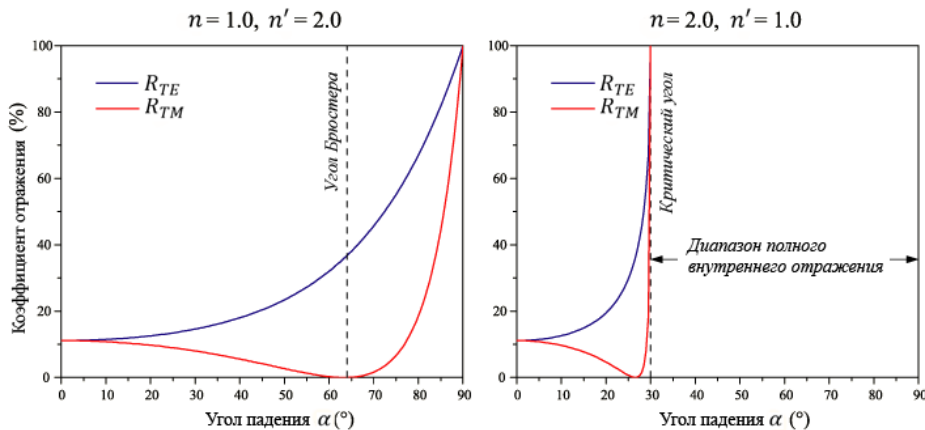


Рисунок 32. Зависимость коэффициента отражения от угла падения.

Рассмотрим зависимость коэффициентов отражения от угла падения при $n'/n > 1$ и $n'/n < 1$, соответствующие графики представлены на рисунке 32.

При нормальном падении ($\alpha = 0$) коэффициенты отражения одинаковы, что естественно, т.к. при таком падении стирается отличие между ТМ и ТЕ-волнами.

При увеличении угла падения появляется различие между этими коэффициентами – для всех углов падения, отличных от 0 и $\pi/2$, имеем $R_{TE} > R_{TM}$. Поэтому в общем случае при отражении и преломлении направление вектора линейной поляризации меняется, а круговая поляризация превращается в эллиптическую. Более того, прежде неполяризованный свет при отражении становится частично поляризованным (с преимущественно ТЕ-поляризацией).

Из рисунка 32 можно заметить, что существует угол падения, при котором $R_{TM} = 0$ – одна из поляризаций исчезает! Если на поверхность под этим углом падает волна, имеющая произвольную поляризацию, то отраженная волна будет поляризована линейно, вектор \mathbf{E} в ней будет перпендикулярен плоскости падения. Этот угол называют *углом Брюстера*; обозначим его через α_B . Условие, которому он должен удовлетворять

$$\operatorname{tg}(\alpha_B) = n'/n.$$

Полное внутреннее отражение

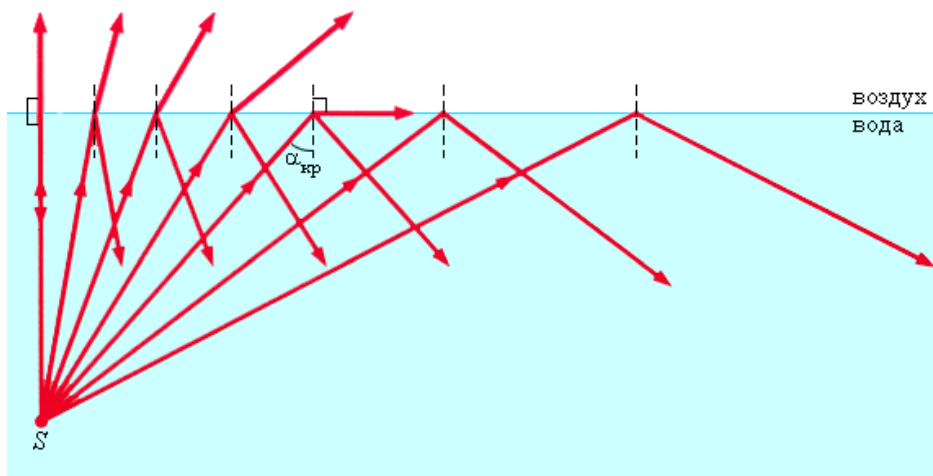


Рисунок 33. Поведение отраженного и преломленного лучей при различных углах падения.

Давайте рассмотрим случай $n'/n < 1$ – волна падает из оптически более плотной среды в менее плотную (например, из воды в воздух). При этом существует диапазон углов падения, для которых формальное применение закона Снеллиуса дает комплексный угол преломления. Действительно, если $\sin(\alpha) > n'/n$, согласно этому закону

$$\sin(\alpha') = \sin(\alpha) \frac{n}{n'} > 1.$$

Разумеется, все реальные углы, образуемые световыми лучами с осями координат, остаются вещественными, но меняется физическая сущность происходящих процессов – отсутствует уходящая от границы преломленная волна. Такой эффект называется *полным внутренним отражением*, т.е. при достаточно “косом”

падении волна не может вырваться из более плотной среды, вся световая энергия возвращается обратно в первую среду, коэффициент отражения $R = 1$. Критический угол падения, при котором наступает полное внутреннее отражение, определяется из условия

$$\sin(\alpha') = 1 \Rightarrow \sin(\alpha_{\text{кр}}) = n'/n.$$

Поведение отраженного и преломленного лучей при различных углах падения из оптически более плотной среды и момент возникновения полного внутреннего отражения изображено на рисунке 33.

Эффект полного внутреннего отражения находит многочисленные применения в оптических устройствах. Например, в таких устройствах как бинокли и перископы используются призмы, поворачивающие световой поток; это их действие основано на данном эффекте. Также полное отражение используется в оптических волокнах (световодах), представляющих собой тонкую нить из прозрачного вещества, по которой свет может распространяться на значительные расстояния без заметного затухания. Световоды широко применяются в оптических линиях связи, медицине и других областях.

Полное внутреннее отражение, однако, не означает, что поля во второй среде нет. Это поле может быть записано как

$$E'(x, z, t) = E e^{-z/l} \cos(k \sin(\alpha) x - \omega t), \quad l = \frac{n'}{k} \sqrt{\sin^2(\alpha) - \sin^2(\alpha_{\text{кр}})},$$

направление осей координат согласно рис. 31. Это выражение означает, что электромагнитное поле во второй среде представляет собой неоднородную волну, распространяющуюся параллельно границе раздела, с амплитудой, экспоненциально убывающей по мере удаления от границы. Величина l , по порядку близкая к длине волны, характеризует эффективную глубину проникновения поля во вторую среду. Таким образом, при полном внутреннем отражении электромагнитное поле во второй среде существует только в тонком приповерхностном слое.

Эффект полного внутреннего отражения может нарушаться, если во второй среде недалеко от границы поместить третью среду. Если две стеклянные призмы поместить близко друг от друга, то часть неоднородной приграничной волны попадает во вторую призму и будет в ней распространяться как обычная плоская волна в направлении от границы; коэффициент отражения первичной волны будет меньше единицы. Изменяя толщину воздушного зазора, можно менять соотношение мощностей отраженной и прошедшей волн.

Посмотрим на эффект нарушения полного внутреннего отражения со следующей стороны: слой воздуха между призмами играет роль непроницаемого потенциального барьера для квантов света. Однако, если толщина этого барьера сравнима с длиной волны, то какая-то доля световой энергии все-таки может проникнуть во вторую среду. В дальнейшем, при изучении квантовой механики, мы столкнемся с туннельным эффектом, аналогичным эффекту нарушения полного отражения.

Соотношения между фазами волн

Осталось разобраться с поведением фаз волн при отражении и преломлении. Можно заметить, что согласно формулам Френеля величина E'' может иметь знак, противоположный величине E , что означает противоположные направления векторов \mathbf{E} и \mathbf{E}'' для ТЕ-волны и противоположные направления этих векторов вдоль оси x для ТМ-волны. Поскольку

$$-\cos(\dots) = \cos(\dots - \pi),$$

это можно интерпретировать как сдвиг фазы волны на π при отражении. Об этом эффекте часто говорят как о потере полуволны при отражении. Соотношения между фазами падающей и отраженной волн различных поляризаций приведены в таблице 1.

Таблица 1. Соотношения между фазами падающей и отраженной волн.

	ТМ волна	ТЕ волна
$n' > n$	отличаются на π при $\alpha < \alpha_B$ совпадают при $\alpha > \alpha_B$	отличаются на π при любых углах падения
$n' < n$	совпадают при $\alpha < \alpha_B$ отличаются на π при $\alpha > \alpha_B$	совпадают при любых углах падения

Фаза преломленной волны всегда совпадает с фазой падающей волны независимо от поляризации, угла падения или от соотношения n и n' .

2.2 Геометрическая оптика

Давайте от плоских волн в однородных средах перейдем к более общему случаю неоднородных сред. Прямой путь – решение системы уравнений Максвелла с какими-то (как правило, заранее неизвестными) граничными и начальными условиями – крайне сложен. Есть, однако, приближение, которое значительно упрощает расчеты – это приближение *геометрической оптики*. Это приближение изучает принципы построения изображений в оптических системах и законы распространения света без учета его волновых свойств. Иначе говоря, в этом приближении длина волны принимается пренебрежимо малой:

$$\lambda \rightarrow 0.$$

В геометрической оптике основным понятием является луч – геометрическая линия, вдоль которой распространяется световая энергия. Основные свойства лучей: они независимы друг от друга, т.е. не взаимодействуют друг с другом, и в однородной среде распространяются прямолинейно (если ничто им не препятствует). Очевидно, что световые лучи – это идеализация, удобная для различных геометрических построений и расчетов. Давайте более подробно рассмотрим, как возникает такое понятие как световой луч и из чего следуют его свойства.

Мы знаем, что в однородной среде поле любой электромагнитной волны $\mathbf{E}(\mathbf{r}, t)$ описывается волновым уравнением

$$\Delta \mathbf{E} - \frac{\varepsilon \mu}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0.$$

В неоднородной среде это уравнение усложняется. Однако, если интересоваться только усредненным потоком энергии волны (что справедливо для видимого света), можно показать, что в приближении геометрической оптики оно остается верным для длины вектора \mathbf{E} :

$$\Delta E - \frac{1}{v^2} \frac{\partial^2 E}{\partial t^2} = 0, \quad v = \frac{c}{n(\mathbf{r})}.$$

Далее будем рассматривать монохроматические волны. Будем искать решение этого уравнения при помощи метода комплексных амплитуд

$$E(\mathbf{r}, t) = \text{Re}(\hat{E}(\mathbf{r}, t)), \quad \hat{E}(\mathbf{r}, t) = \hat{E}(\mathbf{r})e^{-i\omega t}.$$

Подставляя $\hat{E}(\mathbf{r}, t)$ в волновое уравнение, получаем, что комплексная амплитуда $\hat{E}(\mathbf{r})$ должна удовлетворять уравнению

$$\Delta \hat{E}(\mathbf{r}) + k^2 \hat{E}(\mathbf{r}) = 0,$$

называемому *уравнением Гельмгольца*. Если представить комплексную амплитуду в виде

$$\hat{E}(\mathbf{r}) = A(\mathbf{r})e^{ik_0\Phi(\mathbf{r})}, \quad k_0 = \frac{\omega}{c},$$

и подставить это выражение в уравнение Гельмгольца, то в левой части получим комплексное выражение. Оно равно нулю, только если и его действительная, и мнимая части равны нулю, т.е. получаем два уравнения, в которых входят функции $A(\mathbf{r})$ и $\Phi(\mathbf{r})$ – амплитуда и фаза волны в различных точках пространства. Эти уравнения можно значительно упростить, если использовать приближение геометрической оптики, при этом получаются следующие уравнения:

$$(\text{grad } \Phi)^2 = n^2,$$

$$A\Delta\Phi + 2 \text{grad } A \cdot \text{grad } \Phi = 0.$$

При их получении было использовано, что амплитуда волны и ее первые пространственные производные мало изменяются на протяжении длины волны – это и есть более строгое условие применимости геометрической оптики. Отступления от геометрической оптики возникают, например, в следующих случаях: 1) на границе геометрической тени; 2) вблизи фокуса, т.е. геометрической точки схождения лучей; 3) при распространении света в среде с резко меняющимся показателем преломления (например, в мутной среде); 4) при распространении света в сильно поглощающих средах (например, металлах).

Фазовая функция комплексной амплитуды $\Phi(\mathbf{r})$ имеет собственное название – *эйконал*, а уравнение, которым она описывается, называется *уравнением эйконала*, оно является основным уравнением геометрической оптики. Условие

$$\Phi(\mathbf{r}) = \text{const}$$

определяет *волновые поверхности* или *волновые фронты* – поверхности, колебания электрического поля в точках которой происходят синфазно. Если вспомнить, что градиент функции всегда ортогонален поверхности постоянного уровня этой функции, то уравнение эйконала можно записать следующим образом:

$$\text{grad } \Phi = n\mathbf{s},$$

где \mathbf{s} – единичный вектор нормали к фронту волны, направленный в сторону ее распространения.

С течением времени волновые фронты монохроматической волны распространяются в пространстве, переходя друг в друга. Это распространение определяется из условия постоянства полной фазы колебаний $k_0\Phi(\mathbf{r}) - \omega t = \text{const}$. Дифференцируя это условие, находим

$$\omega dt = k_0 d\Phi = \frac{\omega}{c} |\text{grad } \Phi| ds = \frac{\omega}{c} n ds.$$

Мы получили, что в геометрической оптике скорость распространения волнового фронта равна

$$\frac{ds}{dt} = v,$$

т.е. эта скорость такая же, как у плоской волны; кроме того, эта скорость не зависит от распределения амплитуды поля.

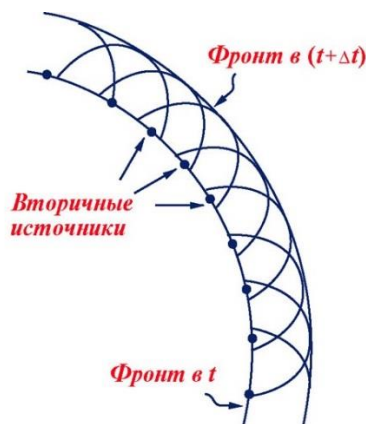


Рисунок 34. Принцип Гюйгенса.

Полученный результат позволяет построить волновой фронт в момент времени $t + dt$, если известно его положение в момент t . Для этого из каждой точки исходного волнового фронта следует отложить в направлении нормали отрезок длиной $v dt$. Соединив концы всех таких отрезков, мы и получим новое положение волнового фронта. Вместо этого можно из каждой точки волнового фронта, как из центра, описать сферы радиусом $v dt$. Огибающая таких сфер так же будет волновым фронтом в момент времени $t + dt$ – оба построения совершенно эквивалентны. Таким образом, точки на поверхности первичного волнового фронта можно рассматривать как источники вторичных сферических волн. Такой метод

нахождения распространения волнового фронта называется *принципом Гюйгенса*, его иллюстрация приведена на рисунке 34.

Воспользовавшись полученными результатами, запишем второе уравнение геометрической оптики в виде

$$A\Delta\Phi + 2n\frac{\partial A}{\partial s} = 0.$$

Определим *световой луч* как ортогональную траекторию к семейству волновых фронтов. Взятие производной по s можно понимать в смысле дифференцирования по длине луча, тогда это уравнение можно проинтегрировать вдоль луча и получить следующий результат:

$$A(l) = A_0 \exp\left(-\int_0^l \frac{\Delta\Phi}{2n} dl\right),$$

где A_0 – амплитуда в точке луча, от которой отсчитывается длина l . Эта формула показывает, что для определения волнового поля во всех точках луча достаточно знать его значение в какой-либо одной точке этого луча. Следовательно, уравнения геометрической оптики ничего не могут сказать относительно изменения амплитуды поля при переходе от одного луча к соседнему. Они допускают любые изменения амплитуды от луча к лучу, необходима только достаточная медленность такого изменения.

Таким образом, в приближении геометрической оптики и траектория всякого луча, и световое поле на луче совершенно не зависят от других лучей. Отсюда следует основное представление геометрической оптики о том, что энергия света распространяется вдоль независимых лучей, точнее – вдоль “лучевых трубок”, образованных лучами. Можно выделить отдельную световую трубку, поставив на пути распространяющейся волны узкую диафрагму. Только диафрагма не должна быть особенно узкой, а световая трубка слишком длинной. Дело в том, что на краях диафрагмы и вблизи боковых границ трубки амплитуда поля меняется резко, т.е. условия применимости геометрической оптики не выполняются. Возникает дифракция света, приводящая к уширению светового пучка.

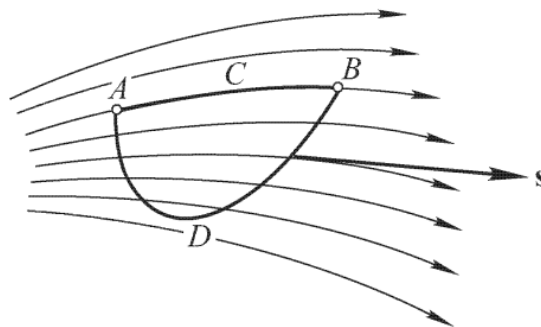


Рисунок 35. К доказательству принципа Ферма.

В однородной среде траектория светового луча, как хорошо известно, представляет собой прямую линию. В неоднородной среде эту траекторию можно

найти исходя из *принципа Ферма*. Его простейшая форма: луч света распространяется в пространстве между двумя точками по тому пути, по которому время его прохождения меньше, чем по любому из других соединяющих эти точки путей. Для доказательства этого принципа рассмотрим две лежащие на одном и том же луче точки A и B , и соединим их двумя линиями, одна из которых совпадает с лучом, другая – отличается от него – см. рисунок 35. В силу свойств градиента интеграл

$$\frac{1}{c} \int_A^B n(l) \mathbf{s} \cdot d\mathbf{l} = \frac{1}{c} \int_A^B \text{grad } \Phi \cdot d\mathbf{l} = \frac{1}{c} (\Phi(B) - \Phi(A))$$

одинаков для любого пути, соединяющего эти две точки. На линии ACB векторы \mathbf{s} и $d\mathbf{l}$ сонаправлены, поэтому $\mathbf{s} \cdot d\mathbf{l} = dl$, на линии ADB $\mathbf{s} \cdot d\mathbf{l} = dl \cos(\mathbf{s}, d\mathbf{l}) < dl$, поэтому

$$t_{ACB} = \int_{ACB} dt = \int_{ACB} \frac{dl}{v(l)} = \frac{1}{c} \int_{ACB} n(l) dl < t_{ADB} = \frac{1}{c} \int_{ADB} n(l) dl.$$

Величина $\Phi(B) - \Phi(A)$ имеет размерность длины и называется *оптической длиной пути*. Видно, что это есть расстояние, на которое свет распространился бы в вакууме за время его прохождения от A до B .

Существует более строгая формулировка принципа Ферма: любое малое изменение пути не приводит в первом порядке к изменению времени прохождения. Математическая запись этой формулировки

$$\delta t = 0.$$

2.3 Интерференция

Для понимания дальнейшего материала необходимо уяснить следующее: частоты электромагнитных волн оптического диапазона составляют порядка 10^{14} Гц, так что зарегистрировать мгновенные значения электрического или магнитного полей таких волн нельзя. Любой измерительный прибор – будь то глаз, фотопленка или ячейка ПЗС матрицы – выдает сигнал, зависящий лишь от усредненной по времени энергии света. Вводят *интенсивность электромагнитного излучения* – усредненную по времени величину плотности потока мощности, переносимую электромагнитной волной в направлении своего распространения:

$$I(\mathbf{r}, t) = \langle S(\mathbf{r}, t) \rangle = \frac{1}{\tau_0} \int_t^{t+\tau_0} S(\mathbf{r}, t') dt',$$

где τ_0 – разрешающее время измерительного прибора, которое практически всегда много больше периода электромагнитных колебаний, S – величина вектора Пойнтинга

$$\mathbf{S} = \frac{c}{4\pi} \mathbf{E} \times \mathbf{H}.$$

Проще говоря, чем больше интенсивность света в какой-либо точке, тем более яркой будет выглядеть эта точка на измерительном приборе. Для волн, чей средний поток мощности остается постоянным во времени, интенсивность будет зависеть только от положения точки наблюдения $I = I(\mathbf{r})$.

Для монохроматической линейно поляризованной волны с амплитудой напряжённости электрического поля E_0 интенсивность равна:

$$I = \frac{c}{8\pi} \sqrt{\varepsilon/\mu} E_0^2.$$

Дополнительный множитель $1/2$ появляется вследствие операции усреднения.

На предыдущих лекциях мы изучали структуру и динамику электромагнитной волны. А что если в пространстве распространяются две волны? Тогда, согласно принципу суперпозиции, результирующие поля \mathbf{E} и \mathbf{H} в каждой точке пространства есть векторная сумма соответствующих полей каждой из волн:

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2, \quad \mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2.$$

Посмотрим на поток энергии, описываемый вектором Пойнтинга

$$\mathbf{S} = \frac{c}{4\pi} \mathbf{E} \times \mathbf{H} = \mathbf{S}_1 + \mathbf{S}_2 + \mathbf{S}_{12}, \quad \mathbf{S}_{12} = \frac{c}{4\pi} (\mathbf{E}_1 \times \mathbf{H}_2 + \mathbf{E}_2 \times \mathbf{H}_1),$$

$$I = I_1 + I_2 + I_{12}.$$

Дополнительный член I_{12} называют *интерференционным членом*. Если он не равен нулю, то говорят, что волны интерферируют друг с другом (от английского to interfere – пересекаться, служить препятствием, помехой). Интенсивность суммарной волны оказывается отличной от суммы интенсивностей составляющих волн и, следовательно, принцип суперпозиции для интенсивности несправедлив. Дадим определение: физическое явление изменения (модуляции) средней плотности потока энергии, обусловленное суперпозицией волн, называется *интерференцией*.

В оптике интерференция света была объяснена английским физиком Юнгом на основе гипотезы о волновой природе света, что послужило несомненным подтверждением этой гипотезы.

Интерференция двух монохроматических волн

В большинстве задач, связанных с интерференцией и дифракцией света, исследуется пространственное положение максимумов и минимумов интенсивности света и их относительная величина, постоянные множители не учитываются. По этой причине часто полагают:

$$I(\mathbf{r}) = \langle E^2(\mathbf{r}, t) \rangle.$$

Рассмотрим сложение волн от двух источников, излучающих линейно поляризованные волны с частотой ω . Обозначим точку наблюдения как P , электрические поля от источников в этой точке можно записать как

$$\hat{\mathbf{E}}_1 = \mathbf{E}_{10}(P)e^{i\varphi_1(P)-i\omega t}, \quad \hat{\mathbf{E}}_2 = \mathbf{E}_{20}(P)e^{i\varphi_2(P)-i\omega t}.$$

$\mathbf{E}_{10}, \mathbf{E}_{20}$ – амплитуды, φ_1, φ_2 – фазы соответствующих волн в точке P . Суммарное поле

$$\hat{\mathbf{E}}(\mathbf{r}, t) = (\mathbf{E}_{10}(P)e^{i\varphi_1(P)} + \mathbf{E}_{20}(P)e^{i\varphi_2(P)})e^{-i\omega t} = \hat{\mathbf{E}}(P)e^{-i\omega t},$$

$\hat{\mathbf{E}}(P)$ – комплексная амплитуда поля в точке P .

Находим интенсивность

$$\begin{aligned} I = \langle E^2 \rangle &= \frac{1}{2} \langle |\hat{\mathbf{E}}(P)|^2 \rangle = \frac{1}{2} \langle \hat{\mathbf{E}} \hat{\mathbf{E}}^* \rangle = \\ &= \frac{1}{2} \langle (\mathbf{E}_{01}e^{i\varphi_1} + \mathbf{E}_{02}e^{i\varphi_2})(\mathbf{E}_{01}e^{-i\varphi_1} + \mathbf{E}_{02}e^{-i\varphi_2}) \rangle = \\ &= \frac{1}{2} \langle \mathbf{E}_{01}^2 \rangle + \frac{1}{2} \langle \mathbf{E}_{02}^2 \rangle + \frac{1}{2} \langle \mathbf{E}_{01} \cdot \mathbf{E}_{02} (e^{i(\varphi_1 - \varphi_2)} + e^{i(\varphi_2 - \varphi_1)}) \rangle. \end{aligned}$$

Видно, что при перпендикулярной поляризации $I_{12} = 0$, интерференция отсутствует. Отметим также, что интерференция также отсутствует, если частоты источников не равны друг другу. При параллельной поляризации выражение для интенсивности приобретает следующий вид:

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \langle \cos(\Delta\varphi) \rangle.$$

Разность фаз $\Delta\varphi(P) = \varphi_1(P) - \varphi_2(P)$ в точке наблюдения определяется как разностью фаз источников, так и тем, что свет до нее проходит от разных источников разное расстояние.

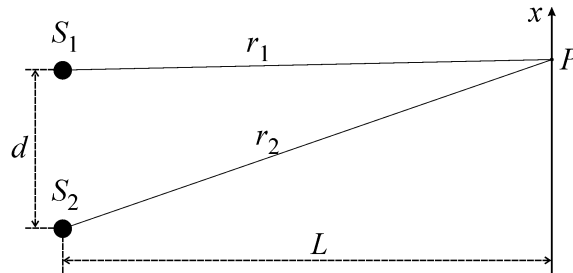


Рисунок 36. Освещение экрана двумя источниками.

Рассмотрим простейший случай двух полностью синфазных точечных источников S_1 и S_2 , излучающих сферические волны (см. рис. 36). Тогда $\Delta\varphi$ полностью определяется разностью расстояний

$$\Delta\varphi = k(r_1 - r_2) = \frac{2\pi}{\lambda} n(r_1 - r_2),$$

где r_1, r_2 – расстояния от источников до точки наблюдения, λ – длина волны соответствующей частоты в вакууме. Заметим, что величина $n(r_1 - r_2)$ есть разность оптических длин и называется *оптической разностью хода*. Так как $\Delta\varphi$ не зависит от времени, то усреднение по времени дает ту же самую величину. В зависимости от положения точки наблюдения изменяется величина $\Delta\varphi$, и значение $\cos(\Delta\varphi)$ может меняться от -1 до $+1$. Соответственно, в каких-то точках на экране интенсивность максимальна, в каких-то – минимальна:

$$I_{\max} = I_1 + I_2 + 2\sqrt{I_1 I_2} = (\sqrt{I_1} + \sqrt{I_2})^2 \quad \text{при } \Delta\varphi = 0, \pm 2\pi, \pm 4\pi, \dots$$

$$I_{min} = I_1 + I_2 - 2\sqrt{I_1 I_2} = (\sqrt{I_1} - \sqrt{I_2})^2 \quad \text{при } \Delta\varphi = \pm\pi, \pm3\pi, \dots$$

Число m , определяемое выражением

$$m(P) = \frac{\Delta\varphi(P)}{2\pi}$$

называют *порядком интерференционной полосы*. Условию $m(P) = \text{const}$ отвечает обычно какая-то линия (полоса) на экране, интенсивность света в разных точках на одной линии одинакова. Светлым полосам соответствуют целые порядки, темным – полуцелые.

Качество интерференционной картины (ее контрастность, мера видимости светлых и темных полос) определяется величиной, называемой *видностью*. Выражение для видности

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}.$$

В разных областях экрана эта величина может различаться, при ее вычислении I_{max} и I_{min} надо брать в непосредственной близости от точки наблюдения. Для рассматриваемого случая

$$V = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2} = \text{const};$$

в дальнейшем мы увидим, что на практике получается меньшая величина, чем дает это выражение.

Рассмотрим случай синфазных источников, дающих одинаковые интенсивности на экране $I_1 = I_2 = I_0$. В этом случае

$$I(P) = 2I_0(1 + \cos(\Delta\varphi(P))) = 4I_0 \cos^2\left(\frac{\Delta\varphi(P)}{2}\right), \quad V = 1.$$

Когерентность

В дальнейшем нами будет широко использоваться понятие когерентности. *Когерентность* это согласованность нескольких (многих) колебательных или волновых процессов во времени, проявляющаяся при их сложении, которое и обеспечивает ненулевую видность интерференционной картины. Источники, обеспечивающие такие колебания, называют когерентными источниками.

Мы рассмотрели случай полностью синфазных источников. Если у источников есть постоянная разность фаз, то

$$\Delta\varphi = (\varphi_{01} - \varphi_{02}) + \frac{2\pi}{\lambda} n(r_1 - r_2),$$

интерференционная картина в целом не изменится, а только сдвинется на экране. Заключаем, что любые два строго монохроматических источника являются когерентными.

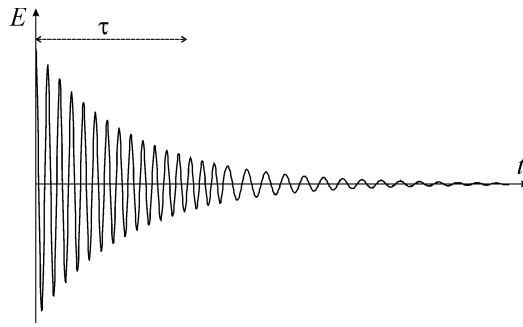


Рисунок 37. Схематичный вид цуга излучения отдельного атома.

Рассмотрим излучение реального источника света, состоящего из большого числа атомов — элементарных излучателей. Атомы могут переходить в возбужденное состояние за счет столкновений между собой или другими способами. При переходе из возбужденного состояния обратно в основное (или другое возбужденное) атомы излучают свет. Процесс излучения света отдельным атомом обычно занимает время $\tau \sim 10^{-8}$ с, за это время атом испускает цуг волн, схематичный вид которого показан на рисунке 37. В цуге излучения видимого диапазона содержится очень большое (порядка 10^7) число периодов колебаний. В самом простом случае колебания, продолжающиеся в течение этого интервала времени, можно рассматривать как гармонические; более точное описание картины излучения требует учета затухания колебаний, изменения частоты света из-за хаотического теплового движения атомов (доплеровский эффект), столкновения атомов в процессе излучения. После излучения атом через какой-то промежуток времени опять может перейти в возбужденное состояние, излучить свет и т.д.

Различные атомы обычных (не лазерных) источников начинают излучать свет независимо друг от друга в случайные моменты времени, т.е. фазы колебаний электромагнитного поля излучения различных атомов не скоррелированы друг с другом. Достаточно понятно, что волны, создаваемые обычными источниками, являются квазимонохроматическими и представляют собой наложения случайных последовательностей отдельных цугов волн. Также понятно, что если амплитуда такой волны будет оставаться более-менее постоянной, то фаза волны будет случайной функцией времени $\varphi(t)$. Эта функция такова, что за время τ она меняется слабо (такое время называется *временем когерентности*). С другой стороны, за время разрешения измерительного прибора τ_0 эта функция с равной вероятностью принимает любое значение из промежутка от $-\pi$ до π .

Сейчас становится очевидным, что два различных естественных источника света когерентными не являются. Действительно, поскольку фаза каждого источника случайным образом изменяется во времени, разность фаз источников также будет изменяться случайно. Положение максимумов и минимумов величины вектора Пойнтинга со временем перемещается по экрану; период этого движения много меньше времени разрешения регистрирующего прибора, так что он может зарегистрировать только какую-то равномерную среднюю освещенность. Математически это выражается тем, что

$$\langle \cos(\Delta\varphi(t)) \rangle = 0,$$

видность $V = 0$.

Все интерференционные схемы основаны на сложении ранее каким-либо способом разделенного излучения одного источника. При этом, несмотря на хаотическое изменение фазы первичной волны, разность фаз в расщепленных волнах остается постоянной – случайные изменения фазы световых колебаний в этих волнах протекают согласованно – и интерференционная картина может быть устойчивой.

В оптике существуют два основных способа разделения исходной волны:

1. деление волнового фронта, к которому относятся схема опыта Юнга, бипризма Френеля, зеркало Ллойда и т.д.;
2. деление амплитуды волны – кольца Ньютона, интерферометр Майкельсона, линии равной толщины / равного наклона, ...

Рассмотрим работу каждого способа на характерном примере.

Схема опыта Юнга

В схеме опыта Юнга два точечных синфазных источника получаются при освещении внешним точечным источником света экрана с двумя малыми отверстиями. Та часть волнового фронта этого внешнего источника, которая вырезается одним из этих отверстий, и будет служить первым из этих источников, другая часть волнового фронта, вырезанная другим отверстием, будет вторым источником. Если расстояния от внешнего источника до отверстий одинаково, то эти вторичные источники будут полностью синфазными. Вследствие малых размеров отверстий на них будет значительная дифракция, волны от них будут сильно расходиться, частично перекрываться на экране наблюдения, и в этой области и возможно наблюдение интерференционной картины.

Пусть экран освещается двумя точечными синфазными источниками одинаковой интенсивности, расстояние между которыми d много меньше расстояния от них до экрана L (см. рис. 35). Как мы уже знаем, распределение интенсивности на экране описывается выражением

$$I(P) = 4I_0 \cos^2 \left(\frac{1}{2} \frac{2\pi}{\lambda} n(r_1(P) - r_2(P)) \right).$$

Если x – координата точек на экране вдоль плоскости наблюдения, то формулы для расстояний можно упростить:

$$r_2 = \sqrt{L^2 + \left(x - \frac{d}{2}\right)^2} \approx L + \frac{1}{2} \frac{\left(x - \frac{d}{2}\right)^2}{L}, \quad r_1 = \sqrt{L^2 + \left(x + \frac{d}{2}\right)^2} \approx L + \frac{1}{2} \frac{\left(x + \frac{d}{2}\right)^2}{L},$$
$$r_1 - r_2 \approx \frac{xd}{L}.$$
$$I(x) = 4I_0 \cos^2 \left(\frac{\pi}{\lambda} \frac{xd}{L} n \right).$$

Вдоль плоскости наблюдения светлые полосы чередуются с темными, координаты светлых полос

$$x_m = \frac{mL\lambda}{nd}.$$

Расстояние между полосами равно длине волны в веществе, увеличенной в L/d раз, что делает их заметными даже без сильного увеличения.

Линии равного наклона

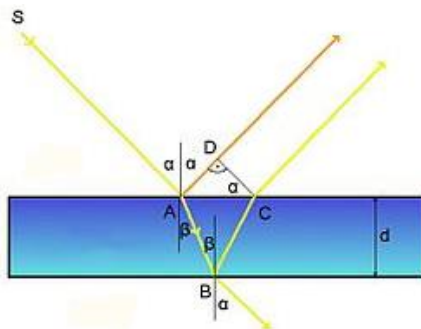


Рисунок 38. Преломление и отражение света на плоскопараллельной пластинке.

Направим луч света под углом на тонкую стеклянную пластинку постоянной толщины, как показано на рисунке 38. Мы изучали, что при падении на поверхность вещества волна частично отражается, частично преломляется, проходя внутрь вещества. Начальная энергия волны делится между этими двумя волнами. Пусть преломленная волна полностью отражается от нижней поверхности пластины, и, преломляясь второй раз, выходит из нее. Можно показать, что угол наклона выходящей волны совпадает с углом наклона отраженной волны – они параллельны, но разница пути вносит разность фаз между ними. Эта разность фаз определяется толщиной пластины, показателем преломления ее вещества и первоначальным углом падения волны. Так же свою лепту вносит сдвиг на π фазы второй волны при отражении. Если потом каким-либо образом свести эти волны вместе, то поскольку они порождены одной волной, возможно наблюдать устойчивую интерференционную картину. Для наблюдения этой картины экран должен находиться либо на достаточно большом расстоянии от пластины, либо он должен быть размещен в фокальной плоскости линзы, которая собирает отраженный от пластины свет.

Пусть на поверхность пластины падает рассеянный свет, в котором есть волны с различными углами падения. Соответственно, и на линзу лучи света будут падать под разными углами, линза же будет собирать этот свет в разные точки на экране в зависимости от угла падения. Поскольку разность хода отраженного и преломленного лучей зависит от угла падения, то и интенсивность на экране будет периодически меняться от точки к точке. Если лучи под разными углами падения падают со всех сторон, то на экране после собирающей линзы получим темные и светлые кольца. Получающиеся интерференционные полосы носят название полос равного наклона, т.к. они образованы светом, падающим под одним углом падения к нормали.

В данной схеме энергию одной волны мы разделили на две волны, т.е. как бы амплитуда исходной волны разделилась на две части – отсюда и название такого способа.

Пространственная когерентность

Рассмотрим эффекты, которые уменьшают видность интерференционной картины в приведенных выше схемах.

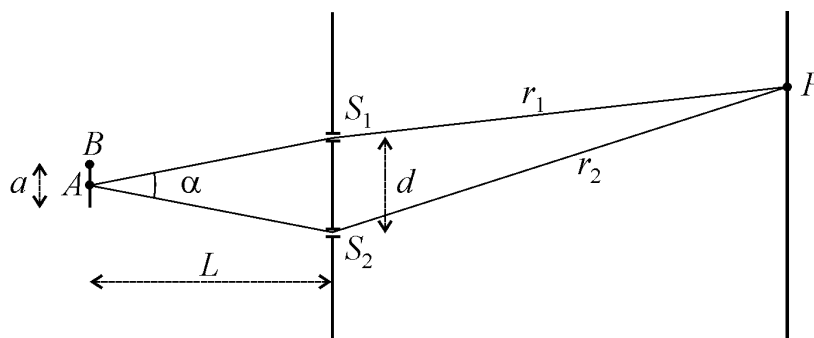


Рисунок 39. Освещение экрана со щелями протяженным источником.

Во-первых, рассмотрим влияние поперечного размера исходного источника света в опыте Юнга. Каждый малый элемент площади источника можно рассматривать как отдельный точечный источник, дающий свою интерференционную картину на экране. Поскольку для разных малых элементов площади расстояния до отверстий в экране будут различаться, вторичные волны уже не будут синфазными. Таким образом, картины интерференции от разных источников оказываются сдвинуты между собой. Поскольку отдельные источники не когерентны между собой, интенсивности от них на экране складываются, и общая картина получается размытой. Видность этой картины тем меньше, чем больше размер исходного источника.

Пусть точка A исходного источника находится на оптической оси системы, как показано на рисунке 39. Тогда расстояния AS_1 и AS_2 от нее до отверстий в экране, формирующих вторичные источники S_1 и S_2 , совпадают, эти источники являются синфазными. Пусть точка B отстоит от оптической оси, тогда расстояния BS_1 и BS_2 будут различаться, различаться будут и фазы вторичных источников, порожденных исходящей из B волной. Если оптическая разность хода $BS_2 - BS_1 = \lambda/2$, то максимумы интерференционной картины от B накладываются на минимумы картины от A и наоборот, общая картина размывается. Следовательно, условие получения относительно четкой картины

$$BS_2 - BS_1 \leq \lambda/2.$$

Если a – поперечный размер источника, d – расстояние между вторичными источниками, L – расстояние между исходным источником и экраном с отверстиями, то это условие можно представить в виде (при $a, d \ll L$):

$$\frac{ad}{L} \leq \lambda.$$

Можно определить максимально возможное расстояние d :

$$d \leq d_{\text{пк}} = \frac{\lambda L}{a} = \frac{\lambda}{\beta},$$

где $\beta = a/L$ – угловой размер источника. Длину $d_{\text{пк}}$ называют *длиной пространственной когерентности* (иногда – *поперечной длиной когерентности*), она определяет размер области волнового фронта, в котором излучение протяженного источника когерентно.

Временная когерентность

Второй эффект, уменьшающий видность картины интерференции, связан с конечной шириной $\Delta\omega$ спектра естественных источников света. Можно показать, что для таких источников ширина спектра и время когерентности τ связаны между собой соотношением неопределенности

$$\Delta\omega \cdot \tau \sim 2\pi.$$

Если отношение $\Delta\omega/\omega$ мало, то о излучении такого источника говорят как о *квазимонохроматическом излучении*.

Квазимонохроматическая волна может быть рассмотрена как сумма монохроматических волн с близкими значениями частот. Каждая монохроматическая волна создает на экране интерферометра свою картину полос. Для разных частот эти полосы смещены друг относительно друга, причем это смещение тем сильнее, чем больше порядок полосы интерференции m . Поскольку волны разной частоты не интерферируют друг с другом, то общая интенсивность света на экране есть сумма интенсивностей волн каждой их частот – полное распределение освещенности экрана определяется простым наложением картин, даваемых отдельными монохроматическими волнами. Отсюда легко понять, что при малых разностях хода интерферирующих волн (от нуля до несколько длин волн) положение полос в картинах, создаваемых отдельными монохроматическими составляющими, практически одинаково, и полосы наблюдаемой общей картины интерференции отчетливо видны. По мере увеличения разности хода из-за различия в длинах волн происходит смещение картин отдельных полос относительно друг друга; для больших значений m картина интерференции квазимонохроматической волны оказывается полностью размытой.

Пусть источник света является квазимонохроматическим, в его спектре присутствуют волны с длинами волн от λ до $\lambda + \Delta\lambda$. Найдем максимально наблюдаемый порядок интерференции. Интерференционная картина полностью исчезнет, когда максимум m волны $\lambda + \Delta\lambda$ совпадет с максимумом $m + 1$ волны λ . Так как это одна и та же точка, оптическая разность путей интерферирующих лучей должна быть одинакова для обеих этих волн. Выразим эту оптическую разность через m и λ :

$$m = \frac{\Delta\varphi(P)}{2\pi} = \frac{1}{\lambda} n(r_1 - r_2), \quad n(r_1 - r_2) = m\lambda.$$

Следовательно

$$(m + 1)\lambda = m(\lambda + \Delta\lambda).$$

Отсюда максимально наблюдаемый порядок интерференции равен

$$m_{max} = \frac{\lambda}{\Delta\lambda}.$$

Найдем максимальную оптическую разность хода, при которой наблюдается интерференционная картина для квазимонохроматического света:

$$n(r_1 - r_2)_{max} = m_{max}\lambda = \frac{\lambda^2}{\Delta\lambda}.$$

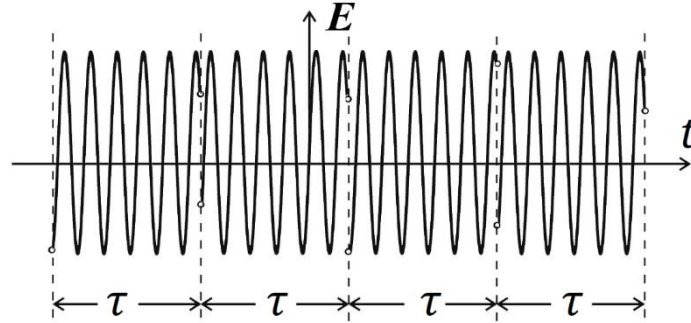


Рисунок 40. Простейшая модель волны, создаваемой естественным источником света.

Рассмотрим полученный результат с другой точки зрения. Выше мы описывали поле, создаваемое естественным источником света. Его простейшей идеализацией будет следующая модель: в течение времени когерентности τ поле представляет собой цуг волны с постоянной амплитудой и не сбивающейся фазой, при переходе от одного цуга к другому амплитуда остается постоянной, а фаза сбивается случайным образом – см. рисунок 40. Именно такой можно рассматривать волну первичного источника, такими будут и расщеплённые волны в начале своего пути. Однако, встречаясь в какой-либо точке на экране, эти волны пройдут разный оптический путь, соответственно время пути у них будет различаться. Очевидно, что если в точке наблюдения будут в основном перекрываться части одного и того же цуга, то они создадут хорошо видимую картину интерференции, если же в такой точке будут перекрываться различные цуги исходной волны, то интерференции не будет.

Отсюда следует, что для наблюдения интерференции оптическая разность хода не должна превышать длину одного цуга. Для нахождения этой длины рассмотрим цепочку равенств

$$\omega = \frac{2\pi}{T} = 2\pi \frac{v}{\lambda} = \frac{2\pi c}{\lambda n}, \quad \Delta\omega = 2\pi \frac{c}{n \lambda^2} \Delta\lambda, \quad \tau \approx \frac{2\pi}{\Delta\omega} = \frac{n \lambda^2}{c \Delta\lambda}.$$

Получаем длину цуга (точнее, длину, на которой фаза волны остается более-менее постоянной)

$$\tau v = \tau \frac{c}{n} = \frac{\lambda^2}{\Delta\lambda} = d_{BK}.$$

Мы видим, что этот результат согласуется с полученным ранее. Расстояние d_{BK} называют *длиной временной когерентности* (иногда – *продольной длиной когерентности*).

Рассмотрим три характерных примера. Для белого света, занимающего весь видимый диапазон от 400 нм до 800 нм, получаем среднюю длину волны $\lambda = 600$ нм, $\Delta\lambda = 400$ нм, и $d_{\text{вк}} \cong 900$ нм. Таким образом, длина когерентности белого света оказывается порядка длины световой волны. Достаточно монохроматический источник света с длиной волны $\lambda = 500$ нм имеет $\Delta\lambda = 10$ нм, длина когерентности равна 25 мкм. Для лазерного излучения схожей длины волны $\Delta\lambda$ составляет порядка 10^{-3} нм, а длина когерентности – 25 см.

2.4 Дифракция

Дифракция – явление отклонения волн от прямолинейного распространения, если только это не вызвано законами геометрической оптики (отражением, преломлением, рефракцией). Дифракция, в частности, приводит к огибанию световыми волнами препятствий и проникновению света в область геометрической тени. Как мы увидим далее, чем сильнее ограничены поперечные размеры (апертура) оптических элементов, тем сильнее проявляется дифракция. Как следствие, нет оптических приборов, полностью свободных от дифракции. В частности, именно дифракционные эффекты определяют разрешающую способность спектральных приборов.

Давайте разберем пример – пусть плоская монохроматическая волна падает перпендикулярно на экран, в котором вырезана длинная щель определенной ширины. Геометрическая оптика определяет, что за экраном лучи света, прошедшие через щель, будут распространяться прямолинейно, по нормали к плоскости экрана, без изменения своей интенсивности. За пределами вырезанной области света не будет, амплитуда поля равна нулю. Мы также определили, что условие применимости законов геометрической оптики есть малое изменение амплитуды поля в пространстве – очевидно, что здесь это условие нарушается. Более строго говоря, такое “четко обрезанное” поле не удовлетворяет волновому уравнению на своем краю.

Чтобы определить, как именно ведет себя поле за экраном, где приближение геометрической оптики работает, а где – нет, необходимо развить теорию дифракции, что и сделал в свое время французский физик Огюстен Жан Френель.

Принцип Гюйгенса-Френеля. Интеграл Кирхгофа

Строгое решение задачи дифракции – нахождение распределения поля электромагнитной волны после ее взаимодействия с препятствием – сводится к отысканию решения системы уравнений Максвелла с соответствующими начальными и граничными условиями. Это очень сложная как в формулировке, так и в решении задача. К счастью, многие важные задачи могут быть решены приближенным методом, в основе которого лежит *принцип Гюйгенса-Френеля*. Этот принцип состоит из двух положений:

1. Каждая малая по сравнению с длиной волны площадка поверхности волнового фронта может рассматриваться как источник вторичных сферических

волн. Фронт волны в последующие моменты времени представляет собой огибающую фронтов этих вторичных волн (см. рис. 34).

2. Вторичные волны интерферируют между собой, т.е. результирующее поле в любой точке пространства есть сумма полей, создаваемых этими вторичными источниками.

Немецкий физик Густав Кирхгоф дал математическое обоснование принципа Гюйгенса-Френеля и нашел те приближенные условия, когда он может считаться справедливым при решении задач дифракции. Этими условиями являются малость длины волны λ по сравнению с характерными размерами препятствия (как вариант, с характерным размером отверстия в непрозрачном экране), а так же малость λ по сравнению с расстоянием от препятствия до точки наблюдения.

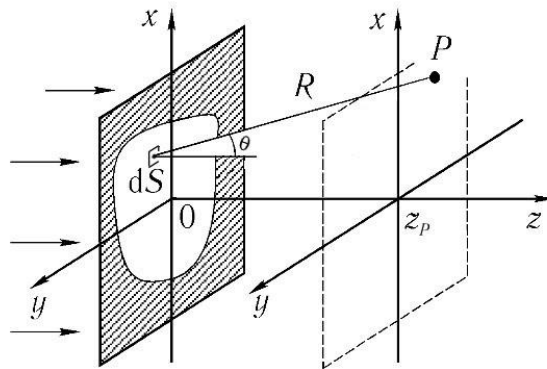


Рисунок 41. Дифракция света на отверстии.

Рассмотрим дифракцию волны на отверстии произвольной формы в непрозрачном экране (см. рис. 41). Это отверстие вырезает из волнового фронта падающей волны какую-либо поверхность. Пусть x, y, z – координаты точки на вырезанной поверхности волнового фронта, x_p, y_p, z_p – координаты точки наблюдения P , в которой вычисляем значение поля. Возьмем малый участок поверхности волнового фронта вокруг рассматриваемой точки площадью dS . Какое поле создаст этот участок в точке наблюдения? Комплексная амплитуда этого поля должна быть:

- а. пропорциональна комплексной амплитуде поля в рассматриваемой точке на поверхности волнового фронта $E(x, y, z)$;
- б. пропорциональна площади рассматриваемого участка dS ;
- в. как и в любой сферической волне, при удалении от источника комплексная амплитуда должна изменяться как e^{ikR}/R , где R – расстояние от рассматриваемой точки на поверхности волнового фронта до точки наблюдения;
- г. содержать какую-либо функцию угла θ между нормалью к поверхности волнового фронта и направлением на точку наблюдения; эта функция должна быть максимальна (равна единице) для $\theta = 0$ и убывать при увеличении θ .

Полная комплексная амплитуда поля в этом случае есть сумма полей, создаваемых всеми участками вырезаемой отверстием поверхности волнового фронта, т.е. эта сумма переходит в интеграл

$$E(P) = A \iint E(S) \frac{e^{ikR}}{R} K(\theta) dS.$$

Этот интеграл называется *интегралом Кирхгофа*. Константу A можно определить из условия, что при бесконечных размерах отверстия и постоянной амплитуде $E(S)$ поле в точке наблюдения $E(P) = E(S)$. Отсюда

$$A = \frac{1}{i\lambda}.$$

Кирхгоф показал, что если рассматриваемая поверхность есть волновой фронт точечного источника, то зависимость от угла θ имеет следующий вид:

$$K(\theta) = (1 + \cos(\theta))/2.$$

Сейчас мы рассмотрели случай, когда поверхностью, по которой производится интегрирование, является часть волнового фронта первичной волны. В качестве такой поверхности можно взять любую гладкую поверхность, закрывающую собой отверстие; фазы вторичных источников будут уже не одинаковы, но сами источники, конечно, останутся когерентными. Функция K в этом случае должна также содержать зависимость от угла между нормалью к поверхности и направлением распространения первичной волны.

Приближения интеграла Кирхгофа

Интеграл Кирхгофа может быть представлен в приближенном виде, существенно облегчающем вычисления, в двух важных частных случаях. В первом из этих случаев рассматривается следующая задача дифракции – точка наблюдения находится достаточно далеко от отверстия, волновой фронт падающей волны можно считать приблизительно плоским, совпадающим с плоскостью отверстия, а угол θ между нормалью к волновому фронту и направлением на точку наблюдения мал. Тогда соотношения между координатами можно записать как:

$$z \approx 0, \quad z_P \gg |x - x_P|, |y - y_P|,$$

$$R = \sqrt{(x - x_P)^2 + (y - y_P)^2 + (z - z_P)^2} \approx z_P + \frac{(x - x_P)^2 + (y - y_P)^2}{2z_P},$$

$$K(\theta) \approx 1.$$

Такое приближение называется *параксиальным*. При этом приближении интеграл Кирхгофа значительно упрощается и становится равным

$$E(P) = \frac{e^{ikz_P}}{i\lambda z_P} \iint E(x, y) \exp\left(ik \frac{(x - x_P)^2 + (y - y_P)^2}{2z_P}\right) dx dy,$$

где интегрирование идет по площади отверстия в экране. Дифракция, рассматриваемая в параксиальном приближении, называется *дифракцией Френеля*.

Второй частный случай – точка наблюдения лежит настолько далеко от отверстия в экране, что вектора, проведенные из различных точек отверстия к точке наблюдения можно считать параллельными. В этом случае

$$R = \sqrt{z_p^2 + x_p^2 + y_p^2 - 2(xx_p + yy_p) + x^2 + y^2} \approx R_0 - \frac{xx_p + yy_p}{R_0},$$

где $R_0 = \sqrt{z_p^2 + x_p^2 + y_p^2}$ – расстояние между точкой наблюдения и центральной точкой отверстия.

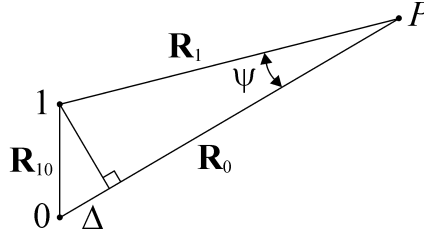


Рисунок 42. Нахождение разности расстояний.

Покажем геометрический смысл проделанного разложения. Пусть $\mathbf{R}_0 = x_p \mathbf{e}_x + y_p \mathbf{e}_y + z_p \mathbf{e}_z$ – вектор, направленный из центральной точки отверстия (точки 0) в точку наблюдения, а \mathbf{R}_1 направлен в эту точку из какой-либо другой точки отверстия (точки 1) – см. рисунок 42. Введем вектор $\mathbf{R}_{10} = x_1 \mathbf{e}_x + y_1 \mathbf{e}_y$, направленный из точки 0 в точку 1. Из рисунка видно, что

$$R_0 - \Delta = R_1 \cos(\psi), \quad \Delta = \mathbf{R}_{10} \cdot \frac{\mathbf{R}_0}{R_0} = \frac{x_1 x_p + y_1 y_p}{R_0}.$$

При удалении точки наблюдения от отверстия $\psi \rightarrow 0$, т.е. \mathbf{R}_0 и \mathbf{R}_1 становятся все более параллельны друг другу, а

$$R_1 \approx R_0 - \Delta.$$

Оценим, когда рассматриваемое разложение можно представить в таком виде. В этом разложении мы отбрасываем член $(x^2 + y^2)/2R$; необходимо, чтобы при этом интеграл не изменял сильно свое значение. Для выполнения этого приближения достаточно, чтобы

$$\exp\left(k \frac{x^2 + y^2}{2R_0}\right) \sim 1 \Rightarrow k \frac{D^2}{2R_0} < \frac{\pi}{4}, \quad D^2 = \max(x^2 + y^2) \Rightarrow R_0 > 4 \frac{D^2}{\lambda}.$$

Дифракция при выполнении этого условия называется *дифракцией Фраунгофера*. По-прежнему полагая угол θ малым, получаем в этом приближении

$$E(P) = \frac{e^{ikz_p}}{i\lambda z_p} \iint E(x, y) \exp\left(-ik \frac{xx_p + yy_p}{R_0}\right) dx dy.$$

При дифракции Фраунгофера волны, порожденные различными участками волнового фронта, идут параллельно друг другу, поэтому эту дифракцию называют также *дифракцией в параллельных лучах*. Таким образом, дифракция Фраунгофера определяется интерференцией плоских волн, синхронно излучаемых поверхностью волнового фронта в заданном направлении. Отсюда также следует вывод об устойчивом угловом распределении поля при данной дифракции – при увеличении расстояния от экрана до плоскости наблюдения дифракционная картина увеличивается подобно самой себе.

Качественное рассмотрение дифракции

Прежде чем начать анализировать решение задачи дифракции, определяемое интегралом Кирхгофа, разберемся с задачей нахождения суммы многих колебаний с различными фазами. Требуется найти

$$E(t) = \sum_i E_i(t) = \sum_i E_{0i} \cos(\omega t + \varphi_i).$$

Мы уже разбирали, что итоговое движение будет так же колебанием с той же частотой

$$E(t) = E_0 \cos(\omega t + \varphi_0),$$

а его амплитуду E_0 проще всего найти, если воспользоваться методом комплексных амплитуд:

$$\begin{aligned} \hat{E}(t) &= \sum_i \hat{E}_i(t) = \sum_i E_{0i}(t) e^{i(\omega t + \varphi_i)} = e^{i\omega t} \sum_i E_{0i} e^{i\varphi_i} = \hat{E}_0 e^{i\omega t}, \\ \hat{E}_0 &= \sum_i E_{0i} e^{i\varphi_i} = |\hat{E}_0| e^{i\varphi_0}, \quad E_0 = |\hat{E}_0|. \end{aligned}$$

Получили, что амплитуда E_0 суммы колебаний есть модуль суммы комплексных амплитуд отдельных колебаний. Каждое комплексное число можно представить как вектор на комплексной плоскости, а сумму таких чисел – как векторное сложение соответствующих векторов.

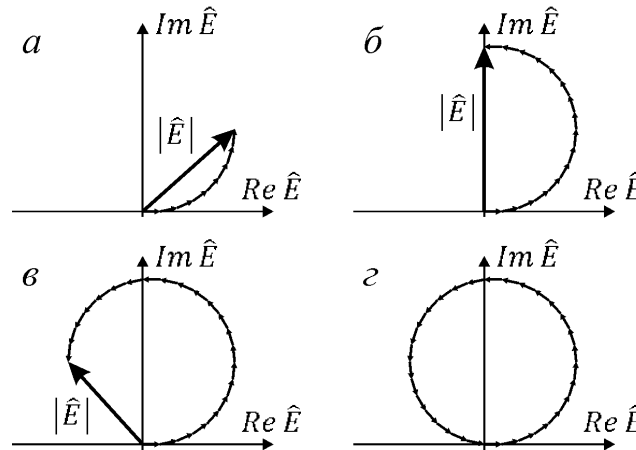


Рисунок 43. Нахождение амплитуды результирующего колебания при сложении многих колебаний. Максимальная разность фаз колебаний: а) $\pi/2$; б) π ; в) $3\pi/2$; г) 2π .

Пусть амплитуды отдельных колебаний одинаковы, их фазы линейно увеличиваются, а наименьшая из фаз φ_1 равна нулю. Как будет изменяться амплитуда итоговых колебаний при сложении все большего и большего числа колебаний? Как видно из рисунка 43, пока мы складываем колебания, фазы которых отличаются не больше чем на π , то итоговая амплитуда возрастает. В дальнейшем, при прибавлении колебаний, фазы которых отличаются от начальной на величину от π до 2π , эта амплитуда уменьшается. Она достигает нуля, если сум-

мируются колебания с фазами, разность между которыми равномерно распределена от 0 до 2π . При дальнейшем увеличении слагаемых колебаний картина повторяется, E_0 периодически изменяется от нуля до максимального значения.

Пусть теперь амплитуда отдельных колебаний слабо убывает с увеличением фазы. Тогда при достижении разности фаз в 2π суммарная амплитуда уже будет малой, но ненулевой – вследствие падения амплитуды колебания из области фаз от π до 2π не смогут полностью скомпенсировать колебания из области фаз от 0 до π (см. рис. 44). По мере дальнейшего возрастания числа слагаемых колебаний со все большими фазами и все меньшими амплитудами максимум итоговой амплитуды становится все меньше, а минимум – все больше, эта зависимость схематически показана на рисунке 45. В пределе очень многих колебаний итоговая амплитуда равна половине от ее максимального значения.

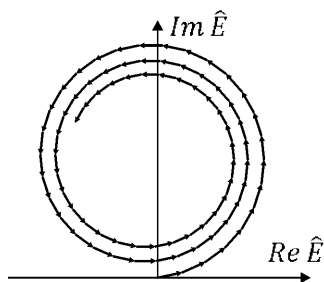


Рисунок 44. Сложение комплексных амплитуд затухающих колебаний.

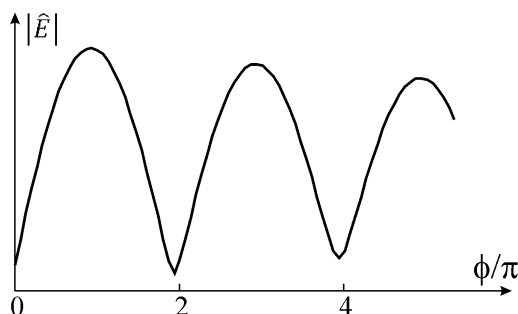


Рисунок 45. Зависимость амплитуды результирующего колебания от максимальной разности фаз складывающихся колебаний.

Возвращаясь обратно к задаче дифракции, становится очевидно, что интеграл Кирхгофа выражает подобную ситуацию – вторичные волны, приходящие в точку наблюдения, имеют различные фазы; требуется найти сумму этих волн. Давайте дадим определение: *зона Френеля* – область поверхности волнового фронта, фазы волн в точке наблюдения от различных участков которой различаются не более чем на π . Для этого разность расстояний от различных точек зоны Френеля до точки наблюдения не должна превышать $\lambda/2$. Из рассмотренного ранее следует, что волны от одной зоны Френеля, складываясь, усиливают друг друга, волны от двух соседних зон Френеля друг друга ослабляют.

Пусть характерный размер (радиус) отверстия в экране равен D . Расположим точку наблюдения на оси отверстия на расстоянии z_p , тогда разность максимального и минимального расстояний от разных точек экрана до точки наблюдения будет порядка

$$r_{max} - r_{min} \sim \sqrt{z_p^2 + D^2} - z_p \approx \frac{D^2}{2z_p}.$$

При сравнении этой разности с $\lambda/2$ появляется безразмерное число

$$F = \frac{D^2}{z_p \lambda},$$

которое характеризует число зон Френеля, открытых отверстием в экране; как мы видим, это число зависит от расстояния z_p . Если в точке наблюдения это число является четным, в этой точке будет наблюдаться минимум интенсивности света, если нечетным – максимум.

В зависимости от значения параметра Френеля можно пространство за экраном разбить на три области. Для случая $F \gg 1$ расстояние от экрана до плоскости наблюдения z_p должно быть

$$z_p \ll \frac{D^2}{\lambda}.$$

Это случай геометрической оптики. Открыто очень много зон Френеля, итоговая амплитуда поля в каждой точке плоскости наблюдения постоянна – не меняется при изменении размеров отверстия в экране или изменении расстояния z_p , пока выполняется соответствующее условие. Эта амплитуда равна амплитуде поля перед экраном (или равна нулю, если точка наблюдения лежит в области геометрической тени). Таким образом, в этом случае дифракционная картина не изменяется при изменении z_p .

Для случая $F \ll 1$ расстояние z_p должно быть

$$z_p \gg \frac{D^2}{\lambda},$$

а фазы вторичных волн от различных участков волнового фронта в точке на оси отверстия практически совпадают. Это есть условие дифракции Фраунгофера. Отверстие в экране составляет только малую часть зоны Френеля; говорят, что открыта только часть первой зоны Френеля. Дифракционная картина изменяется подобно самой себе в зависимости от z_p . Интенсивность поля в подобных точках дифракционной картины монотонно спадает по мере удаления точки наблюдения от экрана и монотонно возрастает при увеличении размера отверстия.

Случай $F \sim 1$ является промежуточным. При этом

$$z_p \sim \frac{D^2}{\lambda},$$

открыто несколько зон Френеля, дифракцию необходимо рассматривать в приближении Френеля. Интенсивность поля в точке наблюдения периодически меняется от минимальной до максимальной при изменении положения этой точки или изменении размеров отверстия (если при этом последовательно открываются и закрываются зоны Френеля). Таким образом, дифракционная картина существенным образом зависит от z_p .

Дифракция Френеля на щели

Рассмотрим падение плоской монохроматической волны на длинную прямоугольную щель. Направим координату x вдоль щели, тогда по этой координате пределы интегрирования будут приближенно равны от минус до плюс бесконечности. Если разделить подынтегральную функцию в интеграле Кирхгофа по переменным x и y , то по x можно сразу проинтегрировать:

$$\begin{aligned} E(P) &= \frac{e^{ikz_P}}{i\lambda z_P} E_0 \int_{-\infty}^{+\infty} \exp\left(ik \frac{(x - x_P)^2}{2z_P}\right) dx \int_{-d/2}^{d/2} \exp\left(ik \frac{(y - y_P)^2}{2z_P}\right) dy = \\ &= \frac{e^{ikz_P}}{\sqrt{i\lambda z_P}} E_0 \int_{-d/2}^{d/2} \exp\left(ik \frac{(y - y_P)^2}{2z_P}\right) dy. \end{aligned}$$

Введем переменную

$$v = \sqrt{\frac{k}{\pi z_P}} (y - y_P),$$

$$E(P) = \frac{e^{ikz_P}}{\sqrt{2i}} E_0 \int_{v_1}^{v_2} e^{i\frac{\pi v^2}{2}} dv = \frac{e^{ikz_P}}{\sqrt{2i}} E_0 (C(v) + iS(v)) \Big|_{v_1}^{v_2},$$

$$v_1 = \sqrt{\frac{k}{\pi z_P}} \left(-\frac{d}{2} - y_P\right), \quad v_2 = \sqrt{\frac{k}{\pi z_P}} \left(\frac{d}{2} - y_P\right).$$

Здесь введены так называемые *интегралы Френеля*:

$$C(v) = \int_0^v \cos\left(\frac{\pi \alpha^2}{2}\right) d\alpha, \quad S(v) = \int_0^v \sin\left(\frac{\pi \alpha^2}{2}\right) d\alpha.$$

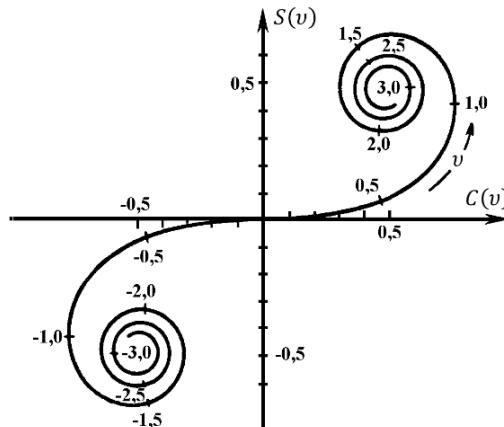


Рисунок 46. Спираль Корню.

Комплексная функция вещественного переменного $z(v) = C(v) + iS(v)$ описывает в параметрическом виде на комплексной плоскости кривую, изобра-

женную на рисунке 46 и называемую *спиралью Корню*. Итоговую амплитуду колебаний опять можем найти в виде модуля вектора на комплексной плоскости, соединяющего две точки $z(v_1)$ и $z(v_2)$ на этой спирали.

Найдем распределение интенсивности тени от края полубесконечного экрана. Совместим начало системы координат с краем экрана, в этом случае

$$v_1 = v_1(y_P) = -\sqrt{\frac{k}{\pi z_P}} y_P, \quad v_2 = \infty, \quad C(v_2) = S(v_2) = \frac{1}{2}.$$

$$E(y_P) = \frac{e^{ikz_P}}{\sqrt{2i}} E_0 \left[\left(\frac{1}{2} - C(v_1) \right) + i \left(\frac{1}{2} - S(v_1) \right) \right],$$

$$I(y_P) = \frac{E_0^2}{2} \left[\left(\frac{1}{2} - C(v_1) \right)^2 + \left(\frac{1}{2} - S(v_1) \right)^2 \right].$$

График распределения интенсивности на плоскости наблюдения, нормированной на интенсивность I_0 вдали от края экрана, представлен на рисунке 47. Осталось отметить, что величина $\sqrt{\lambda z_P}$, которая задает масштаб дифракционной картины, определяет ширину зон Френеля.

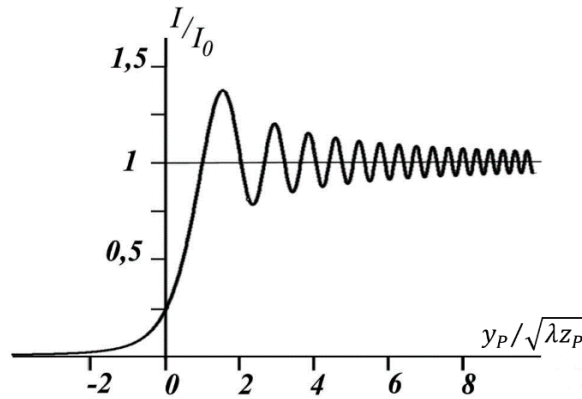


Рисунок 47. Распределение интенсивности при дифракции на крае полубесконечного экрана.

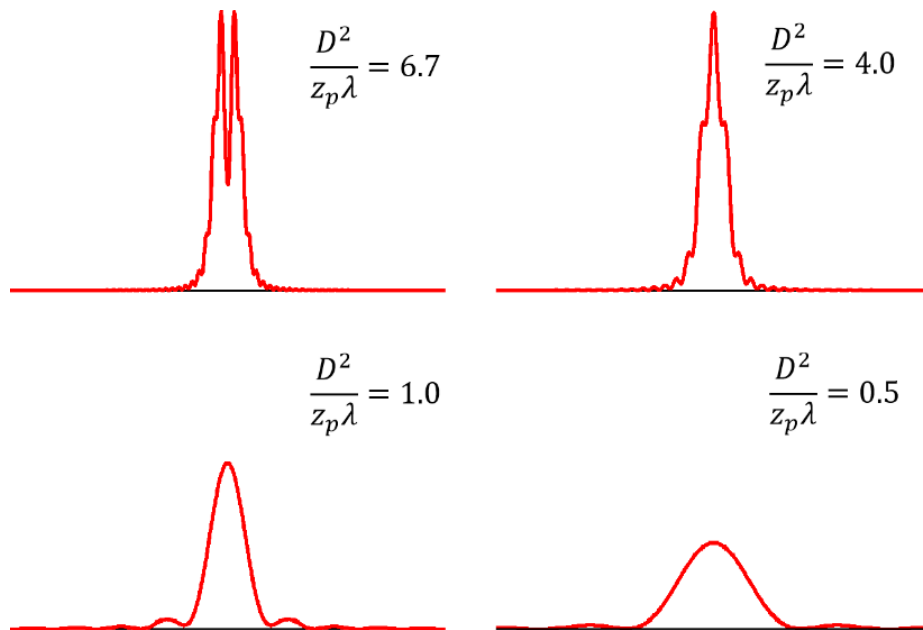


Рисунок 48. Переход от дифракции Френеля к дифракции Фраунгофера.

На примере дифракции на щели конечных размеров можно нагляднее всего показать переход от дифракции Френеля к дифракции Фраунгофера. Распределение интенсивности на плоскости наблюдения для разных значений параметра Френеля схематически изображено на рисунке 48. Видно, что при дифракции Френеля распределение интенсивности существенно меняется при изменении этого параметра, при дифракции же Фраунгофера форма распределения интенсивности не изменяется.

Дифракция Фраунгофера на прямоугольном отверстии

Рассмотрим, как преобразуется ранее полученный результат с ростом расстояния от экрана. Воспользуемся интегралом Кирхгофа в приближении Фраунгофера:

$$E(P) = \frac{E_0}{\lambda z_P} \int_{-D_x/2}^{D_x/2} \exp\left(-ik \frac{x x_P}{R_0}\right) dx \int_{-D_y/2}^{D_y/2} \exp\left(-ik \frac{y y_P}{R_0}\right) dy =$$

$$= \frac{E_0 D_x D_y}{\lambda z_P} \frac{\sin(\alpha_x)}{\alpha_x} \frac{\sin(\alpha_y)}{\alpha_y},$$

где опущены фазовые множители и введены обозначения

$$\alpha_x = \frac{k x_P D_x}{R_0}, \quad \alpha_y = \frac{k y_P D_y}{R_0}.$$

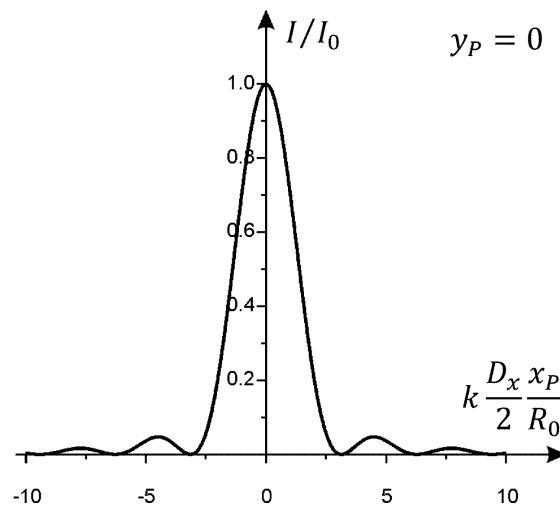


Рисунок 49. Распределение интенсивности на экране при дифракции Фраунгофера на прямоугольном отверстии.

Интенсивность можно представить в виде

$$I(P) = I_0 \frac{\sin^2(\alpha_x)}{\alpha_x^2} \frac{\sin^2(\alpha_y)}{\alpha_y^2},$$

I_0 – интенсивность в центре дифракционной картины. Распределение интенсивности на плоскости наблюдения как функция от x_P при фиксированном значении $y_P = 0$ показано на рисунке 49. Видно, что подавляющая доля световой энергии сосредоточена в центральной области, ограниченной по α от $-\pi$ до π . Можно

сказать, что на большом расстоянии от щели волна расширяется с характерным углом расходимости Θ , определяемым из условия

$$\alpha_x = k \frac{D_x}{2} \frac{x_P}{R_0} = \frac{2\pi D_x}{\lambda} \frac{x_P}{2} \sin(\Theta) \approx \pi \Rightarrow \Theta \approx \frac{\lambda}{D_x}.$$

Это соотношение верно не только щели, но и приближенно выполняется для любых форм отверстий в экране.

Все полученные результаты подытожены (качественно, без сохранения масштабов) на рисунке 50. На этом рисунке представлен переход от геометрической оптики к дифракции Френеля, а потом – к дифракции Фраунгофера, показан характерный угол расходимости пучка света далеко от отверстия.

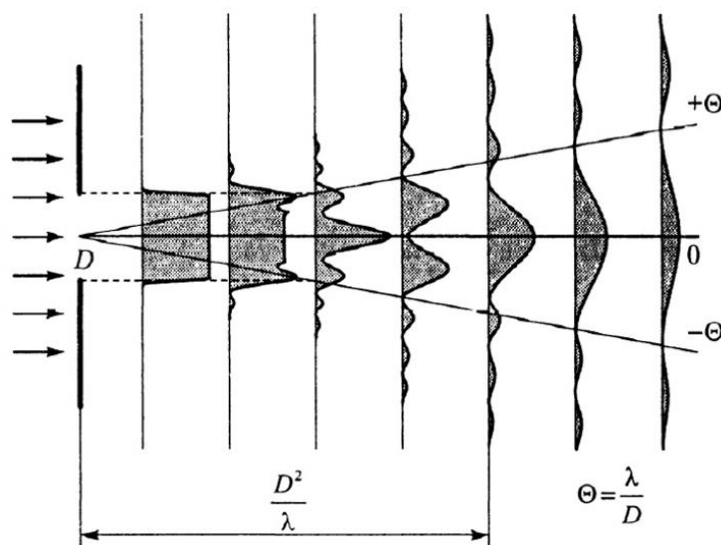


Рисунок 50. Распределение интенсивности света в пространстве при дифракции на прямоугольном отверстии.

2.5 Дисперсия света

Рассмотрение линейных систем во временной и частотной областях

Прежде чем мы перейдем к изучению такого явления как дисперсия света, давайте рассмотрим две простые электрические цепи, представленные на рисунке 51. Обе этих схемы представляют собой четырехполюсники, т.е. имеют входы и выходы. Пусть на входы этих схем подается сигнал $U_{вх}(t)$; какой сигнал будет на выходе каждой из схем?

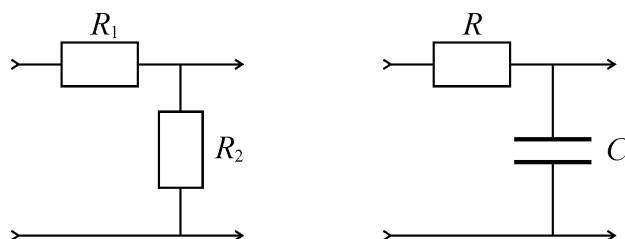


Рисунок 51. Линейные электрические схемы - четырехполюсники.

Первая из этих схем представляет собой простейший делитель из двух сопротивлений. Можно сразу написать

$$U_{\text{вых}}(t) = K \cdot U_{\text{вх}}(t), \quad K = R_2 / (R_1 + R_2) = \text{const.}$$

Для второй схемы – RC-цепи – напряжение на выходе будет определяться исходя из решения соответствующего дифференциального уравнения; можно показать, что

$$U_{\text{вых}}(t) = \int_0^{\infty} U_{\text{вх}}(t - \tau) \frac{\exp(-\tau/RC)}{RC} d\tau.$$

Обсудим полученные результаты – во-первых, видно, что обе этих схемы являются линейными – если входной сигнал увеличить, скажем, в два раза, то и выходной сигнал увеличится в два раза. Во-вторых, в отличие от первой схемы, в которой сигнал на выходе зависит от сигнала на входе только в тот же самый момент времени, во второй схеме сигнал на выходе зависит от сигнала на входе во все предшествующие моменты времени. Можно сказать, что функция

$$f(\tau) = \frac{\exp(-\tau/RC)}{RC}$$

есть некая функция “памяти” системы. В делителе из сопротивлений такой “памяти” нет, точнее говоря, для такой цепи функция $f(\tau)$ представляет собой дельта-функцию.

Результаты, которые мы получили, показывают, как выходной сигнал зависит от времени – говорят, что эти системы были рассмотрены во *временной области*. Как мы увидели, даже для простейшей RC-цепи $U_{\text{вых}}(t)$ представляет собой сложный интеграл. Можно существенно упростить анализ линейных систем, если перейти в *частотную область*. Пусть на вход линейной системы мы подаем гармонический сигнал

$$\hat{U}_{\text{вх}}(t) = U_{\text{вх},0} e^{i(\omega t + \varphi)}.$$

В этом случае, как мы уже знаем, выходной сигнал определяется по очень простой формуле:

$$\hat{U}_{\text{вых}}(t) = K \cdot \hat{U}_{\text{вх}}(t), \quad K = \frac{Z_{\text{вых}}}{\sum Z_i}.$$

Здесь $Z_{\text{вых}}$ – импеданс элемента, с которого снимается выходной сигнал. Для делителя из сопротивлений по-прежнему $K = R_2 / (R_1 + R_2) = \text{const}$; для RC-цепи

$$K = \frac{Z_C}{Z_R + Z_C} = \frac{1}{1 + i\omega RC} = K(\omega).$$

Мы видим, что для систем с памятью передаточная функция K является функцией частоты – это общее правило для линейных систем!

Дисперсия света в среде

Вспомним, что под действием электрического поля происходит смещение связанных зарядов в диэлектрике или поворот электрических диполей (если диэлектрик является полярным), и диэлектрическая среда *поляризуется* – каждый малый элемент среды с объемом dV начинает проявлять себя как электрический диполь с дипольным моментом

$$d\mathbf{P} = \mathbf{P}dV.$$

Вектор $\mathbf{P}(\mathbf{r})$ описывает плотность дипольных моментов среды, он называется вектором поляризации; этот вектор можно рассматривать как “ответ” среды на внешнее воздействие. Если постоянное электрическое поле \mathbf{E} , вызывающее поляризацию, не является очень сильным, то этот “ответ” линейно зависит от \mathbf{E} :

$$\mathbf{P} = \chi\mathbf{E}.$$

Коэффициент χ называется диэлектрической восприимчивостью, его величина, конечно же, зависит от среды. Вектор электрической индукции \mathbf{D} , по определению равный

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P},$$

в этом случае также будет линейно зависеть от \mathbf{E} :

$$\mathbf{D} = \mathbf{E} + 4\pi\chi\mathbf{E} = (1 + 4\pi\chi)\mathbf{E} = \varepsilon\mathbf{E}.$$

Коэффициент ε – это хорошо знакомая диэлектрическая проницаемость среды.

Приведенная выше линейная зависимость между \mathbf{P} и \mathbf{E} правильна для стационарных полей, сохраняет свой вид для медленно изменяющегося поля \mathbf{E} , а вот в случае его достаточно быстрого изменения она становится неправильной. Действительно, такая зависимость говорит, что ответ среды определяется полем \mathbf{E} только в тот же самый момент времени, в то время как диполи среды просто не будут успевать переориентироваться вслед за изменением \mathbf{E} ! Ответ среды останется линейным, но зависеть он будет от поля во все предыдущие моменты времени:

$$\mathbf{P}(\mathbf{r}, t) = \int_0^{\infty} \mathbf{E}(\mathbf{r}, t - \tau) f(\tau) d\tau.$$

Из всего предыдущего материала можно сделать очень важный вывод: диэлектрическая восприимчивость среды, как и ее диэлектрическая проницаемость, зависят от частоты переменного электрического поля, т.е. от частоты проходящей через среду электромагнитной волны:

$$\chi = \chi(\omega), \quad \varepsilon = \varepsilon(\omega).$$

Поскольку показатель преломления среды зависит от ε , то он также будет функцией частоты волны:

$$n = \sqrt{\varepsilon} = n(\omega).$$

Это явление – зависимость параметров среды от частоты электромагнитной волны – и называется *дисперсией*. Наиболее известна дисперсия света, при которой свет, имеющий разные длины волн (следовательно, разные частоты), преломляется на границе среды с дисперсией под разными углами. В частности, именно вследствие этого явления происходит разложение белого света при его прохождении через стеклянную призму.

Осталось обсудить следующий вопрос – как мы видели на примере электрических цепей, функция $K(\omega)$ является комплексной функцией. Ее модуль определяет отношение амплитуд выходного и входного сигналов, а фаза – сдвиг фаз между этими сигналами. Изучая дисперсию электромагнитных волн, мы также получаем комплексную величину n . Мы увидим, что ее вещественная часть будет определять скорость распространения волны в среде (и, следовательно, угол преломления на границе среды), а мнимая часть будет описывать затухание волны при ее прохождении через среду.

Электронная теория дисперсии Лоренца

Отыскание явного вида зависимости $\mathbf{P} = \mathbf{P}(\mathbf{E})$ сводится к решению уравнения, описывающего движение зарядов. Для этого рассмотрим атом как осцилляторную систему. Поскольку ядро атома обладает массой как минимум в 2000 раз большей, чем электроны, его можно считать неподвижным при наложении внешнего переменного поля, будем рассматривать движение только электронов.

Пусть в отсутствие внешнего поля молекулы или атомы среды нейтральны, появление же поля смещает электроны на расстояние \mathbf{r} , образуя в среде ненулевую плотность дипольных моментов

$$\mathbf{P} = -Ner,$$

где N – число атомов в единице объема вещества. Это смещение, конечно же, определяется полем \mathbf{E} :

$$\mathbf{r} = \mathbf{r}(\mathbf{E}).$$

Для нахождения этой зависимости воспользуемся уравнением движения Ньютона

$$m \frac{d^2 \mathbf{r}}{dt^2} = \sum_i \mathbf{F}_i.$$

Примем, что кроме внешней силы со стороны электрического поля $\mathbf{F} = -e\mathbf{E}$, на электрон действуют еще две внутриатомные силы:

1. Сила, удерживающая его в положении равновесия. Примем, что это квазиупругая сила $\mathbf{F} = -k\mathbf{r}$;
2. Сила, вызывающая затухание колебаний. Примем, что эта сила описывается аналогично силе трения $\mathbf{F} = -\eta\dot{\mathbf{r}}$.

Уравнение движения электрона приобретает вид

$$\ddot{\mathbf{r}} + \gamma\dot{\mathbf{r}} + \omega_0^2\mathbf{r} = -\frac{e}{m}\mathbf{E}.$$

Здесь $\gamma = \frac{\eta}{m}$ – параметр, описывающий затухание, $\omega_0^2 = \frac{k}{m}$ – квадрат собственной частоты осциллятора, в данном случае электрона в атоме. Будем рассматривать периодическую вынуждающую силу $\mathbf{E}(t) \sim e^{i\omega t}$. Частное решение уравнения в этом случае

$$\mathbf{r} = -\frac{e}{m} \frac{\mathbf{E}}{\omega_0^2 - \omega^2 + i\omega\gamma}.$$

Отсюда можно найти вектор поляризации среды

$$\mathbf{P} = \frac{Ne^2}{m} \frac{1}{\omega_0^2 - \omega^2 + i\omega\gamma} \mathbf{E}.$$

Следовательно, диэлектрическая восприимчивость среды и диэлектрическая проницаемость будут иметь вид

$$\chi_e = \chi_e(\omega) = \frac{Ne^2}{m} \frac{1}{\omega_0^2 - \omega^2 + i\omega\gamma}, \quad \varepsilon(\omega) = 1 + \frac{4\pi Ne^2}{m} \frac{1}{\omega_0^2 - \omega^2 + i\omega\gamma}.$$

Диэлектрическая проницаемость получилась комплексной, следовательно, показатель преломления среды и волновое число тоже будут комплексными величинами. Обозначим их действительную и мнимую части следующим образом:

$$n = n' + in'', \quad k = k' + ik'' = \frac{\omega}{c} (n' + in'').$$

Пусть плоская монохроматическая волна распространяется в среде вдоль оси z . Тогда поле волны можно представить в виде

$$E = E_0 \exp(i(kz - \omega t)) = E_0 \exp(-k''z) \exp(i(k'z - \omega t)).$$

Следовательно, мнимая часть волнового вектора определяет затухание волны в среде, а действительная – скорость ее распространения. Интенсивность света зависит от z как

$$I(z) = I_0 e^{-bz},$$

где $b = 2k'' = 2\frac{\omega}{c}n''$. Эта зависимость известна как *закон Бугера*, она определяет поглощение света при его распространении в среде. Показатель преломления и коэффициент поглощения выражаются через комплексную диэлектрическую проницаемость среды следующим образом:

$$n'(\omega) = \operatorname{Re} \left[\sqrt{\varepsilon(\omega)} \right], \quad b(\omega) = 2\frac{\omega}{c} \operatorname{Im} \left[\sqrt{\varepsilon(\omega)} \right].$$

Если ввести плазменную частоту как

$$\omega_p = \sqrt{\frac{4\pi Ne^2}{m}},$$

то диэлектрическая проницаемость запишется в виде

$$\varepsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 + i\omega\gamma} =$$

$$= 1 + \frac{\omega_p^2(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + \omega^2\gamma^2} - i \frac{\omega_p^2\omega\gamma}{(\omega_0^2 - \omega^2)^2 + \omega^2\gamma^2}.$$

Анализ этого выражения проведем, разделив среды на разреженные и плотные.

Разреженные среды

В разреженных средах плотность атомов в веществе N мала. Это позволяет, воспользовавшись известной формулой математического анализа

$$(1 + x)^\alpha \approx 1 + \alpha x \text{ при } x \ll 1,$$

представить показатель преломления и коэффициент поглощения диспергирующей среды в виде

$$n'(\omega) = 1 + \frac{1}{2} \frac{\omega_p^2(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + \omega^2\gamma^2}, \quad b(\omega) = \frac{\omega_p^2\omega^2\gamma/c}{(\omega_0^2 - \omega^2)^2 + \omega^2\gamma^2}.$$

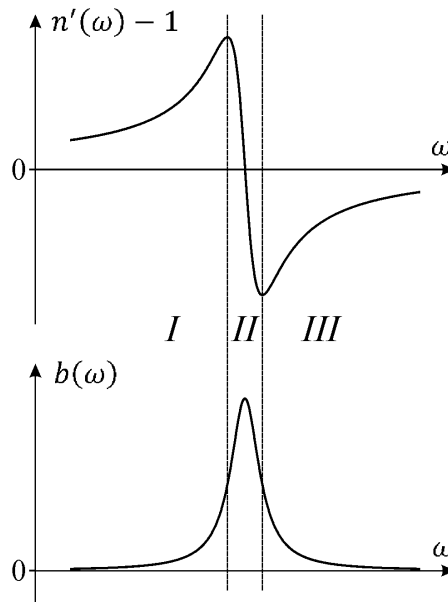


Рисунок 52. Зависимость показателя преломления и коэффициента поглощения от частоты.

Графики зависимости этих величин от частоты представлены на рисунке 52. В этих графиках наблюдается знакомая картина резонанса – вблизи собственной частоты системы значительно увеличивается поглощение волны средой. Разобьем весь диапазон частот на три части. В областях I и III, вдали от собственной частоты, коэффициент поглощения мал, показателя преломления слабо отличается от единицы, а его величина увеличивается с ростом частоты: $\frac{\partial n}{\partial \omega} > 0$ – это условие определяет *область нормальной дисперсии*. В области II $\frac{\partial n}{\partial \omega} < 0$ – это *область аномальной дисперсии*. Величина коэффициента поглощения в этой области уже значительна, поэтому соответствующий ей диапазон частот также называется *полосой поглощения*.

Если атомы и молекулы вещества имеют несколько резонансных частот, то будет несколько полос поглощения, в каждой из которых будет наблюдаться аномальная дисперсия.

Вдали от области поглощения, при $|\omega - \omega_0| \gg \gamma$, дисперсионная формула для $n'(\omega)$ приобретает вид

$$n'(\omega) = 1 + \frac{1}{2} \frac{\omega_p^2}{\omega_0^2 - \omega^2}.$$

Предполагая, что $\omega \ll \omega_0$, эту формулу можно разложить в ряд по степеням малого параметра ω/ω_0 . Переходя от частоты к длине волны, получаем *дисперсионную формулу Коши*:

$$n'(\omega) - 1 \approx \frac{1}{2} \frac{\omega_p^2}{\omega_0^2} \left(1 + \frac{\omega^2}{\omega_0^2} \right) = A \left(1 + \frac{B}{\lambda^2} \right),$$

которая хорошо описывает дисперсию газов в видимой и ИК-областях спектра.

Плотные среды

При рассмотрении дисперсии в разреженных средах мы полагали, что на атом воздействует только поле световой волны. Если же среда достаточно плотная, то расстояния между атомами невелики, и необходимо учитывать их взаимное влияние. Для области прозрачности, когда $|\omega - \omega_0| \gg \gamma$, справедливо следующее выражение для показателя преломления среды:

$$\frac{n^2 - 1}{n^2 + 1} = \frac{1}{3} \frac{\omega_p^2}{\omega_0^2 - \omega^2}.$$

Данное выражение называют *формулой Лоренца-Лоренца* в честь датского физика Людвиг Лоренца и голландского физика Хендрика Лоренца, получивших ее независимо друг от друга.

Если показатель преломления вещества близок к единице, то $n^2 + 1 \approx 2$ и формула Лоренца-Лоренца переходит в формулу Коши. Стоящая в правой части формулы Лоренца-Лоренца величина ω_p^2 пропорциональна концентрации атомов N , а следовательно, и плотности вещества ρ . Поэтому во многих случаях справедлив закон постоянной удельной рефракции:

$$\frac{1}{\rho} \frac{n^2 - 1}{n^2 + 1} = \text{const.}$$

Например, для воздуха при увеличении давления в 200 раз удельная рефракция изменяется на доли процента.

2.6 Фазовая и групповая скорости

Говоря о плоской монохроматической волне, мы записывали ее зависимость от координаты и времени как

$$f(z, t) = f_0 \cos(kx - \omega t).$$

Найдем скорость перемещения точки волны с определенной фазой (это может быть один из максимумов или минимумов волны). Эта скорость (ее называют *фазовой скоростью*) определяется из условия

$$kx - \omega t = \text{const}, \quad k dx - \omega dt = 0, \quad \frac{dx}{dt} = \frac{\omega}{k} = v_{\text{фаз}}.$$

Для сред с дисперсией мы видели, что в эту формулу должна входить реальная часть волнового числа, и получали зависимость волнового числа от частоты в виде

$$k'(\omega) = \frac{\omega}{c} \text{Re} \left[\sqrt{\varepsilon(\omega)} \right].$$

Можно обратить это уравнение и записать его в виде

$$\omega = \omega(k),$$

такое уравнение носит название *дисперсионного уравнения*.

Монохроматическая волна обладает бесконечной длительностью, следовательно, не несет в себе какой-либо информации. Как мы хорошо знаем, для передачи информации нужно осуществить модуляцию волны – амплитудную, частотную, фазовую, импульсную и т.д. Спектр получающегося сигнала при этом будет неизбежно шире, чем бесконечно узкий спектр монохроматического сигнала.

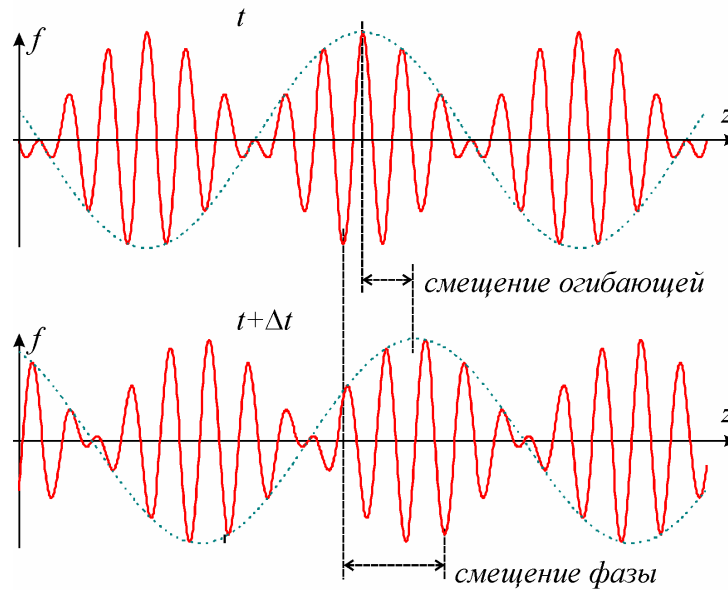


Рисунок 53. Биения в два последующих момента времени.

Рассмотрим представленный на рисунке 53 сигнал, спектр которого включает только две близко стоящие частоты:

$$\begin{aligned} f(z, t) &= f_0 \cos(k_1 x - \omega_1 t) + f_0 \cos(k_2 x - \omega_2 t) = \\ &= 2f_0 \cos\left(\frac{k_2 - k_1}{2} x - \frac{\omega_2 - \omega_1}{2} t\right) \cos\left(\frac{k_1 + k_2}{2} x - \frac{\omega_1 + \omega_2}{2} t\right). \end{aligned}$$

В результате наложения двух периодических сигналов с близкими частотами получаются биения. Второй множитель описывает волну с частотой, близкой к частотам двух исходных волн; ее фазовая скорость так же близка к их фазовым скоростям. Часто эту волну называют несущей волной. Первый множитель есть низкочастотная модуляция огибающей несущей волны, скорость перемещения

максимума этой огибающей так же определяется из уравнения постоянства ее фазы

$$\Delta k x - \Delta \omega t = \text{const}, \quad \frac{dx}{dt} = \frac{\omega_2 - \omega_1}{k_2 - k_1} \approx \frac{d\omega}{dk} \equiv v_{\text{гр}}.$$

Здесь мы ввели *групповую скорость* как скорость перемещения максимума амплитудной огибающей квазимонохроматического *волнового пакета* – совокупности обладающих разными частотами волн, которые описывают обладающее волновыми свойствами возмущение в среде. Это возмущение в общем случае ограничено во времени и пространстве. Следовательно, именно групповая скорость определяет скорость перемещения информации в среде. Значение групповой скорости определяется из дисперсионного уравнения. Можно показать, что групповая скорость определяется выражением (*формула Релея*)

$$v_{\text{гр}} = \frac{v_{\text{фаз}}}{1 + \frac{\omega}{n} \frac{\partial n}{\partial \omega}} = v_{\text{фаз}} - \lambda \frac{dv_{\text{фаз}}}{d\lambda}.$$

Если среда дисперсии не имеет, то групповая скорость совпадает с фазовой, в случае нормальной дисперсии $v_{\text{гр}} < v_{\text{фаз}}$, в случае аномальной дисперсии $v_{\text{гр}} > v_{\text{фаз}}$. На рисунке 53 изображена ситуация, при которой групповая скорость в два раза меньше фазовой скорости.

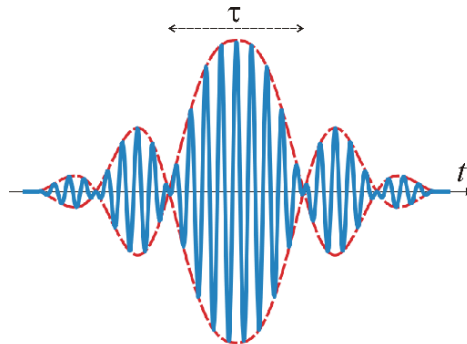


Рисунок 54. Ограниченный во времени волновой пакет.

Рассмотрим ситуацию, когда нам надо передать в среде с дисперсией импульс, занимающий достаточно короткий промежуток времени τ . Характерный вид подобного волнового пакета представлен на рисунке 54. Мы знаем, что ширина спектра $\Delta\omega$ такого сигнала достаточно широка, поскольку

$$\tau \Delta\omega \sim 2\pi.$$

Разные области широкого спектра могут обладать разной групповой скоростью, эта разность пропорциональна второй производной частоты $\frac{d^2\omega}{dk^2}$. Если ее значение не равно нулю, что импульс достаточно скоро расплывется, и передать его относительно невозмущенным на достаточно большое расстояние нельзя. Поэтому учет влияния дисперсии вместе с учетом затухания являются обязательными при проектировании оптоволоконных линий связи.

В области аномальной дисперсии групповая скорость, вычисленная по приведенным выше формулам, может оказаться больше c . Это не противоречит теории относительности, ибо групповая скорость выражает скорость передачи информации лишь тогда, когда волновой импульс в процессе распространения практически не изменяет своей формы. В области же аномальной дисперсии импульс сильно деформируется, и данный физический смысл групповой скорости теряется.

2.7 Оптическая спектроскопия

Прошлые лекции мы посвятили изучению интерференции, дифракции и дисперсии света. Давайте разберем, как эти явления помогают в получении информации о спектре электромагнитного излучения.

При рассмотрении дисперсии мы рассмотрели резкое увеличение поглощения света вблизи одной резонансной частоты. Более точно, у веществ в природе таких частот несколько, эти частоты разные для разных веществ, и именно на этих частотах и происходит поглощение или излучение света веществом. Это приводит к тому, что в непрерывном спектре пропускаемого через вещество света появляются темные линии. Интенсивность этих линий зависит от количества вещества и его состояния. *Оптическая спектроскопия* – изучение спектров взаимодействия электромагнитного излучения в видимом диапазоне волн с веществом. Именно этим методом был получен огромный объем информации о том, как устроено вещество на атомном и молекулярном уровне, какие свойства приобретают молекулы вещества в том или другом состоянии. Такой анализ достаточно прост, очень точен и чувствителен, и широко применяется в разных областях науки и техники.

Для определения спектральных характеристик используют спектральные приборы. Исторически первыми и наиболее распространенными приборами являются спектрометры с пространственным разделением излучения по длинам волн, для чего в их схему включен диспергирующий элемент (элемент, обладающий выраженной дисперсией). Схема такого спектрального прибора приведена на рисунке 55. Наиболее часто в качестве такого элемента используют дифракционные решетки, спектральные призмы и интерферометры Фабри-Перо.

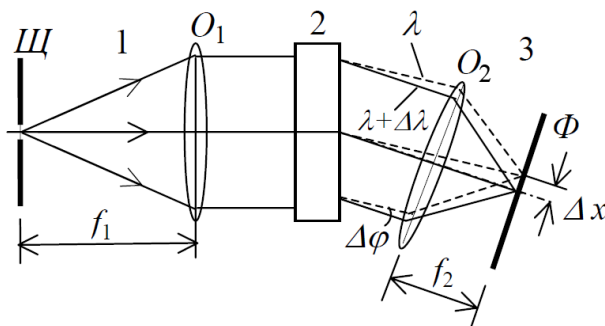


Рисунок 55. Схема спектрального прибора с пространственным разделением излучения: 1 – коллиматор с входной щелью Щ и объективом O_1 ; 2 – диспергирующий элемент; 3 – фокусирующая система с объективом O_2 ; Φ – фокальная плоскость.

Спектральная призма

Наиболее известным диспергирующим элементом является спектральная призма – т.е. многогранник, сделанный из прозрачного вещества, обладающего значительной дисперсией. При прохождении через призму поток лучей меняет свое направление, причем угол выхода лучей зависит от величины показателя преломления вещества призмы на данной длине волны.

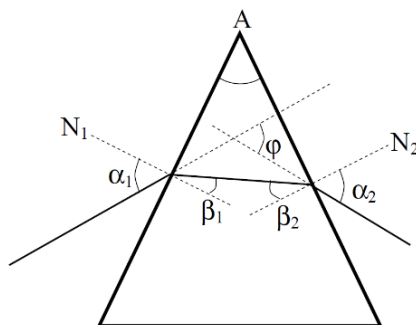


Рисунок 56. Спектральная призма.

Пусть на боковую сторону призмы под углом α_1 падает луч света (см. схему на рис. 56¹). Используя геометрические построения и закон Снеллиуса, можно записать:

$$\varphi = \alpha_1 + \alpha_2 - A, \quad A = \beta_1 + \beta_2, \quad \sin(\alpha_1) = n \sin(\beta_1), \quad \sin(\alpha_2) = n \sin(\beta_2).$$

Полагая величины углов α_1 , α_2 , β_1 , β_2 малыми, можно считать, что значения синусов этих углов приближенно равны самим этим углам. Отсюда находим угол отклонения луча в линзе и его производную по длине волны

$$\varphi = (n - 1)A, \quad \frac{d\varphi}{d\lambda} = D_\varphi = A \frac{dn}{d\lambda}.$$

Определенная таким образом величина D_φ называется *угловой дисперсией* и является важной характеристикой любого диспергирующего элемента. Именно она определяет разницу углов, на которую элемент преломляет волны с разной длиной волны.

Зададимся вопросом – какая минимальная разница длин волн может быть зарегистрирована при помощи призмы? Казалось бы, призма обеспечит преломление двух волн с длинами волн λ и $\lambda + \Delta\lambda$ под разными углами (хотя и разница между этими углами и будет мала). Отнесем экран, на котором будем наблюдать световую картину после призмы, подальше от нее. На большом расстоянии преломленные под разными углами волны разойдутся достаточно сильно, чтобы их можно было различить – и так для какой угодно малой разницы $\Delta\lambda$! Конечно, это рассуждение является неправильным – в нем не учитывается влияние дифракции. В данном случае экран наблюдения находится далеко от призмы, и можно

¹ Этот и последующие рисунки текущей главы взяты из учебного пособия Н. А. Маслова и других “Оптические приборы”, издательство Новосиб. гос. ун-та, Новосибирск, 2008.

воспользоваться приближением Фраунгофера. Вспомним, что при этом интенсивность света на экране для одной монохроматической волны задается следующей формулой:

$$I(\alpha) = I_0 \left(\frac{\sin(\alpha)}{\alpha} \right)^2, \quad \alpha = \frac{k\varphi d}{2},$$

где φ – величина угла, под которым наблюдается дифракция, относительно направления выхода исследуемой волны, d – характерный размер боковой поверхности призмы. Эта формула задает центральный максимум (при $\varphi = 0$) какой-то конечной ширины и последующие боковые минимумы и максимумы (см. рис. 49). Волны разных частот некогерентны, поэтому интенсивности от волн с разной длиной волны накладываются друг на друга (суммируются их интенсивности). Очевидно, что если эти длины волн очень близки друг к другу, то их центральные максимумы также будут близко друг к другу, и вследствие конечной ширины центральных максимумов их будет очень трудно отличить. С увеличением расстояния между призмой и экраном положение их центральных максимумов расходитя, но и ширина максимумов тоже увеличивается, дифракционная картина увеличивается подобно (см. раздел о дифракции Фраунгофера), и улучшения различимости не наблюдается. Как и было сказано ранее, именно дифракция ограничивает разрешающую способность спектральных приборов.

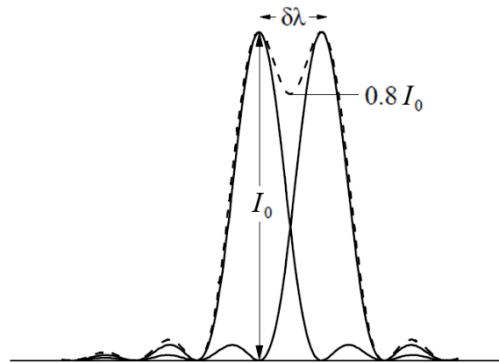


Рисунок 57. Критерий Релея.

Для условия разрешения двух близко стоящих спектральных линий в спектрометрах вводится *критерий Релея* – такие линии могут быть разрешены, если центральный максимум одной из этих линий находится не ближе к центральному максимуму другой линии, чем ее первый минимум. Иллюстрация предельного разрешения двух линий согласно этому критерию приведена на рисунке 57. Провал в распределении интенсивности между максимумами составляет в этом случае порядка 20 процентов.

Возвращаясь к призме, первый минимум наблюдается при $\alpha = \pi$, отсюда согласно критерию Релея для волн с длинами волн λ и $\lambda + \Delta\lambda$ находим минимальную разность $\Delta\lambda$, которую еще можно зарегистрировать

$$\frac{kd}{2} \Delta\varphi_{min} = \pi, \quad \Delta\lambda_{min} = \left(\frac{d\varphi}{d\lambda} \right)^{-1} \Delta\varphi_{min} = \frac{\lambda}{d} \frac{1}{A \frac{dn}{d\lambda}}.$$

Величину $\Delta\lambda_{min}$ называют *пределом разрешения*. Вводят *разрешающую силу* (или *спектральное разрешение*) спектрального прибора:

$$R = \frac{\lambda}{\Delta\lambda_{min}}.$$

Для призмы

$$R = Ad \frac{dn}{d\lambda}.$$

Величина Ad есть характерная длина основания призмы, обычно она не превышает 10 см, для оптического стекла $\frac{dn}{d\lambda} \sim 1,5 \cdot 10^3 \text{ см}^{-1}$. Отсюда теоретический предел разрешающей силы призмы $R \sim 10^4$, в реальности же эта сила меньше.

Дифракционная решетка

Дифракционная решетка представляет собой периодическую структуру прозрачных и непрозрачных щелей или штрихов, нанесенных на плоскую или цилиндрическую поверхность. Решетка может работать или на “просвет”, или в отраженном свете.

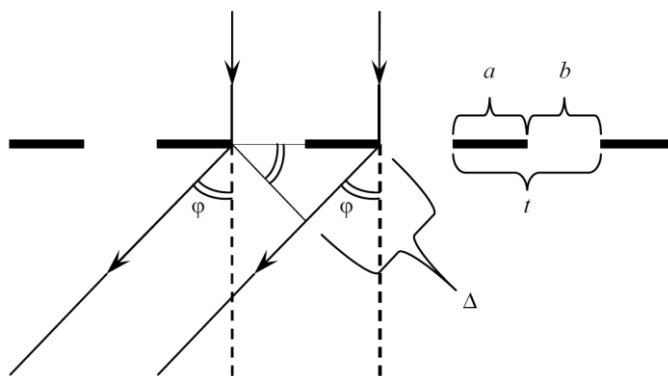


Рисунок 58. Дифракционная решетка, работающая на просвет.

Рассмотрим первый случай, когда свет проходит через решетку; схема такой решетки изображена на рисунке 58. Часть светового потока проходит через щели, часть – задерживается непрозрачными промежутками между щелями. Щели решетки узкие, следовательно, на них наблюдается сильная дифракция светового потока. Как и в случае призмы, экран наблюдения или находится достаточно далеко от решетки, или свет на него собирается линзой, находящейся от него на фокусном расстоянии. В обоих случаях реализуется дифракция в параллельных лучах, т.е. дифракция Фраунгофера. Картина на экране определяется интерференцией плоских волн, синхронно излучаемых поверхностью дифракционной решетки в заданном направлении. Комплексная амплитуда поля в точке наблюдения описывается соответствующим интегралом. Щели очень длинные, соответственно, по одному из направлений можно проинтегрировать с бесконечными пределами, после интегрирования интеграл приобретает вид

$$E(\varphi) = A \int E(x) \exp(-ikx \sin(\varphi)) dx,$$

где φ – угол между направлением на точку наблюдения и нормалью к поверхности решетки.

Поскольку $E(x) = 0$, если x принадлежит непрозрачным промежуткам между щелями, то этот интеграл разбивается на сумму интегралов по отдельным щелям решетки. Опустив его вычисление, сразу приведем угловое распределение интенсивности света

$$I(\varphi) = I_0 I_1(u) I_2(v), \quad I_1(u) = \left(\frac{\sin(u)}{u} \right)^2, \quad I_2(v) = \left(\frac{\sin(Nv)}{N \sin(v)} \right)^2,$$

$$u = \pi \frac{b \sin(\varphi)}{\lambda}, \quad v = \pi \frac{t \sin(\varphi)}{\lambda} = \frac{t}{b} u.$$

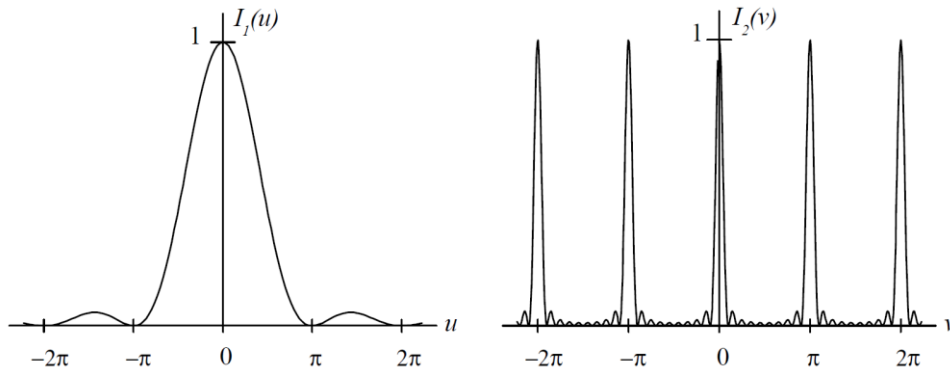


Рисунок 59. Графики функций $I_1(u)$ и $I_2(v)$ (при $N = 10$).

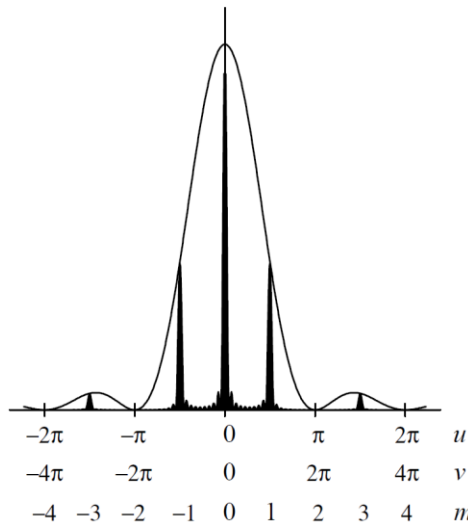


Рисунок 60. Дифракционная картина при $a = b$.

Зависимость $I(\varphi)$ – достаточно интересная функция. Она является произведением двух функций $I_1(u)$ и $I_2(v)$, графики которых приведены на рисунке 59. При $v = 0, \pm\pi, \pm2\pi, \dots$ в функции $I_2(v)$ присутствуют резкие максимумы, называемые *главными*, видны также сильно подавленные $N - 1$ побочных максимумов, в остальных точках значение этой функции мало. Ширина главных максимумов тем меньше, чем больше число периодов решетки N . Сама дифракционная картина будет также зависеть от отношения ширины щели к периоду решетки. Для случая $a = b$ эта картина представлена на рисунке 60. При $\varphi = 0$ наблюда-

ется центральный главный максимум, амплитуда остальных главных максимумов с ростом φ резко уменьшается, но вблизи центрального есть по крайней мере один или несколько боковых главных максимумов (зависит от величины отношения b/t) с еще значительной амплитудой. Номер (порядок) m главных максимумов определяется из условия

$$v = m\pi, \quad t \sin(\varphi_m) = m\lambda.$$

Таким образом, если сквозь решетку направить поток монохроматического света, то на экране мы увидим не одно, а несколько светлых пятен, вытянутых в ряд перпендикулярно штрихам решетки. Центральное пятно, соответствующее прямому прохождению света, не зависит от длины волны пропускаемого света, а вот положение боковых пятен, образованных боковыми главными максимумами, зависит! Таким образом, дифракционная решетка обладает ненулевой угловой дисперсией $D_\varphi \neq 0$ для боковых максимумов.

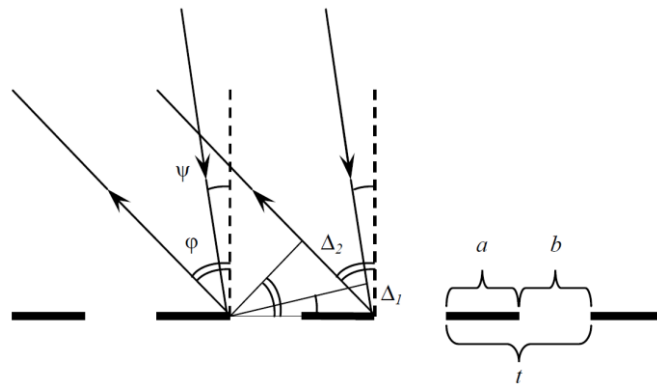


Рисунок 61. Отражающая дифракционная решетка.

Рассмотрим отражающую дифракционную решетку при падении на нее света под углом ψ к нормали плоскости решетки (см. рис. 61). Интенсивность света определяется той же самой формулой, но в этом случае

$$u = \frac{\pi b}{\lambda} \sin(\varphi + \psi) \cos(\psi), \quad v = \frac{\pi t}{\lambda} (\sin(\varphi) + \sin(\psi)).$$

Условие главных максимумов для такой решетки

$$t(\sin(\varphi) + \sin(\psi)) = m\lambda,$$

нулевой максимум наблюдается для угла $\varphi = -\psi$.

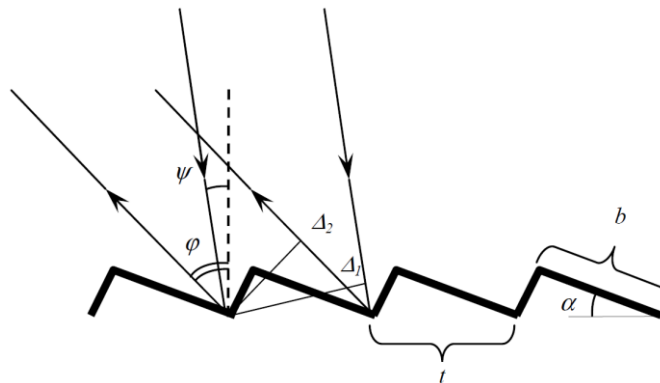


Рисунок 62. Отражающая дифракционная решетка.

Как покажем далее, угловая дисперсия тем сильнее, чем больше номер m главного максимума, для которого она наблюдается. С другой стороны, как уже отмечалось, интенсивность главных максимумов быстро падает с увеличением их номера. Этот недостаток устранен в другом типе дифракционных решеток, а именно в отражательных решетках с профилированной поверхностью. Схема поверхности такой решетки представлена на рисунке 62. Такую решетку также называют *эшелетт*. Разность хода лучей от соседних периодов в ней зависит от высоты ступеньки штриха, зависимость $v(\varphi)$ не изменяется по сравнению с предыдущим случаем, а

$$u = \frac{\pi b}{\lambda} \sin(\varphi + \psi + 2\alpha) \cos(\psi).$$

Центральный максимум функции $I_1(u)$ располагается при $\varphi = -(\psi + 2\alpha)$. Этот угол соответствует зеркальному отражению от наклонной грани штриха и носит название *угол блеска*. Подбирая величину α , можно добиться совпадения главного максимума $I_1(u)$ и главного максимума функции $I_2(v)$ с номером m , как показано на рисунке 63. Отметим, что это совпадение наблюдается при определенной длине волны, для других длин волн уже полного совпадения не будет, амплитуда выбранного главного максимума для них уменьшается, что необходимо учитывать при спектрометрических измерениях.

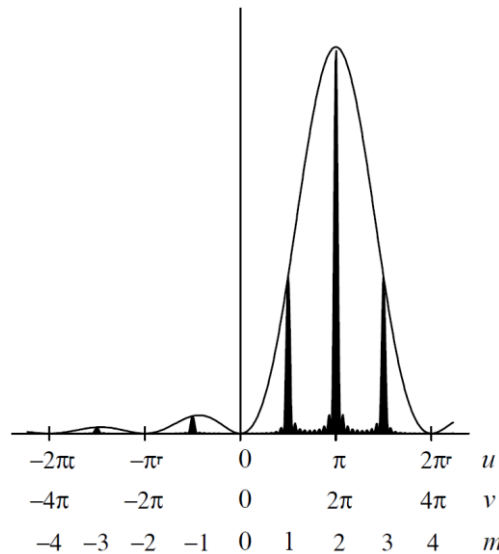


Рисунок 63. Дифракционная картина для эшелетта.

Рассчитаем основные характеристики дифракционной решетки. Для нахождения угловой дисперсии продифференцируем условие главных максимумов:

$$d[t(\sin(\varphi) + \sin(\psi))] = t \cos(\varphi) d\varphi = m d\lambda, \quad D_\varphi = \frac{d\varphi}{d\lambda} = \frac{m}{t \cos(\varphi)}.$$

Угловая дисперсия решетки пропорциональна порядку главного максимума и обратно пропорциональна расстоянию между штрихами (т.е., пропорциональна плотности штрихов).

Для разрешающей силы необходимо привлечь критерий Релея. Положение главного максимума определяется функцией $I_2(\nu)$, в которой расстояние от максимума до минимума равно π/N . Найдем, на сколько изменится угол φ при изменении ν на эту, достаточно малую величину:

$$d\nu_{min} = \frac{\pi}{N} = d \left[\frac{\pi t}{\lambda} (\sin(\varphi) + \sin(\psi)) \right] = \frac{\pi t}{\lambda} \cos(\varphi) d\varphi_{min}, \quad d\varphi_{min} = \frac{\lambda}{tN \cos(\varphi)}.$$

Отсюда находим

$$\Delta\lambda_{min} = \left(\frac{d\varphi}{d\lambda} \right)^{-1} \Delta\varphi_{min} = \frac{\lambda}{Nm}, \quad R = Nm.$$

Так как число N периодов решетки очень велико, то ее разрешающая сила может достигать 10^6 .

Как мы выяснили ранее, при использовании дифракционной решетки на экране наблюдается несколько вытянутых в ряд узких светлых участков, отвечающих разным порядкам m . При изменении длины волны света эти участки сдвигаются. Если есть сразу две волны с разными, но близкими длинами волн, то отвечающие одинаковому порядку светлые участки, образованные этими волнами, находятся близко друг к другу, и далеко от других светлых участков. При увеличении разности длин волн эти близкие когда-то участки разъезжаются, приближаются к другим, и расшифровка дифракционной картины становится затруднительной. *Областью свободной дисперсии* называется диапазон частот $G = \lambda_{max} - \lambda_{min}$, в котором главный максимум порядка m волны λ_{max} не перекрывает главный максимум порядка $m + 1$ волны λ_{min} . Уравнение на границу этой области относительно ее средней частоты λ определяется из условия совпадения углов φ этих волн:

$$m \left(\lambda + \frac{G}{2} \right) = (m + 1) \left(\lambda - \frac{G}{2} \right).$$

Отсюда область свободной дисперсии дифракционной решетки равна

$$G = \frac{\lambda}{m+1/2}.$$

Спектральный интервал, занимаемый исследуемым излучением, не должен превышать этой величины.

Часть III.

Квантовая физика

3.1 Кризис классической физики

К концу XIX века ученым-физикам казалось, что они знают все основные законы, описывающие окружающий мир. Или, как говорилось, здание физики почти построено. В самом деле, механика Ньютона раз за разом находила свое полное экспериментальное подтверждение, наиболее блестящим из которых явилось открытие планеты Нептун. За все века до этого астрономические открытия были прямым следствием наблюдений за звездным небом. Одним из результатов из таких наблюдений стало обнаружение небольших периодических возмущений орбиты Урана, которые на тот момент не имели объяснения. Была выдвинута гипотеза, что эти возмущения есть проявление влияния другой, неизвестной планеты. На основе механики Ньютона, после сложных, тщательных вычислений французский математик Леверье нашел параметры ее орбиты, а немецкий астроном Галле нашел ее на небе в точности в указанном месте. Как поэтически сказано, Нептун – планета, открытая “на кончике пера”. Кроме механики, и другие области физики – термодинамика, оптика, новоизобретенная электродинамика были проработаны достаточно хорошо, и давали полное согласие с экспериментами. На всем горизонте физики, по мнению лорда Кельвина, высказанному им на ставшей знаменитой лекции в Королевском институте 27 апреля 1900, виднелись лишь две тучки; давайте их рассмотрим.

Физики привыкли, что волны, которые могут распространяться, основаны на связанных колебаниях частиц какой-либо среды. И для электромагнитных волн также ввели свою, особую среду – эфир, в которую должен быть погружен наш мир. Естественно, были сделаны попытки обнаружить этот самый эфир. Эти попытки были основаны на том, что Земля движется в пространстве, следовательно, движется сквозь эфир, и, как следствие, должно быть различие в скоростях распространения света вдоль и поперек этого движения. Эксперименты же, наиболее известным из которых является опыт Майкельсона-Морли, показали полнейшую независимость скорости света от направления, что и стало одной из тучек, упомянутых лордом Кельвиным. Было сделано много гипотез о причине отрицательного результата (в частности, что Земля увлекает “эфирный ветер” за собой), в конечном же счете от идеи эфира полностью отказались, приняли, что электромагнитное поле является самостоятельным видом материи, а из обсуждения этой проблемы в конечном итоге родилась специальная теория относительности.

Тучкой номер два являлась неудачная попытка нахождения спектра излучения нагретого тела, исходя из законов классической термодинамики. Попытка решить эту проблему привела к созданию квантовой физики. Остается удивляться проницательности лорда Кельвина: он точно нащупал две болевые точки

современной ему науки, за которыми скрывались фундаментальные основы физики сегодняшнего дня.

3.2 Равновесное тепловое излучение

Из опыта известно, что нагретые тела излучают электромагнитные волны в широком диапазоне частот. Излучая, тела остывают. В то же время, тело может поглощать падающее на него электромагнитное излучение, нагреваясь при этом. При термодинамическом равновесии температура тела постоянна, т.е. в единицу времени оно излучает столько же энергии, сколько и поглощает. Вводят *излучательную способность* тела $r(\omega, T)$ – это энергия, излучаемая нагретым до температуры T телом с единицы поверхности за единицу времени в единичном интервале частот вблизи частоты ω . Аналогичным образом вводят *поглощательную способность тела* $a(\omega, T)$ – доля от падающей энергии электромагнитного излучения, которая поглощается телом. Соотношение между этими величинами устанавливает *закон излучения Кирхгофа*: отношение излучательной способности любого тела к его поглощательной способности одинаково для всех тел при данной температуре для данной частоты и не зависит от их формы и химической природы. Это отношение равно излучательной способности абсолютно черного тела $f(\omega, T)$, поскольку его поглощательная способность по определению равна единице:

$$\frac{r(\omega, T)}{a(\omega, T)} = f(\omega, T).$$

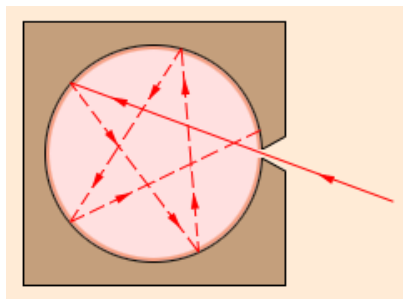


Рисунок 64. Замкнутая полость с небольшим отверстием.

Абсолютно черных тел в природе не существует, но хорошей моделью такого тела является замкнутая полость с небольшим отверстием, поперечные размеры которого малы по сравнению с размерами самой полости; эта полость показана на рисунке 64. Свет, проходящий внутрь полости, после многократных отражений практически полностью поглотится. Спустя определенное время температура внутри полости стабилизируется, излучение с внутренних стенок придет в термодинамическое равновесие со стенками, следовательно, выходящее из полости излучение можно считать приближенно равным $f(\omega, T)$. Экспериментально полученные зависимости равновесной плотности энергии теплового излучения $u(\omega, T) = 4f(\omega, T)/c$ абсолютно черного тела от длины волны излучения при разных температурах тела приведены на рисунке 65.

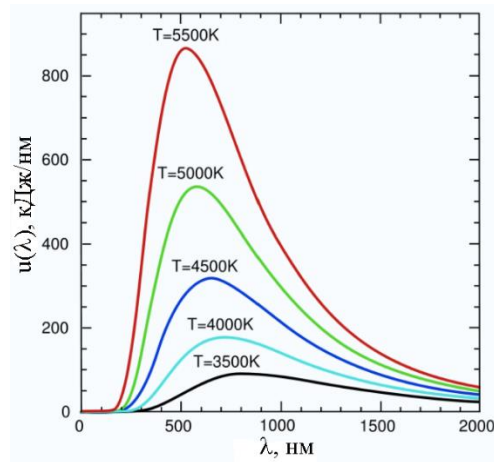


Рисунок 65. Равновесная плотность энергии теплового излучения абсолютно черного тела.

Австрийскими физиками Стефаном и Больцманом было экспериментально и теоретически получено, что полная мощность излучения с единицы площади поверхности абсолютно черного тела равна

$$R = \sigma T^4,$$

где σ – постоянная Стефана-Больцмана. Данная зависимость называется *законом Стефана-Больцмана*, величина R есть площадь под одной из кривых рисунка с соответствующим значением T .

Можно заметить, что с ростом температуры максимум излучательной способности смещается в сторону более коротковолновой области. Действительно, наш опыт говорит, что при высоких температурах тела излучение идет преимущественно в видимом диапазоне частот, при более низких – в инфракрасном. Немецкий физик Вин сформулировал *закон смещения Вина*

$$\lambda_{max} = \frac{b}{T},$$

где b есть постоянная Вина.

Оба упомянутых закона выведены в рамках классической физики и полностью согласуются с экспериментальными данными, но продвинуться далее, а именно теоретически получить сам вид $f(\omega, T)$ оказалось при этом невозможным. Действительно, строгое следование классической физике приводит к *формуле Рэлея-Джинса*

$$f(\omega, T) = kT \frac{\omega^2}{4\pi^2 c^2},$$

k – постоянная Больцмана. Эта формула предполагает неограниченное, квадратичное возрастание спектральной плотности излучения при возрастании его частоты. Это не только противоречит экспериментальным данным, но и бессмысленно – согласно этой формуле, вся тепловая энергия любого тела должна немедленно перейти в коротковолновое излучение. Такое гипотетическое явление было названо *ультрафиолетовой катастрофой*. Оказывается, формула Рэлея-Джинса удовлетворительно описывает только длинноволновую часть спектра.

На основе имеющихся данных немецкий физик Макс Планк в 1900 году составил формулу, которая великолепно подходит к экспериментальным результатам во всей области частот

$$f(\omega, T) = \frac{\omega^2}{4\pi^2 c^2} \frac{\hbar \omega}{\exp\left(\frac{\hbar \omega}{kT}\right) - 1}.$$

В этой формуле используется так называемая *приведенная постоянная Планка* $\hbar = \frac{h}{2\pi} = 1.05 \cdot 10^{-34}$ Дж·с, она выражается через собственно *постоянную Планка* $h = 6,62 \cdot 10^{-34}$ Дж·с. При $\frac{\hbar \omega}{kT} \ll 1$ формула Планка переходит в формулу Релея-Джинса.

Если бы Планк ограничился составлением этой формулы, он бы вошел в историю как человек, угадавший закон излучения. Но он пошел далее, попытался получить свою формулу из каких-либо теоретических соображений, и на этом пути первым приоткрыл занавес над квантовой физикой. Как мы уже знаем, исходя только из соображений классической физики этого сделать нельзя, и Планку пришлось сделать допущение, что полная энергия осцилляторов с частотой ω может быть разделена на точное число равных частей (элементов, или квантов) величиной $\hbar \omega$. Именно с этого допущения и началось постепенное становление квантовой физики.

3.3 Фотоэффект

Физическая сущность квантов проявилась при экспериментальном и теоретическом изучении фотоэффекта. *Фотоэффект* есть испускание (эмиссия) электронов в вакуум или другую разреженную среду с поверхности твердого тела или жидкости при падении на нее светового потока. Открыл это явление тот самый немецкий физик Генрих Герц, который доказал существование электромагнитных волн, экспериментально подтвердив электромагнитную теорию света Максвелла.

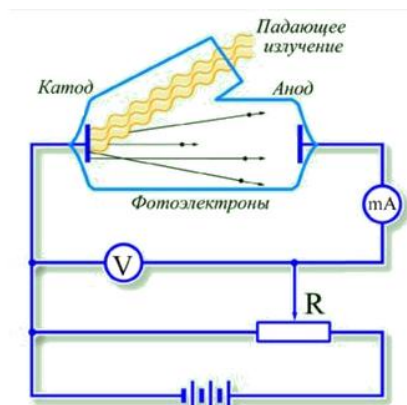


Рисунок 66. Схема установки для изучения фотоэффекта.

Электроны, вышедшие под действием светового потока из тела (в рассматриваемом случае, металла), называются *фотоэлектронами*, а сам электрод – *фотокатодом*. Если рядом поместить другой электрод (анод) и приложить к ним

напряжение соответствующей полярности, то фотоэлектроны устремятся к аноду, и в цепи потечет ток; схема эксперимента изображена на рисунке 66. Эта схема есть вакуумный фотодиод, его вольтамперная характеристика приведена на рисунке 67. Видно, что ток прекращается только при отрицательном, запирающем напряжении величиной U_3 . Можно заключить, что фотоэлектроны после выхода из проводника имеют ненулевую начальную скорость, максимальное значение которой определяется из равенства

$$\frac{mv_{max}^2}{2} = eU_3.$$

При увеличении ускоряющего потенциала ток выходит на насыщение, когда все эмитируемые электроны достигают анода.

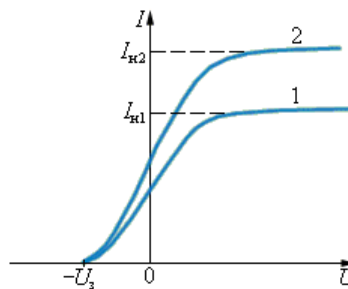


Рисунок 67. Вольтамперная характеристика фотодиода при двух различных интенсивностях падающего на фотокатод света.

Были установлены следующие экспериментальные законы фотоэффекта в области слабых токов:

1. При фиксированной частоте падающей на фотокатод световой волны величина тока насыщения прямо пропорциональна ее интенсивности;
2. Максимальная начальная кинетическая энергия фотоэлектронов не зависит от интенсивности света, но линейно зависит от его частоты;
3. Для каждого материала фотокатода существует длина волны излучения, при превышении которой эмиссия фотоэлектронов полностью прекращается. Эту длину волны называют *красной границей фотоэффекта*.

Эти закономерности никак не могли быть объяснены классической физикой. Для построения теории фотоэффекта Альберт Эйнштейн в 1905 году создал квантовую теорию света. Он предположил, что при распространении света энергия не распределяется непрерывно, а складывается из конечного числа квантов энергии (фотонов), движущихся в вакууме со скоростью света. Энергия отдельного фотона, соответствующая свету с частотой ω , равна

$$E = \hbar\omega.$$

Фотоны обладают свойствами настоящих частиц и характеризуются массой и импульсом. Используя формулу $E = mc^2$, можно записать, что для фотонов

$$m = \frac{\hbar\omega}{c^2}, \quad p = mc = \hbar \frac{\omega}{c} = \hbar k.$$

Фотон, попадая в проводник, может столкнуться со свободным электроном и передать ему свою энергию. Если энергия электрона после этого окажется достаточно велика, то он сможет преодолеть силы, удерживающие его в металле, и выйти в вакуум. Соответствующий закон сохранения энергии

$$\hbar\omega = A + \frac{mv_{max}^2}{2},$$

A есть работа выхода, которую нужно отдать электрону при выходе из металла. Это уравнение, помимо остальных свойств фотоэффекта, определяет его красную границу, когда

$$\hbar \frac{2\pi c}{\lambda_{кг}} = A.$$

Теория фотоэффекта Эйнштейна стала первой полноценной квантовой теорией физики, именно за ее создание Эйнштейн в 1921 году получил Нобелевскую премию.

3.4 Модель атома Бора

Еще одно несоответствие между классической физикой и экспериментальными данными возникло при исследовании микромира, в данном случае – внутреннего устройства атомов. В начале XX века были две основные гипотезы о структуре атома (см. рис. 68):

- а. Модель Томсона, предложенная в 1904 году английским физиком Джозефом Джоном Томсоном, который ранее, в 1897 году, открыл электрон. Согласно этой модели атом состоит из электронов, помещённых внутри распределенного положительного заряда, который компенсирует отрицательные заряды электронов, подобно отрицательно заряженным «изюминкам» в положительно заряженном «пудинге». Электроны, как предполагалось, были распределены по всему атому.
- б. Планетарная модель с сосредоточенным в центре атома положительным зарядом (ядре атома), вокруг которого, как планеты вокруг Солнца, движутся электроны.

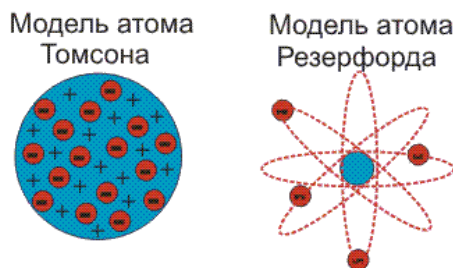


Рисунок 68. Сравнение моделей атомов Томсона и Резерфорда.

В 1911 году опыт британского физика Резерфорда отверг модель Томсона в пользу планетарной модели. Однако, согласно законам электродинамики, любое заряженное тело,двигающееся с ускорением, должно излучать электромагнитные волны; движение по окружности не есть исключение. Оценки показывают,

что излучая, электроны полностью потеряют всю свою энергию и упадут на ядро за очень короткий промежуток времени.

Для спасения планетарной модели, а также для объяснения наблюдавшихся закономерностей в спектре атома водорода, датский ученый Нильс Бор в 1913 году ввел следующие утверждения (постулаты):

1. Атом может длительное время находиться только в определенных, так называемых *стационарных* состояниях. Эти состояния должны удовлетворять *правилу квантования*. Энергии стационарных состояний образуют дискретный спектр.
2. Излучение испускается или поглощается атомом в виде светового кванта только при переходе электрона из одного стационарного состояния в другое, причем выполняется правило частот

$$\hbar\omega = E_n - E_m,$$

где E_n, E_m — энергии соответствующих состояний.

Правило квантования гласит, что стационарными являются лишь те орбиты, при движении по которым момент импульса электрона равен целому числу постоянных Планка \hbar :

$$L = mvr = n\hbar.$$

Закон Ньютона для движения по окружности принимает вид

$$\frac{1}{4\pi\epsilon_0} \frac{e^2}{r^2} = \frac{mv^2}{r}.$$

Отсюда исключаем скорость и находим радиусы стационарных орбит атома водорода и соответствующую им энергию

$$r_n = 4\pi\epsilon_0 \frac{\hbar^2}{me^2} n^2, \quad E_n = \frac{m}{2} \left(\frac{1}{4\pi\epsilon_0} \frac{e^2}{mr_n} \right) - \frac{1}{4\pi\epsilon_0} \frac{e^2}{r_n} = - \frac{1}{(4\pi\epsilon_0)^2} \frac{me^4}{2\hbar^2} \frac{1}{n^2} = - \frac{R_y}{n^2}.$$

$R_y = 13,6$ эВ называется *постоянной Ридберга*. При $n = 1$ состояние имеет наименьшую энергию, такое состояние называют *основным*. При $n \rightarrow \infty$ уровни, сгущаясь, приближаются снизу к предельному значению $E = 0$.

Существует также правило квантования, называемое правилом квантования Бора-Зоммерфельда. Это правило применяется для систем, которые можно представить как осциллятор; причем допускается медленное (адиабатическое) изменение параметров этого осциллятора.

Теория Бора подошла к объяснению внутриатомных процессов с принципиально новых позиций, став при этом первой полуквантовой теорией атома. Кроме того, она в точности объяснила закономерности спектра водородоподобных атомов. В то же время, ей присущи значительные недостатки: она не может объяснить интенсивность спектральных линий, справедлива только для водородоподобных атомов и не работает для атомов других элементов, из двух уравнений, лежащих в ее основе, одно (уравнение движения электрона) является классическое, другое (уравнение квантования орбит) — квантовое. Теория Бора была

в дальнейшем заменена современной квантовой механикой, основанной на более общих и непротиворечивых исходных положениях, однако правила квантования широко используются и в наши дни как приближенные соотношения.

3.5 Волны де Бройля. Волновая функция

Выше мы выяснили, что теория Бора, хотя и являлась важным шагом на пути становления квантовой механики, не является последовательно квантовой теорией. Как же построить такую теорию, от чего оттолкнуться?

Вспомним, что мы изучали о природе света – с одной стороны, подробно разобрали его волновые свойства, с другой – увидели, что ему свойственно вести себя как потоку частиц. Говорят, что свет проявляет *корпускулярно-волновой дуализм*. Именно эта идея дуализма и послужила основой для создания квантовой механики; первый шаг в этом направлении сделал 1923 году французский физик Луи де Бройль. Он выдвинул гипотезу, что дуализм имеет универсальный характер, т.е. не только фотоны имеют волновые свойства, но и все материальные частицы. Любой частице, имеющей энергию E и импульс \mathbf{p} , можно сопоставить какой-то волновой процесс с частотой и волновым вектором

$$\omega = \frac{E}{\hbar}, \quad \mathbf{k} = \frac{\mathbf{p}}{\hbar}.$$

Эта связь получили название *уравнений* (или *соотношений*) *де Бройля*, а соответствующая длина волны – *длина волны де Бройля*, она имеет следующее выражение через импульс:

$$\lambda = \frac{h}{p}.$$

В качестве примера, если ускорить электрон разностью потенциалов порядка 100 В, то его длина волны будет соответствовать рентгеновской области спектра. А вот для макроскопических объектов эта длина волны просто мизерна, движение таких тел описывается классической механикой.

Согласно гипотезе де Бройля, движение частиц описывается некой волновой функцией Ψ . Свободно двигающейся частице соответствует плоская монохроматическая волна, распространяющаяся в том же направлении, ее волновая функция имеет соответствующий вид:

$$\Psi(\mathbf{r}, t) = A \exp\left(-\frac{i}{\hbar}(Et - \mathbf{p} \cdot \mathbf{r})\right).$$

Фазовая скорость волны де Бройля определяется из условия

$$Et - \mathbf{p} \cdot \mathbf{r} = \text{const}, \quad v_{\phi} = \frac{E}{p} = \frac{c^2}{v} > c.$$

Вспомним, что скорость перемещения информации и энергии характеризует не фазовая, а групповая скорость. Найдем ее

$$v_{\text{гр}} = \frac{dE}{dp} = \frac{pc^2}{E} = \frac{mvc^2}{mc^2} = v,$$

групповая скорость волны де Бройля совпадает со скоростью частицы; при выводе было использовано соотношение между энергией и импульсом частицы

$$E^2 - p^2 c^2 = m^2 c^4, \quad 2EdE - 2c^2 p dp = 0.$$

Вслед за Бором де Бройль рассмотрел атом водорода, исходя из того, что на длине стационарной орбиты должно укладываться целое число длин волн:

$$2\pi r_n = n\lambda, \quad \lambda = \frac{h}{mv}, \quad mvr_n = n\hbar.$$

В итоге получилось уравнение, совпадающее с правилом квантования Бора.

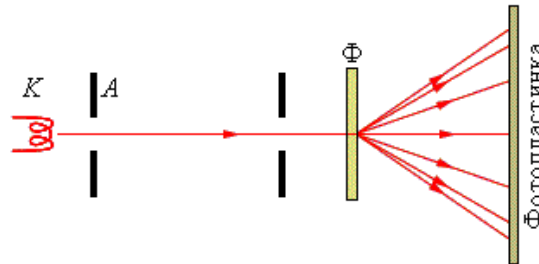


Рисунок 69. Схема опыта Томсона.

Теория де Бройля была для своего времени достаточно смелой и совершенно неочевидной, пока не появились ее экспериментальные подтверждения. Сперва, в 1927 году, наблюдалась дифракционная картина при отражении пучка электронов от кристалла никеля. В 1928 году английский физик Томсон (сын ученого Томсона, который открыл электрон) наблюдал дифракцию электронов, прошедших сквозь тонкую золотую фольгу, схема эксперимента приведена на рисунке 69. Дифракционные картины при различных временах экспозиции показаны на рисунке 70. Такой же результат наблюдался и для одиночных электронов, т.е. волновые свойства присущи каждой частице в отдельности.



Рисунок 70. Картина дифракции электронов при длительной (a) и при короткой экспозиции (b).

3.6 Уравнение Шредингера

Второй шаг в становлении квантовой механики сделал австрийский физик Эрвин Шредингер. Узнав о гипотезе де Бройля и о предварительных подтверждающих ее результатах, он в 1926 году смог вывести уравнение, которому должна подчиняться волновая функция де Бройля.

Напомним, что классическое волновое уравнение имеет вид

$$\Delta\Psi(\mathbf{r}, t) = \frac{1}{v^2} \frac{\partial^2 \Psi(\mathbf{r}, t)}{\partial t^2}.$$

Для монохроматических волн его решение представимо в виде

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r})e^{-i\omega t}.$$

Подставляя это выражение в волновое уравнение, получаем для комплексной амплитуды $\psi(\mathbf{r})$ уравнение Гельмгольца:

$$\Delta\psi(\mathbf{r}) + k^2\psi(\mathbf{r}) = 0, \quad k^2 = \omega^2/v^2.$$

Заметим, что при переходе из одной среды в другую изменяется волновое число k , частота же ω всегда постоянна.

Шредингер предположил, что это уравнение должно описывать и волновую функцию де Бройля. Плоская волна соответствует свободному движению частицы, в присутствии внешних сил волновая функция усложняется вслед за характером движения частицы.

Рассмотрим движение нерелятивистских частиц в потенциальных полях. Соотношение де Бройля $E = \hbar\omega$ показывает, что при постоянной E частота ω тоже есть константа. Следовательно, уравнение Гельмгольца можно применять, если полная энергия частицы остается постоянной, а именно, в потенциальных полях:

$$E = \frac{p^2}{2m} + U(\mathbf{r}) = \text{const.}$$

В таком случае волновая функция представима в виде

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r})e^{-i\frac{E}{\hbar}t}.$$

Потенциальная энергия $U(\mathbf{r})$ описывает силы, действующие на частицу. Зная ее, можно найти выражение для волнового числа

$$k^2 = \frac{p^2}{\hbar^2} = \frac{2m}{\hbar^2}(E - U).$$

Подставляя это выражение в уравнение Гельмгольца, получаем *стационарное уравнение Шредингера* для комплексной амплитуды волновой функции

$$\Delta\psi(\mathbf{r}) + \frac{2m}{\hbar^2}(E - U(\mathbf{r}))\psi(\mathbf{r}) = 0.$$

Другая (операторная) форма стационарного уравнения Шредингера есть

$$\hat{H}\psi(\mathbf{r}) = E\psi(\mathbf{r}).$$

\hat{H} есть *оператор Гамильтона*, имеющий вид

$$\hat{H} = -\frac{\hbar^2}{2m}\Delta + U(\mathbf{r}),$$

т.е. им задается линейная операция, действующая на функцию ψ . Уравнение Шредингера в этом случае есть уравнение на отыскание собственных значений и собственных функций оператора Гамильтона.

Обобщим уравнение Шредингера на случай зависимости от времени. Представим волновую функцию в виде

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r})e^{-i\frac{Et}{\hbar}},$$

где $\psi(\mathbf{r})$ должно удовлетворять стационарному уравнению Шредингера и возьмем частную производную по времени от обеих частей этого выражения:

$$\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) = -\frac{i}{\hbar}E\psi(\mathbf{r})e^{-i\frac{Et}{\hbar}} = -\frac{i}{\hbar}\hat{H}\psi(\mathbf{r})e^{-i\frac{Et}{\hbar}}.$$

В окончательном виде получаем общую форму уравнения Шредингера в нестационарном случае:

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) = \hat{H}\Psi(\mathbf{r}, t) = \left(-\frac{\hbar^2}{2m}\Delta + U(\mathbf{r}, t)\right)\Psi(\mathbf{r}, t).$$

Как видно, уравнение Шредингера является уравнением первого порядка по времени. Поэтому для его решения необходимо задать одно начальное условие, а именно, волновую функцию в некоторый начальный (например, нулевой) момент времени. Тогда, зная волновую функцию системы в начальный момент времени, проинтегрировав, мы сможем определить волновую функцию в любой наперед заданный момент времени.

Таким образом, нестационарное уравнение Шредингера есть основное уравнение квантовой механики и фактически приходит на смену классическому уравнению Ньютона.

3.7 Физический смысл волновой функции

В настоящий момент квантовая механика является одной из наиболее успешных и досконально проверенных теорий в истории науки. Однако, до сих пор существуют несколько ее интерпретации, доминирующей из которых является т.н. *копенгагенская интерпретация*. Согласно ей, физический смысл имеет квадрат модуля волновой функции (в общем виде, комплексной) $|\Psi(\mathbf{r}, t)|^2 = \Psi^* \Psi$, который определяет плотность вероятности нахождения частицы в данной точке в данный момент времени. Отсюда вероятность обнаружения частицы в пределах объема dV , окружающего точку \mathbf{r} , в момент времени t равна

$$dP = |\Psi(\mathbf{r}, t)|^2 dV.$$

Вероятностный характер квантовой механики кардинально отличается от строго определенного, детерминированного характера механики классической. В частности, отсутствует понятие точки положения частицы в пространстве.

Такая интерпретация волновой функции налагает на нее ряд ограничительных условий. Эта функция должна быть конечной (вероятность не может быть больше единицы), однозначной (вероятность не может быть неоднозначной величиной) и непрерывной (вероятность не может изменяться скачком) во всей области изменения пространственных переменных. Исходя из ее физического смысла, на волновую функцию налагаются следующее условие нормировки:

$$\int |\Psi|^2 dV = 1.$$

3.8 Принцип неопределенности Гейзенберга

Важный характер в квантовой механике принимает вопрос о измерениях различных физических величин частицы. Действительно, пусть мы измеряем ее пространственное положение. С одной стороны, в результате измерений мы получаем строго определенное значение одной или нескольких ее координат, с другой стороны, согласно вышеизложенной интерпретации квантовой механики, у частицы вообще нет строго определенного положения в пространстве! Утверждается, что законы квантовой механики дают статистическую вероятность измерения физических величин. Каждое измерение физической величины квантово-механической системы, находящейся в одном и том же состоянии (с одинаковой волновой функцией), дает разный результат, вероятностный характер измерений принципиально неустраним. Это означает, что есть среднее значение наблюдаемой физической величины, и есть разброс данных отдельных измерений вокруг этого среднего значения, характеризуемый среднеквадратическим отклонением. Это отклонение зависит от схемы измерения, оно не может быть сведено к нулю, но, вроде бы, ничего не мешает сделать его пренебрежимо малым.

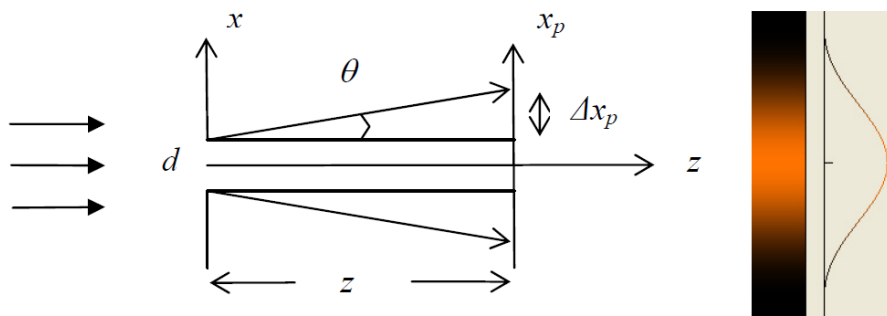


Рисунок 71. Дифракция пучка электронов на щели. Справа – картина дифракции в плоскости экрана и соответствующее распределение интенсивности (показан главный максимум).

Поставим следующий эксперимент (см. рис. 71): пусть на экран с отверстием размера d налетает пучок электронов с импульсом p . За отверстием в силу дифракции волн де Бройля пучок расширяется с характерным углом $\theta \sim \lambda/d$. Можно сказать, что при прохождении через отверстие появляется разброс поперечных импульсов $\Delta p_x \sim p \sin(\theta)$. В то же время, благодаря отверстию мы одновременно локализуем (определяем положение в пространстве) электроны с точностью $\Delta x \sim d$. С учетом выражения для длины волны де Бройля получаем

$$\Delta p_x \Delta x \sim \hbar.$$

Оказывается, это соотношение носит фундаментальный характер и называется *принципом неопределенности Гейзенберга*. Он появляется, поскольку любое измерение квантово-механической системы неизбежно воздействует на эту си-

стему. Сам факт проведения нами измерения, например, местоположения частицы, приводит к изменению ее скорости, причем непредсказуемому (и наоборот). Более строгая формулировка этого принципа гласит, что существует предел точности одновременного измерения координаты и импульса частицы, равный

$$\Delta p_x \Delta x \geq \hbar/2.$$

В эксперименте мы можем повышать точность измерения координаты, но если мы одновременно измеряем и импульс частицы, то точность измерения импульса при этом падает, и наоборот.

Принцип неопределенности Гейзенберга выполняется не только для пары физических величин координата/импульс, но и для многих других. Несколько особняком стоит подобное соотношение для энергии и времени

$$\Delta E \Delta t \geq \hbar/2.$$

Здесь ΔE есть неопределенность энергии квантово-механической системы в состоянии, которое не является полностью стационарным, т.е. обладает конечным временем жизни Δt .

3.9 Операторы в квантовой механике

Ранее мы ввели оператор Гамильтона как линейную операцию, действующую на волновую функцию. Вообще, оператор – это некоторое действие над функцией, однозначно определяющее другую функцию того же типа. К квантовой механике постулируется, что каждой наблюдаемой физической величине L может быть сопоставлен линейный самосопряженный оператор \hat{L} над пространством комплексных функций со скалярным произведением, определенным как

$$\langle \psi | \varphi \rangle = \int \psi^* \varphi dV,$$

интеграл берется по всей области изменения пространственных переменных. Давайте разберем написанное подробнее.

Волновая функция есть комплексная функция, отвечающая определенным, сформулированным выше условиям. Оператор \hat{L} над этой функцией определяет правило, по которому из одной волновой функции получается другая комплексная функция, отвечающая тем же условиям. Линейность оператора означает, что

$$\hat{L}(\psi + \varphi) = \hat{L}\psi + \hat{L}\varphi.$$

Линейный оператор называют *самосопряженным*, или *эрмитовым*, если для любых функций ψ и φ он отвечает условию

$$\langle \hat{L}\psi | \varphi \rangle = \langle \psi | \hat{L}\varphi \rangle.$$

Определим собственные значения L_n и собственные функции ψ_n оператора \hat{L} уравнением

$$\hat{L}\psi_n = L_n\psi_n.$$

Можно показать, что все собственные значения самосопряженного оператора являются вещественными числами, а его собственные функции ортогональны:

$$\langle \psi_n | \psi_m \rangle = 0 \text{ при } m \neq n,$$

$$\langle \psi_n | \psi_m \rangle = 1 \text{ при } m = n,$$

последнее равенство записано с учетом нормировки волновой функции. Более того, доказывается, что собственные функции эрмитового оператора образуют полную систему, т.е. любую функцию Ψ , определенную в той же области переменных и подчиняющуюся тому же классу граничных условий, можно представить в виде

$$\Psi = \sum_n c_n \psi_n.$$

Так же постулируется, что среднее значение наблюдаемой физической величины, измеряемое в квантово-механической системе, находящейся в состоянии, описываемой волновой функцией Ψ , равно

$$\langle L \rangle = \langle \Psi | \hat{L} \Psi \rangle = \int \Psi^* \hat{L} \Psi dV = \sum_n c_n^2 L_n.$$

Величину c_n^2 можно рассматривать как интенсивность, с которой состояние, описываемое собственной функцией ψ_n , представлено в состоянии Ψ . Если же Ψ совпадает с одной из функций ψ_n , все измерения такой системы будут давать величину L_n , $\langle L \rangle = L_n$, а среднее квадратическое отклонение $\Delta L = 0$.

Оператором координаты является умножение волновой функции на эту координату, отсюда следует формула для нахождения среднего положения частицы:

$$\hat{x} = x, \quad \langle x \rangle = \int x |\Psi|^2 dV.$$

Оператор импульса задается выражением

$$\hat{p}_x = -i\hbar \frac{\partial}{\partial x}, \quad \hat{\mathbf{p}} = -i\hbar \text{grad},$$

откуда следует выражение для оператора кинетической энергии

$$\hat{T} = -\frac{\hbar^2}{2m} \Delta.$$

Две физические величины L и M одновременно измеримы с произвольной точностью, если для одного и того же состояния $\Delta L = \Delta M = 0$. Видно, что для этого должны совпадать хоть какие-то их собственные функции. Можно показать, что для этого необходимо, чтобы операторы этих величин были коммутируемы, т.е. для любой функции ψ

$$\hat{L}(\hat{M}\psi) = \hat{M}(\hat{L}\psi).$$

Это есть математическое выражение принципа неопределенности. В частности, не коммутируют между собой операторы координаты и импульса.

3.10 Некоторые частные решения уравнения Шредингера

Рассмотрим решения уравнения Шредингера для некоторых характерных видов потенциальной энергии, которые помогут понять основные особенности поведения волновой функции.

Прохождение частицы через потенциальный барьер

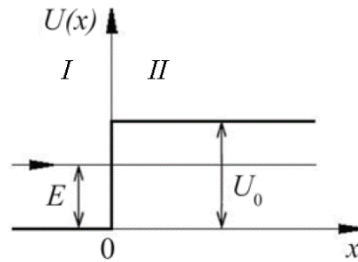


Рисунок 72. Ступенчатый потенциальный барьер.

Первым рассмотрим простейший пример – частица движется в свободном пространстве, а в какой-то точке на нее действует сосредоточенная сила, замедляющая ее движение. Соответствующее распределение потенциальной энергии имеет вид как показано на рисунке 72. *Потенциальным барьером* называют область пространства, где потенциальная энергия частицы больше, чем в окружающих областях. Пусть полная энергия частицы равна E . Возможны два случая:

Случай первый, $E \geq U_0$. Согласно классической механике, частица перейдет из области I в область II и продолжит свое движение, но уже с меньшей кинетической энергией, равной $E - U_0$. Для квантовой механики необходимо записать уравнение Шредингера для каждой из областей, и потом “сшить” получившиеся результаты. Для области I соответствующее уравнение примет вид

$$\frac{d^2}{dx^2} \psi_1(x) + \frac{2m}{\hbar^2} E \psi_1(x) = 0.$$

Получилось линейное дифференциальное уравнение второго порядка с постоянными коэффициентами. Его решением является линейная комбинация вида

$$\psi_1(x) = a_1 e^{+ik_1 x} + b_1 e^{-ik_1 x}, \quad k_1 = \sqrt{2mE/\hbar^2}.$$

Первое из слагаемых получившегося решения описывает волну, распространяющуюся в положительном направлении по x – это волна, падающая на барьер, второе слагаемое – отраженную от барьера волну, распространяющуюся в обратном направлении. Таким образом, в первой области волновая функция является суперпозицией двух волн.

Для второй области можно записать аналогичное уравнение Шредингера

$$\frac{d^2}{dx^2} \psi_2(x) + \frac{2m}{\hbar^2} (E - U_0) \psi_2(x) = 0,$$

но, исходя из физического смысла задачи, в его решении необходимо оставить только преломленную волну, распространяющуюся в положительном направлении

$$\psi_2(x) = a_2 e^{+ik_2 x}, \quad k_2 = \sqrt{2m(E - U_0)/\hbar^2}.$$

На границе потенциального барьера, при $x = 0$, потребуем выполнения условий, накладываемых на волновую функцию:

$$\psi_1(0) = \psi_2(0), \quad \psi'_1(0) = \psi'_2(0).$$

Эти два условия задают следующую систему алгебраических уравнений относительно неизвестных коэффициентов a_1 , b_1 и a_2 :

$$\begin{cases} a_1 + b_1 = a_2 \\ k_1(a_1 - b_1) = k_2 a_2 \end{cases}.$$

Исходя из этой системы, можно найти отношение амплитуд отраженной и падающей волн, а из аналогии с оптикой определить коэффициент отражения от барьера как квадрат этого отношения:

$$\frac{b_1}{a_1} = \frac{k_1 - k_2}{k_1 + k_2}, \quad R = \left(\frac{b_1}{a_1} \right)^2 = \frac{(k_1 - k_2)^2}{(k_1 + k_2)^2}.$$

Так же из аналогии с оптикой можно определить коэффициент прохождения

$$T = 1 - R.$$

С корпускулярной точки зрения, R есть вероятность частице отразиться от барьера, а T – вероятность преодолеть барьер, проникнуть в область II. Исходя из условий задачи, окончательно получаем

$$R = \left(\frac{1 - \sqrt{1 - U_0/E}}{1 + \sqrt{1 - U_0/E}} \right)^2 > 0.$$

В отличие от классической механики, у частицы есть ненулевая вероятность отразиться от барьера, даже если и ее энергия больше высоты этого барьера. Только при $E \gg U_0$ коэффициент отражения $R \rightarrow 0$.

Случай второй, $E < U_0$. Согласно классической механике, переход частице в область II невозможен – частица, достигая барьера, останавливается и поворачивает назад. В квантовой механике в области I уравнение Шредингера и его решение остаются без изменений. В области II волновое число становится чисто мнимым, и решение принимает вид

$$k_2 = i\sqrt{2m(U_0 - E)/\hbar^2} = ik_2^*, \quad \psi_2(x) = a_2 e^{-k_2^* x}.$$

Это решение описывает волну, экспоненциально затухающую по мере ее распространения вглубь барьера. Условия непрерывности волновой функции и ее производной принимают вид

$$\begin{cases} a_1 + b_1 = a_2 \\ ik_1(a_1 - b_1) = -k_2^* a_2 \end{cases}.$$

Отношение амплитуд отраженной и падающей волн становится комплексным, правильное выражение для их интенсивностей есть квадрат модуля этого отношения

$$\frac{b_1}{a_1} = \frac{ik_1 + k_2^*}{ik_1 - k_2^*}, \quad R = \left| \frac{b_1}{a_1} \right|^2 = 1.$$

Таким образом, падающая на барьер частица в этом случае полностью отражается от барьера, как и в классической механике. Тем не менее, в квантовой механике частица может быть обнаружена и во второй области, что в классической механике невозможно. Плотность вероятности ее обнаружения экспоненциально убывает вглубь барьера

$$|\psi_2(x)|^2 = |a_2|^2 e^{-2k_2^* x}.$$

Можно провести аналогию с полным внутренним отражением при падении луча свет на поверхность под углом большим некоторого критического – свет отражается полностью, но поле во второй области ненулевое. Эта аналогия не является случайной, и в том и в другом случаях мы имеем дело с волновыми процессами.

Осталось отметить, что в пределе бесконечно высокого потенциального барьера ($U_0 \gg E$) частица, как и в классической механике, полностью отражается от границы барьера – величина $k_2^* \rightarrow \infty$, $|\psi_2(x)|^2 \rightarrow 0$ для $\forall x > 0$.

Прохождение частицы через барьер конечной ширины

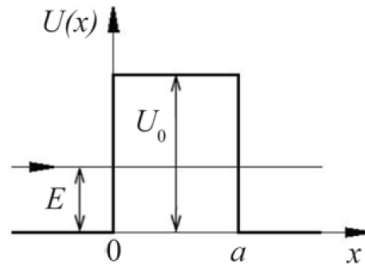


Рисунок 73. Потенциальный барьер конечной ширины.

Рассмотрим важную задачу, когда на пути частицы стоит барьер конечной ширины a (см. рис. 73). Распределение потенциала имеет вид

$$U(x) = \begin{cases} U_0, & 0 < x < a \\ 0, & x \leq 0 \text{ или } x \geq a \end{cases}$$

Возникают три различные области; исходя из физических соображений, в первых двух есть две волны, а в третьей – только одна, соответствующая прошедшей через барьер частице. Волновые функции для соседних областей следует гладко “сшить” между собой при $x = 0$ и $x = a$.

Рассмотрим случай, когда $E > U_0$. Сразу выпишем коэффициент прохождения частицы через барьер

$$T = \frac{4k^2 k_2^2}{(k^2 - k_2^2)^2 \sin^2(k_2 a) + 4k^2 k_2^2}, \quad k = \sqrt{2mE/\hbar^2}, \quad k_2 = \sqrt{2m(E - U_0)/\hbar^2}.$$

Этот коэффициент есть периодическая по E функция; при значениях энергии, удовлетворяющих условию $\sin(k_2 a) = 0$, он равен единице – нет отражения от барьера, барьер становится прозрачен для частиц. При $E \gg U_0$ коэффициент прохождения $T \rightarrow 1$.

При $E < U_0$ барьер абсолютно непрозрачен в классической механике. В квантовой же механике есть отличная от нуля вероятность частице преодолеть барьер

$$T = \frac{4k^2 k_2^{*2}}{(k^2 + k_2^{*2})^2 \operatorname{sh}^2(k_2^* a) + 4k^2 k_2^{*2}} > 0, \quad k_2^* = \sqrt{2m(U_0 - E)/\hbar^2}.$$

Такое явление называют *туннельным эффектом*. Естественно, вероятность этого прохождения экспоненциально убывает с увеличением ширины барьера и с увеличением его высоты. Для практически важных случаев малых вероятностей туннелирования $k_2^* a \geq 1$ можно записать

$$T \approx T_0 \exp\left(-\frac{2}{\hbar} \sqrt{2m(U_0 - E)} a\right).$$

T_0 – множитель, мало отличающийся от единицы. Эту формулу можно распространить на барьер любой формы:

$$T \approx T_0 \exp\left(-\frac{2}{\hbar} \int_{x_1}^{x_2} \sqrt{2m(U(x) - E)} dx\right),$$

x_1 и x_2 есть классические точки остановки, при которых $U(x) = E$.

Туннельный эффект проявляет себя при α -распаде радиоактивных частиц, при автоэлектронной и термоэлектронной эмиссии, в туннельном полупроводниковом диоде. В частности, на этом эффекте основана работа сканирующего туннельного микроскопа.

Частица в бесконечно глубокой потенциальной яме

Рассмотрим потенциальную яму шириной a , ограниченную с обеих сторон бесконечно высоким потенциальным барьером. Распределение потенциальной энергии при этом равно

$$U(x) = \begin{cases} 0, & 0 < x < a \\ \infty, & x \leq 0 \text{ или } x \geq a \end{cases}$$

Внутри ямы частица движется свободно, но выйти из нее не может. Внутри ямы уравнение Шредингера имеет вид

$$\frac{d^2}{dx^2} \psi(x) + \frac{2m}{\hbar^2} E \psi(x) = 0.$$

Исходя из вышеизложенного, граничные условия в данном случае записываются как

$$\psi(0) = \psi(a) = 0.$$

Общее решение уравнения Шредингера

$$\psi(x) = ae^{+ikx} + be^{-ikx}, \quad k = \sqrt{2mE/\hbar^2}.$$

Согласно первому из граничных условий, $a = -b$ и волновые функции стационарных состояний имеют вид

$$\psi(x) = A \sin(kx).$$

Второе условие показывает, что для этих волновых функция допустимы только определенные значения k , а именно

$$k = k_n = \frac{\pi}{a}n, \quad n = 1, 2, 3, \dots$$

Окончательно определяем волновые функции стационарных состояний

$$\psi_n(x) = A \sin(n\pi x/a),$$

значение коэффициента $A = \sqrt{2/a}$ определяется из условия нормировки. Значения энергии этих состояний определяются из выражения для k и равны

$$E_n = \frac{n^2 \hbar^2 \pi^2}{2ma^2}.$$

Следовательно, уровни энергии частицы в бесконечно глубокой потенциальной яме квантованы (или, как говорят, собственные состояния обладают дискретным спектром). Существует минимальная энергия – энергия основного состояния при $n = 1$, причем эта энергия $E_1 \neq 0$, что есть чисто квантовый эффект. Заметим, что при уменьшении a (улучшении локализации частицы) значение E_1 возрастает, что есть проявление принципа неопределенности.

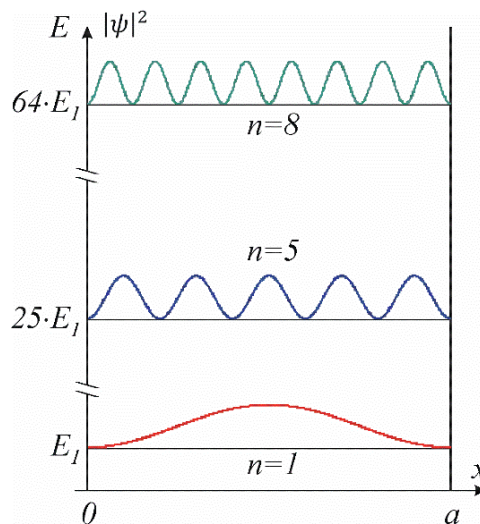


Рисунок 74. Уровни энергии и распределение плотности вероятности в потенциальной яме.

Уровни энергии и распределение $\psi_n^2(x)$ приведены на рисунке 74. Из графиков видно, что при увеличении n плотность вероятности состояний все равномернее распределена по всей ширине ямы. Это есть переход к классической механике, где частица в среднем проводит вблизи каждой точки пространства одинаковое время.

3.11 Квантовый линейный осциллятор

В физике очень часто встречается задача исследования движения вблизи положения устойчивого равновесия. В простейшем случае это одномерный линейный осциллятор – система, в какой-либо точке которой полная действующая на частицу сила равна нулю, по мере отхода от точки равновесия появляется линейно возрастающая сила, которая стремится вернуть частицу в эту точку. Потенциал, описывающий эту силу, имеет квадратичную зависимость от x :

$$U(x) = \frac{kx^2}{2} = \frac{m\omega^2 x^2}{2},$$

где $\omega = \sqrt{k/m}$ есть собственная частота осциллятора, найденная из классической механики, m – масса частицы.

В классической механике частица совершает периодическое движение между двумя точками, где потенциальная энергия становится равной ее полной энергии. В этих точках она полностью останавливается и отражается назад. В квантовой механике можно ожидать, что частица может находиться и дальше этих точек, но вероятность этого должна быстро падать по мере удаления из области $E - U(x) \geq 0$.

Как обычно, изучение квантового осциллятора начнем с записи уравнения Шредингера, в данном одномерном случае принимающего вид

$$\frac{d^2}{dx^2} \psi(x) + \frac{2m}{\hbar^2} \left(E - \frac{m\omega^2}{2} x^2 \right) \psi(x) = 0.$$

Введем безразмерные переменные

$$\eta = \frac{2E}{\hbar\omega}, \quad \xi = \frac{x}{x_0}, \quad x_0 = \sqrt{\frac{\hbar}{m\omega}},$$

в этих переменных исследуемое уравнение примет вид

$$\frac{d^2 \psi}{d\xi^2} + (\eta - \xi^2) \psi = 0.$$

Такое уравнение называется уравнением Чебышева-Эрмита. Мы знаем условия, которые накладываются на решение этого уравнения – квадрат функции ψ должен быть конечно интегрируемым, что бы выполнить условие нормировки, и должен быстро падать в области под барьером. Можно показать, что такие условия могут быть выполнены только при значениях параметра η удовлетворяющих условию

$$\eta = 2n + 1, \quad n = 0, 1, 2, \dots$$

Отсюда получаем, что энергетический спектр частицы в квантовом осцилляторе является дискретным:

$$E_n = \hbar\omega \left(n + \frac{1}{2} \right), \quad n = 0, 1, 2, \dots;$$

это есть общее правило для квантово-механических систем, когда движение частицы ограничено в пространстве. Интересно отметить, что в отличие от бесконечно глубокой плоской ямы, уровни энергии линейного осциллятора эквидистантны, то есть разница в энергии между двумя соседними уровнями постоянна и равна $\hbar\omega$. Однако, в обоих этих случаях, энергия основного состояния не равна нулю (иначе нарушился бы принцип неопределенности Гейзенберга), в рассматриваемом случае $E_0 = \hbar\omega/2$. Это заметным образом проявляется при изучении физики низких температур, поскольку объясняет нулевые колебания атомов конденсированной среды, устанавливающиеся после “вымораживания” нормальных тепловых колебаний кристаллической решётки.

Переходя с одного уровня энергии на другой, осциллятор испускает или поглощает фотон с энергией, равной разнице энергий этих уровней. Поскольку энергетические уровни линейного квантового осциллятора эквидистантны, может показаться, что возможно поглощение или испускание квантов света с энергией $n\hbar\omega$, $n \geq 1$. Однако выходящий за рамки нашего курса анализ показывает, что процессы с $n > 1$ запрещены. Подобные условия, определяющие правила изменения чисел, описывающих волновые функции стационарных состояний, при переходах между энергетическими уровнями называют *правилами отбора*.

Сами волновые функции – решения уравнения Шредингера – имеют вид

$$\psi_n = e^{-\frac{\xi^2}{2}} H_n(\xi), \quad n = 0, 1, 2, \dots,$$

где $H_n(\xi)$ – полиномы Чебышева-Эрмита порядка n . Приведем пару первых волновых функций:

$$\psi_0(x) = \frac{1}{\sqrt{x_0\sqrt{\pi}}} e^{-\frac{x^2}{2x_0^2}},$$

$$\psi_1(x) = \frac{1}{\sqrt{2x_0\sqrt{\pi}}} \frac{2x}{x_0} e^{-\frac{x^2}{2x_0^2}}.$$

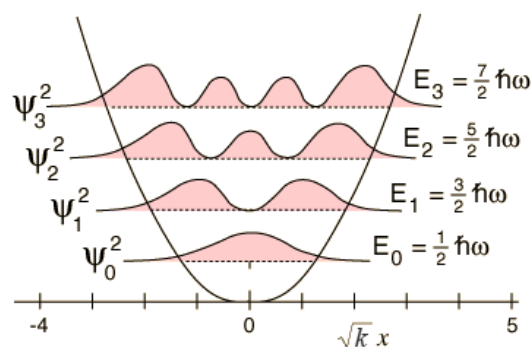


Рисунок 75. Уровни энергии и распределение плотности вероятности в квантовом осцилляторе.

Число нулей функции ψ_n совпадает с значением n . Уровни энергии в квантовом осцилляторе и распределение $\psi_n^2(x)$ приведены на рисунке 75.

При малых n плотности вероятности обнаружения частицы в квантовой и классической механике различаются кардинально, но с ростом n начинают быть похожими – см. рис. 76.

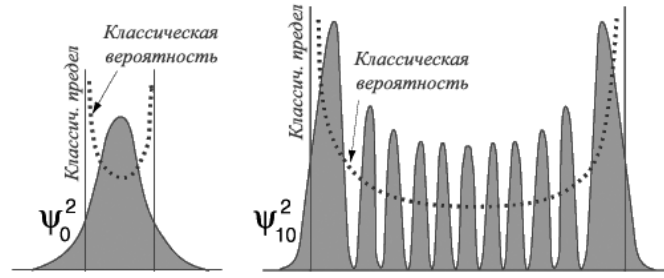


Рисунок 76. Сравнение классического и квантового распределения плотности вероятности.

3.12 Движение в поле центральной силы

Давайте рассмотрим движение в поле центральной силы на примере кулоновской силы. В любой точке пространства *центральная сила* направлена через определенную точку, называемую *центром силы*, а ее величина зависит только от расстояния до центра силы. В сферической системе координат, центр которой совмещен с центром силы, центральную силу можно представить в виде

$$\mathbf{F}(\mathbf{r}) = F(r) \frac{\mathbf{r}}{r}.$$

Центральная сила всегда является потенциальной, т.е. всегда можно ввести потенциальную энергию, причем эта потенциальная энергия будет зависеть только от модуля расстояния от точки наблюдения до центра силы

$$U(\mathbf{r}) = U(r).$$

Рассмотрим движение электрона ($Q = -e$) в кулоновском поле, создаваемом неподвижным зарядом $Q = +e$. При этом

$$U(\mathbf{r}) = -\frac{1}{4\pi\epsilon_0} \frac{e^2}{r}, \quad \mathbf{F}(\mathbf{r}) = -\text{grad } U(\mathbf{r}) = -\frac{1}{4\pi\epsilon_0} \frac{e^2}{r^2} \frac{\mathbf{r}}{r}.$$

Эта модель может быть использована при рассмотрении атома водорода, состоящем из положительно заряженного ядра и электрона. Ядром обычного водорода является протон, который в 1836 раз тяжелее электрона, для тяжелого водорода к протону добавляется еще один или два нейтрона, так что ядро можно считать неподвижным центром силы.

Начнем с записи уравнения Шредингера, для рассматриваемого случая

$$\Delta\psi(r, \theta, \varphi) + \frac{2m}{\hbar^2} (E - U(r)) \psi(r, \theta, \varphi) = 0.$$

В сферической системе координат оператор Лапласа приобретает вид

$$\Delta = \frac{1}{r^2} \left\{ \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{\sin^2(\theta)} \left[\sin(\theta) \frac{\partial}{\partial \theta} \left(\sin(\theta) \frac{\partial}{\partial \theta} \right) + \frac{\partial^2}{\partial \varphi^2} \right] \right\} =$$

$$= \frac{1}{r^2} \left\{ \Delta_r + \frac{1}{\sin^2(\theta)} [\Delta_\theta + \Delta_\varphi] \right\} = \frac{1}{r^2} \{ \Delta_r + \Delta_{\theta,\varphi} \}.$$

Оператор Δ_r действует только на координату r , оператор $\Delta_{\theta,\varphi}$ – только на угловые координаты θ и φ , аналогично для операторов Δ_θ и Δ_φ .

Будем решать это уравнение методом разделения переменных, а именно представим неизвестную волновую функцию в виде

$$\psi(r, \theta, \varphi) = \psi_r(r) \psi_{\theta,\varphi}(\theta, \varphi).$$

Подставляя такое представление в уравнение Шредингера, деля обе его части на ψ и умножая на r^2 получаем

$$\frac{\Delta_r \psi_r(r)}{\psi_r(r)} + \frac{2m}{\hbar^2} (E - U(r)) r^2 + \frac{\Delta_{\theta,\varphi} \psi_{\theta,\varphi}(\theta, \varphi)}{\psi_{\theta,\varphi}(\theta, \varphi)} = 0.$$

Первые два слагаемых получившегося выражения являются функциями только от координаты r , третье слагаемое зависит только от θ и φ . Поскольку r , θ и φ являются независимыми переменными, это возможно только в том случае, если обе получившиеся части равны не зависящей от этих переменных постоянной:

$$\frac{\Delta_r \psi_r(r)}{\psi_r(r)} + \frac{2m}{\hbar^2} (E - U(r)) r^2 = - \frac{\Delta_{\theta,\varphi} \psi_{\theta,\varphi}(\theta, \varphi)}{\psi_{\theta,\varphi}(\theta, \varphi)} = \lambda.$$

Таким образом, из исходного уравнения получили отдельно уравнение на радиальную часть волновой функции, отдельное – на угловую часть.

Функцию $\psi_{\theta,\varphi}(\theta, \varphi)$ так же попробуем представить в виде

$$\psi_{\theta,\varphi}(\theta, \varphi) = \psi_\theta(\theta) \psi_\varphi(\varphi).$$

Подставим это выражение в уравнение на угловую часть волновой функции, разделим обе его части на $\psi_{\theta,\varphi}$ и умножим на $\sin^2(\theta)$; в итоге получим следующее уравнение:

$$\frac{\Delta_\theta \psi_\theta(\theta)}{\psi_\theta(\theta)} + \lambda \sin^2(\theta) + \frac{\Delta_\varphi \psi_\varphi(\varphi)}{\psi_\varphi(\varphi)} = 0.$$

Повторяется аналогичная ситуация – первые два слагаемых зависят от одной независимой координаты, третья часть – от другой. Следовательно, обе получающиеся части равны постоянной, которую обозначим как m^2 :

$$\frac{\Delta_\theta \psi_\theta(\theta)}{\psi_\theta(\theta)} + \lambda \sin^2(\theta) = - \frac{\Delta_\varphi \psi_\varphi(\varphi)}{\psi_\varphi(\varphi)} = m^2.$$

Опять из одного уравнения получаем два: одно для ψ_θ , второе – для ψ_φ , т.е. для каждой из частей волновой функции определили соответствующее ей уравнение.

Начнем с анализа уравнения для ψ_φ , которое имеет вид

$$\frac{d^2 \psi_\varphi(\varphi)}{d\varphi^2} + m^2 \psi_\varphi(\varphi) = 0.$$

Решение этого уравнения есть функция

$$\psi_\varphi(\varphi) = \psi_{\varphi 0} e^{im\varphi}.$$

На это решение накладывается требование симметрии задачи – если мы изменим значение φ на величину 2π , то придем в ту же самую точку. Следовательно, одна из введенных ранее постоянных должна быть целым числом:

$$m = 0, \pm 1, \pm 2, \dots$$

Число m называют *магнитным квантовым числом*.

Уравнение на ψ_θ приобретает вид

$$\sin(\theta) \frac{d}{d\theta} \left(\sin(\theta) \frac{d\psi_\theta(\theta)}{d\theta} \right) + (\lambda \sin^2(\theta) - m^2) \psi_\theta(\theta) = 0.$$

Можно показать, что решение этого уравнения будет непрерывным и конечным при всех значениях угла θ , если только λ представимо в виде

$$\lambda = l(l+1), \quad l - \text{целое число}, \quad l \geq |m|.$$

Число l называют *орбитальным квантовым числом*. Исходя из физических соображений, которое будут рассмотрены далее, обычно фиксируют значение l , и для каждого такого значения магнитное квантовое число m может принимать $2l+1$ значения:

$$m = -l, -(l-1), \dots, -1, 0, 1, \dots, l-1, l.$$

Сами решения на ψ_θ уравнения выражаются через так называемые сферические, или шаровые функции, которые не будем здесь приводить.

Наконец, мы приходим к уравнению на ψ_r , которое принимает вид

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\psi_r(r)}{dr} \right) + \frac{2m}{\hbar^2} \left(E - (U(r) + U_{\text{цб}}(r)) \right) \psi_r(r) = 0, \quad U_{\text{цб}}(r) = \frac{\hbar^2 l(l+1)}{2mr^2}.$$

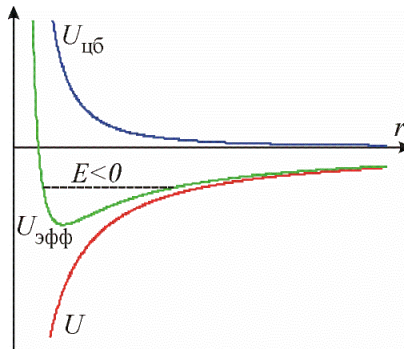


Рисунок 77. Эффективный потенциал.

Добавку $U_{\text{цб}}$ называют *центробежным потенциалом*, а сумму $U + U_{\text{цб}}$ — *эффективным потенциалом*. Для случая кулоновского поля распределение эффективного потенциала изображено на рисунке 77. Видно, что при $E < 0$ частица ограничена в своем движении по r . При больших r частицу ограничивает (не дает уйти на бесконечность) барьер кулоновской силы, при малых r частице не дает приблизиться к центру центробежный барьер. Как мы уже знаем, следует ожидать в этом случае дискретного спектра энергии стационарных состояний.

Действительно, можно показать, что для выполнения условия на конечность волновой функции энергия должна иметь вид

$$E = -\frac{1}{(4\pi\epsilon_0)^2} \frac{me^4}{2\hbar^2} \frac{1}{(n_r + l + 1)^2}, \quad n_r = 0, 1, 2, \dots$$

В этом выражении, кроме орбитального квантового числа l , фигурирует ещё одно квантовое число n_r . Однако, как мы видим, собственные значения энергии зависят только от суммы квантовых чисел $n_r + l + 1$. Сравнение с теорией Бора показывает, что естественно рассматривать эту сумму как одно число $n = n_r + l + 1$, которое называют *главным квантовым числом*; это число принимает значения $n = 1, 2, \dots$. Таким образом, мы из уравнения Шредингера получили спектр атома водорода, совпадающий с ранее полученным результатом, выведенным из постулатов Бора. Сами решения $\psi_r(r)$ имеют весьма сложный вид и приводить мы их не будем, отметим лишь, что они зависят от чисел n и l .

Заметим, что при фиксированном n орбитальное число может принимать значения в пределах $l = 0, \dots, n - 1$, т.е. всего n возможных значений. Магнитное число m также может принимать $2l + 1$ значений для заданного значения l . Итого получаем, что одну и ту же энергию, которая зависит лишь от главного квантового числа n , имеют n^2 различных состояний (с разными квантовыми числами l и m). Говорят, что эти состояния *вырождены* с кратностью вырождения n^2 . Частично это вырождение можно убрать, если поместить атом водорода во внешнее электрическое или магнитное поле. Наблюдаемое при этом в спектре расщепление линий называют эффектами Штарка и Зеемана соответственно.

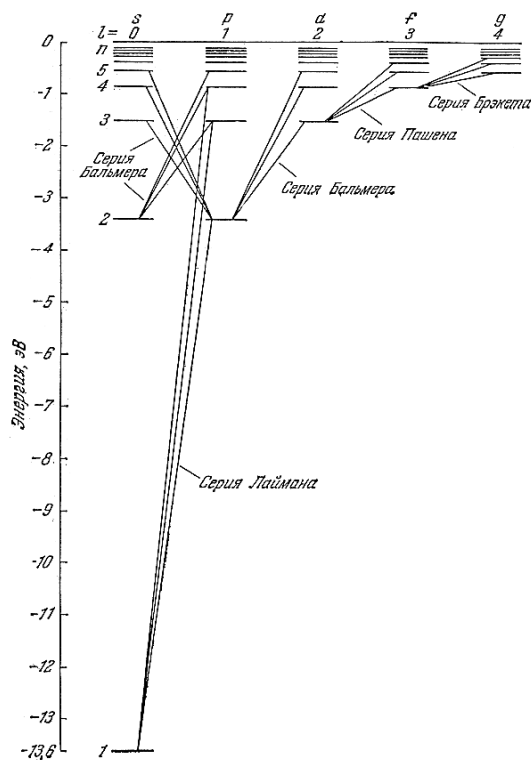


Рисунок 78. Уровни энергии в атоме водорода и переходы между ними.

Таблица уровней энергии в атоме водорода в зависимости от чисел n и l и переходы между ними представлены на рисунке 78. Строчные буквы латинского алфавита используются для обозначения состояния с определенным значением l . При переходе с одного уровня энергии на другой атом испускает или поглощает кванты света. Оказывается, что правилами отбора разрешены только те переходы, при которых $\Delta l = \pm 1$ (точнее, вероятность таких переходов на много порядков выше, чем вероятности других переходов). Отсюда получаем, что спектр водорода характеризуется определенными закономерностями - сериями, что и помогло Бору создать свою теорию.

3.13 Оператор момента импульса

Введем оператор момента импульса

$$\hat{\mathbf{I}} = [\hat{\mathbf{r}} \times \hat{\mathbf{p}}] = \hat{l}_x \mathbf{e}_x + \hat{l}_y \mathbf{e}_y + \hat{l}_z \mathbf{e}_z,$$

где \hat{l}_i – операторы соответствующих проекций момента импульса, например

$$\hat{l}_x = i\hbar \left(z \frac{\partial}{\partial y} - y \frac{\partial}{\partial z} \right).$$

Можно показать, что все собственные функции и собственные значения оператора $\hat{\mathbf{I}}$ равны нулю, а операторы различных проекций не коммутируют между собой. Следовательно, нет такого состояния, в котором бы все три или даже какие-либо две из трех проекций момента импульса имели бы определенное значение (могли бы быть точно измерены одновременно). Иными словами, для любой квантово-механической системы не существует состояния, в котором бы вектор момента импульса имел определенное значение, т. е. был бы полностью определен как по величине, так и по направлению.

В квантовой механике момент импульса характеризуется другими физическими величинами, одной из которых является проекция момента импульса на какое-то одно направление. Если перейти в сферическую систему координат и направить ось этой системы, задаваемую условием $\theta = 0$, вдоль интересующего нас направления, то оператор соответствующей проекции примет вид

$$\hat{l}_z = -i\hbar \frac{\partial}{\partial \varphi}.$$

Для ответа на вопрос, может ли проекция углового момента иметь определенное значение, необходимо рассмотреть задачу на собственные значения этого оператора

$$\hat{l}_z \psi = -i\hbar \frac{\partial \psi}{\partial \varphi} = l_z \psi.$$

Решением такого уравнения является

$$\psi = C(r, \theta) e^{i \frac{l_z}{\hbar} \varphi},$$

где функция $C(r, \theta)$ – произвольная функция от этих координат. Мы уже сталкивались с условием, налагаемым на зависимость волновой функции от координаты φ – при увеличении φ на 2π функция φ должна переходить сама в себя, т.е. быть по φ периодической с периодом 2π . Это условие определяет дискретные значения проекции момента импульса на произвольно выделенное направление

$$l_z = m\hbar, \quad m = 0, \pm 1, \pm 2, \dots$$

Второй величиной, характеризующей момент импульса квантово-механической системы, является квадрат полного момента импульса, соответствующий оператор имеет вид

$$\hat{\mathbf{I}}^2 = \hat{l}_x^2 + \hat{l}_y^2 + \hat{l}_z^2.$$

Этот оператор коммутирует с оператором \hat{l}_z , следовательно, существует состояние, в котором одновременно имеют определенные значения квадрат момента импульса и одна из его проекций на избранное направление. Можно показать, что собственные значения \mathbf{I}^2 оператора $\hat{\mathbf{I}}^2$ имеют вид

$$\mathbf{I}^2 = \hbar^2 l(l+1), \quad l = 0, 1, 2, \dots$$

Оказывается, что между собственными значениями этих двух операторов есть следующая связь:

$$\mathbf{I}^2 > l_z^2,$$

отсюда следует, что

$$|m| < \sqrt{l(l+1)}.$$

При определенном значении l всего возможно $2l+1$ значений m :

$$m = -l, -(l-1), \dots, -1, 0, 1, \dots, l-1, l.$$

Эта дискретность возможных значений проекции и квадрата момента импульса квантово-механической системы называется *пространственным квантованием*. Если сравнить полученные результаты с проведенным выше рассмотрением атома водорода, то становится очевидным, что орбитальное число определяет квадрат момента импульса электрона, а магнитное – величину его проекции.

3.14 Спин

Рассмотрим задачу из классической физики. Пусть частица массой m и зарядом q движется по окружности. С одной стороны, эта частица обладает определенным моментом импульса \mathbf{l} :

$$\mathbf{l} = \mathbf{r} \times \mathbf{p} = m [\mathbf{r} \times \mathbf{v}].$$

С другой стороны, эта частица, как и любой движущийся заряд, создает в окружающем пространстве магнитное поле. При достаточно быстром движении частицу можно представить как циркулирующий по окружности ток, магнитное

поле в этом случае выражается из закона Био-Савара-Лапласа. Аналогичным образом можно поступить, если есть ограниченная в пространстве система, состоящая из движущихся частиц (которые также представляются замкнутыми токами). Прямые вычисления магнитного поля достаточно сложны, однако на больших расстояниях от области, занимаемой движущимися частицами, выражение для поля можно значительно упростить. Здесь можно провести полную аналогию с нейтральной в целом системой заряженных частиц, электрическое поле от которых на большом расстоянии определяется вектором дипольного момента \mathbf{p} :

$$\mathbf{p} = \sum q_i \mathbf{r}_i, \quad \mathbf{E}(\mathbf{r}) = -\frac{\mathbf{p}}{r^3} + 3\mathbf{r} \frac{\mathbf{r} \cdot \mathbf{p}}{r^5},$$

где q_i – величины входящих в систему точечных зарядов, \mathbf{r}_i – радиус-вектора их положения. Аналогичным образом можно ввести вектор магнитного момента \mathbf{m} , который для системы токов равен

$$\mathbf{m} = \frac{1}{2c} \int [\mathbf{r}' \times \mathbf{j}(\mathbf{r}')] dV', \quad \mathbf{H}(\mathbf{r}) = -\frac{\mathbf{m}}{r^3} + 3\mathbf{r} \frac{\mathbf{r} \cdot \mathbf{m}}{r^5}.$$

Для движущейся по окружности радиуса \mathbf{r} частице

$$\mathbf{m} = \frac{q}{2c} [\mathbf{r} \times \mathbf{v}] = \frac{q}{2mc} \mathbf{l}.$$

Таким образом, магнитный момент заряженной частицы пропорционален ее моменту импульса, коэффициент пропорциональности

$$g = \frac{q}{2mc}$$

называется *гиромагнитным отношением*.

Магнитный момент системы токов не только определяет ее собственное магнитное поле, но и показывает, как эта система взаимодействует с внешним магнитным полем. Можно показать, что если магнитное поле является неоднородным (достаточно сильно изменяется в пространстве), то с его стороны на систему как на единое целое действует сила

$$\mathbf{F} = \text{grad}(\mathbf{m} \cdot \mathbf{H}).$$

Таким образом, можно внешним полем воздействовать на в целом нейтральные объекты, такие, как атомы.

В квантовой механике оператор магнитного момента можно определить из оператора момента импульса через гиромагнитное отношение

$$\hat{\mathbf{m}} = g\hat{\mathbf{l}}.$$

При этом свойства оператора магнитного момента оказываются аналогичны свойствам момента импульса, в частности, квантуется проекция магнитного момента на какое-либо выбранное направление:

$$m_z = gl_z = \hbar g m = m_B m, \quad m = 0, \pm 1, \pm 2, \dots,$$

где постоянная m_B носит название *магнетона Бора*.

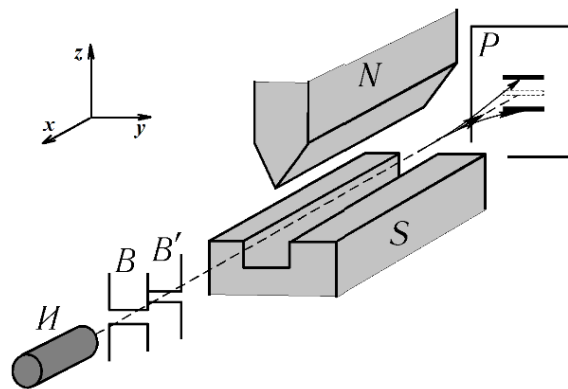


Рисунок 79. Схема опыта Штерна и Герлаха.

Немецкие ученые Штерн и Герлах провели прямой опыт по обнаружению пространственного квантования. Пучок атомов пропусклся вдоль сильно неоднородного магнитного поля, образованного постоянными магнитами, как показано на рисунке 79. Если у атома есть магнитный момент, то со стороны поля на него будет действовать сила

$$F_z = m_z \frac{\partial H_z}{\partial z}.$$

Атомы, обладающие разными проекциями магнитного момента, в результате разойдутся и попадут в разные точки экрана. В классической механике вектор магнитного момента ориентирован произвольно, m_z может принимать всевозможные непрерывно распределенные значения, следовательно, включение магнитного поля приводит к уширению пучка вдоль z . Согласно поведенному рассмотрению, в квантовой механике пучок должен расщепиться на несколько пучков, поскольку m_z может принимать лишь несколько значений. Если орбитальное квантовое число атома равно l , то по теории при расщеплении получится $2l + 1$ пучков – именно столько возможных значений может принимать m_z . Таким образом, ожидаемое число составляющих должно было бы быть всегда нечетным.

Опыт Штерна и Герлаха доказал квантование проекции m_z . Однако их результаты не всегда соответствовали теории, изложенной выше – они получали расщепление как на нечетное, так и на четное число пучков, в частности, для атомов серебра и водорода получалось расщепление на две составляющие.

С учетом результатов этого и других опытов пришли к следующему: У каждой элементарной частицы есть собственный момент импульса, имеющий квантовую природу и не связанный с перемещением частицы как целого. Этот момент импульса называется *спином*. Несмотря на то, что спин не связан с реальным вращением частицы, тем не менее он порождает магнитный момент, а значит, приводит к дополнительному (по сравнению с классической электродинамикой) взаимодействию с магнитным полем.

Спину соответствует свой оператор $\hat{\mathbf{S}}$, который аналогичен оператору $\hat{\mathbf{I}}$. Следовательно, определенные значения в одном и том же состоянии могут иметь квадрат полного спина \mathbf{s}^2 и проекция спина на определенную ось s_z . Если максимальное значение проекции спина s_z равно $\hbar s$, то всего его проекция может принимать $2s + 1$ значений. Это число s и принимается за значение спина частицы.

Опыты Штерна и Герлаха показали, что для электрона число проекций равно 2, откуда $s = 1/2$. Кроме электрона, полуцелый спин имеют протон, нейтрон, мюон, кварки, нейтрино и т.д. Спин $s = 0$ имеют π и K мезоны, бозон Хиггса. Спином $s = 1$ обладают фотоны, W и Z бозоны (переносчики слабого взаимодействия).

3.15 Заполнение электронных оболочек атомов

Атом водорода – единственный из атомов, который может быть точно рассчитан аналитически. Все остальные атомы, начиная с атома гелия, рассчитываются приближенно, методами теории возмущений.

Действительно, кроме взаимодействия каждого электрона с центральным полем ядра, электроны в многозарядных атомах взаимодействуют между собой. Состояние электронной оболочки должно было бы характеризоваться волновой функцией всех координат ее электронов, однако точное нахождение волновой функции системы многих взаимодействующих частиц недоступно методам современной математики. Поэтому приходится вводить приближенное представление о состояниях отдельных электронов – предполагается, что в основном электроны движутся независимо в поле ядра. Волновые функции электрона в атоме водорода мы характеризовали, используя квантовые числа n , l , m (сюда же необходимо добавить квантовое число s_z , учитывающее ориентацию спина). В принимаемом нами приближении в многозарядных атомах волновые функции отдельных электронов так же характеризуются этими квантовыми числами. Взаимодействие электронов между собой рассматривается как возмущение, т.е. малая добавка к этой точно решенной задаче.

Наиболее заметным проявлением этого возмущения является *расщепление* энергетических уровней. В атоме водорода энергия состояния электрона зависит только от числа n , все состояния с этим главным числом, но разными числами l имеют одинаковую энергию (всего таких состояний n). В многозарядных же атомах энергия состояния определяется так же и числом l , хотя эта зависимость обычно гораздо слабее, т.е. одна линия в спектре расщепляется на несколько близлежащих линий. Возможно даже перекрытие линий – энергия состояния с одним n окажется больше, чем энергия состояния с $n + 1$. Состоянию электрона с наименьшей энергией всегда соответствует $n = 1$ (орбитальному и магнитному числам при этом возможно принимать только одно значение $l = m = 0$).

Рассмотрим какой-либо многозарядный атом. Его электронная оболочка содержит электроны, число которых равно, как вы отлично знаете, заряду ядра

(естественно, в единицах заряда электрона). Также хорошо известно, что любая система стремится занять состояние с наименьшей энергией. Вроде бы, отсюда следует, что все электроны в атоме должны находиться в состоянии с $n = 1$, и только при возбуждении, скажем, при поглощении фотона один из электронов перейдет на более высокий энергетический уровень, с $n > 1$. Мы хорошо знаем, что это не так, но почему?

Принцип Паули

В отличие от макроскопических тел, однотипные объекты микромира (электроны, протоны и т.д.) совершенно одинаковы. Как же отличить одну частицу от другой? В классической механике мы принципиально можем проследить за траекторией каждой частицы из первоначального набора нескольких частиц, независимо одинаковые они или нет. А вот в квантовой механике понятия траектории нет, есть волновая функция, имеющая вероятностное истолкование. Если в начальный момент было несколько электронов, каждый из которых был локализован в какой-то области, то спустя какое-то время, обнаружив где-нибудь один из электронов, мы не сможем сказать, какой именно электрон был обнаружен. Одинаковые частицы принципиально неотличимы, и поэтому необходимо говорить о состоянии системы одинаковых частиц, а не о состоянии каждой из частиц в отдельности. Это положение можно формулировать в виде принципа тождественности одинаковых частиц: в системе одинаковых частиц реализуются только такие состояния, которые не меняются при перестановке местами двух любых частиц.

Давайте рассмотрим волновую функцию системы двух одинаковых частиц $\psi(q_1, q_2)$, q_1 – координаты первой частицы, q_2 – второй. Поменяем местами эти две частицы; эту операцию будем описывать как действие оператора перестановки на волновую функцию:

$$\psi(q_2, q_1) = \hat{P}\psi(q_1, q_2).$$

Применив оператор перестановки вторично, получим исходную функцию:

$$\psi(q_1, q_2) = \hat{P}\psi(q_2, q_1) = \hat{P}^2\psi(q_1, q_2), \quad \hat{P}^2 = 1, \quad \hat{P} = \pm 1.$$

Отсюда можно заключить, что для системы, состоящей из двух (и более) тождественных частиц допустимы два типа волновых функций – которые меняют свой знак при перестановке двух частиц, и которые при такой операции знак не меняют:

$$\begin{aligned}\psi_a(q_1, q_2) &= -\psi_a(q_2, q_1), \\ \psi_s(q_1, q_2) &= \psi_s(q_2, q_1).\end{aligned}$$

Первая из таких волновых функций называется *антисимметричной*, вторая – *симметричной*. Тип симметрии оказывает сильное влияние на поведение коллектива частиц. Свойства больших коллективов частиц исследует статистическая физика, которая и определяет распределение тождественных частиц по энергетическим уровням.

Оказывается, что тип симметрии волновой функции системы частиц зависит от их спина. Для системы частиц, обладающих полуцелым спином – электроны, протоны, кварки – волновая функция является антисимметричной. Большой коллектив таких частиц подчиняется статистике Ферми-Дирака, а сами частицы носят название *фермионы*. Для системы частиц, обладающих целым спином – фотоны, π и K мезоны – волновая функция симметрична. Таким частицам соответствует статистика Бозе-Эйнштейна, а сами частицы называются *бозонами*.

Для системы фермионов справедлив *принцип Паули* – в системе тождественных фермионов не может быть двух частиц, находящихся в одном и том же состоянии.

Возвращаясь к рассмотрению заполнения электронных оболочек многозарядных атомов, принцип Паули утверждает, что состояния электронов в ней должно различаться хотя бы одним из квантовых чисел. Рассмотрим, сколько электронов может находиться в состоянии с $n = 1$. При этом возможна только одна комбинация чисел l и m , следовательно, электроны могут различаться только величиной проекции спина $s_z = \pm \frac{1}{2}$, на основном уровне могут находиться всего два электрона, все остальные электроны обязаны иметь $n > 1$.

Совокупность электронов атома с заданным значением n образует *слой*. Иногда слой обозначают прописными буквами латинского алфавита, но чаще пользуются просто номером слоя.

n	1	2	3
Обозначение	K	L	M
Макс. число электронов	2	8	18

Совокупность электронов с заданными значениями n и l образует *оболочку*. Для обозначения оболочек используют строчные буквы латинского алфавита.

l	0	1	2
Обозначение	s	p	d
Макс. число электронов	2	6	10

Конфигурация полной электронной оболочки атома обозначается выражением вроде следующего:

$$2s^2 2p^5.$$

Расшифруем это выражение: $2s^2$ значит, что на втором (L) слое на s оболочке находятся два электрона – она полностью заполнена; $2p^5$ значит, что на втором слое на p оболочке находится пять электронов; подразумевается, что не обозначенные нижележащие слои (в данном случае первый слой) полностью заполнены. Разобранное выражение есть конфигурация электронной оболочки атома фтора, его атомный номер равен 9.