

Предисловие

Идея разработки и издания учебного пособия по ближнепольной оптике принадлежит профессору М.Н.Либенсону. Он одним из первых в России оценил важность нового раздела физики, возникшего в начале 80-х годов и быстро завоевавшего признание. По его инициативе работы по ближнепольной оптике были начаты в ГОИ им. С.И.Вавилова и СПб ГУИТМО (Санкт-Петербург), а затем и в ряде других научных центров. Профессор М.Н.Либенсон любил цитировать выдержки из доклада, представленного на одном из международных конгрессов, в котором появление ближнепольной оптики приравнивалось по значимости к изобретению лазеров и голографии, вместе составивших три этапа научно-технической революции в оптике XX столетия. Он всячески способствовал привлечению к работе студентов и молодых специалистов, способных в будущем реализовать на практике уникальные возможности, открывающиеся при внедрении методов ближнепольной оптики в технику записи и считывания информации, нанолитографию и другие перспективные технологии.

Профессор М.Н.Либенсон наметил план курса и содержание ряда разделов, но не успел довести работу до конца. Мы посвящаем нашу часть совместного труда памяти Михаила Наумовича.

Г.С.Жданов,
Г.А.Марциновский

1. Введение. История возникновения и развития ближнепольной оптики

Стремительное развитие вычислительной техники в конце XX столетия стало возможным благодаря резкому повышению плотности компоновки элементов электронных схем. На смену микротехнологиям пришли нанотехнологии, потребовавшие разработки новых методов анализа поверхности твердых тел.

Среди приборов, предназначенных для исследования поверхности с высоким пространственным разрешением, особое место принадлежит отряду сканирующих зондовых микроскопов (СЗМ). Их первым представителем стал сканирующий туннельный микроскоп (СТМ), принесший его создателям Нобелевскую премию вскоре после опубликования первых результатов. Не меньшую известность приобрел и изобретенный несколькими годами позднее атомно-силовой микроскоп (АСМ), решивший проблему исследования поверхностного рельефа диэлектриков. Оба прибора позволяют не только изучать поверхность, но и модифицировать ее, например, наносить рисунок по заданной программе.

Чрезвычайно заманчива идея сочетать в одном приборе высокое пространственное разрешение СЗМ с многочисленными достоинствами оптических методов: быстроедействие, возможностью перестройки в широких пределах частоты и мощности излучения, отсутствием жестких требований к образцу и среде и т.д. На первый взгляд, эта идея неосуществима, так как согласно установившимся представлениям пространственное разрешение оптических приборов ограничено дифракцией и не может существенно отличаться от длины полуволны света $\lambda/2$. Однако классические законы дифракции, как и других явлений, связанных с волновой природой света, основаны на неявном допущении, что элементы оптической схемы, например, линза объектива и образец, находятся на значительном (по сравнению с λ) расстоянии друг от друга. Если миниатюрный зонд - источник или приемник света - перемещать на расстоянии $z \ll \lambda$ от поверхности образца, то можно рассчитывать на реализацию разрешения, не ограниченного дифракцией, или "сверхразрешения".

На эту возможность впервые обратили внимание в 1928 г. Статья, получившая одобрение Эйнштейна, была опубликована в одном из ведущих физических журналов [1]. К сожалению, идеи, изложенные в этой статье, намного опередили технические возможности того времени и вскоре были забыты. К ним вернулись лишь через десятки лет, ограничившись на первом этапе опытами с микроволнами. В 1984 г. группа ученых из лаборатории IBM в Цюрихе (в которой был создан и первый СТМ) реализовала разрешение

$\lambda/20$ в опытах со светом видимого диапазона [2]. Близкие результаты примерно в то же время были получены и в других исследовательских центрах. Новый прибор получил название “ближнепольный сканирующий оптический микроскоп” (БСОМ) (в иностранной литературе употребляют аббревиатуры NSOM и SNOM).

Первые приборы были технически весьма несовершенными и полученные результаты воспринимались скорее как интересный физический феномен, чем как ценные сведения о поверхности образца. Перелом произошел через несколько лет, когда удалось решить проблему прецизионного контроля ширины зазора острие-образец и существенно улучшить технологию изготовления оптических зондов. Затем фронт работ стал быстро расширяться. Уникальные возможности БСОМ были продемонстрированы в различных областях, включая физику твердого тела, технику записи и считывания информации, биологию и др. [3]. С помощью БСОМ на магнитооптических средах была достигнута плотность записи на порядок превышающая возможности стандартных оптических методов [4]. При сочетании технических приемов СТМ и БСОМ удалось уменьшить размер деталей формируемого на поверхности рельефа до 10 нм [5]. Большой интерес вызвали сообщения о возможности наблюдения отдельных молекул и регистрации их спектров [6].

Эти и многие другие экспериментальные и теоретические работы заложили основы новой области физики – ближнепольной оптики (БПО), объединившей микроскопию, методы анализа и средства модификации поверхности с субволновым (существенно меньшим длины волны) разрешением. Иногда используют другое название - “нанооптика”, подчеркивая тем самым возможность сжатия электромагнитного поля, созданного протяженным источником, в области нанометровых размеров. Исследования в этом направлении ведутся в десятках лабораторий, регулярно проводятся международные конференции, ряд фирм осуществляет серийный выпуск БСОМ. Круг научных и прикладных проблем, связанных с БПО и решаемых ее методами, весьма широк и их число непрерывно растет.

2. Принцип действия и основные узлы БСОМ

Термин “БСОМ” используют по отношению ко всем приборам БПО независимо от того, для решения каких задач они предназначены. К таким задачам относятся обнаружение малых неоднородностей в приповерхностном слое образца, их спектроскопический анализ, модификация рельефа поверхности оптическими методами и т.д.

Важнейшим элементом любого БСОМ является оптический зонд, локализирующий электромагнитное поле в области протяженностью существенно меньшей длины волны. Простейшим зондом является заостренное волокно, на конический отрезок которого сбоку нанесен металлический слой так, чтобы прозрачной осталась только небольшая площадка на вершине.

Зонд может переносить излучение от осветителя к образцу или от образца к детектору. В первом варианте луч лазера запускается в волокно через широкий торец и, проходя через коническую вершину, где он теряет основную часть мощности, освещает образец. За прозрачным образцом расположен детектор, например, фотодиод. Вторым вариантом отличается тем, что осветитель и детектор меняются местами. В этом случае освещен широкий участок образца, но через зонд проходит лишь излучение, рассеянное участком поверхности, ближайшим к вершине. Если образец неоднороден, то интенсивность рассеяния, а значит и ток, регистрируемый приемной схемой, изменяются в процессе сканирования, что позволяет связать положение неоднородности с тем или иным элементом раstra на изображении.

Прецизионное перемещение зонда относительно образца осуществляется с помощью сканера. Обычно он представляет собой отрезок цилиндрической трубки из пьезокерамики, на который нанесено четыре электрода. Подавая в определенной последовательности напряжение на разные пары электродов, можно изгибать цилиндр в разные стороны, а также сжимать или растягивать его вдоль оси, осуществляя тем самым развертку в горизонтальной плоскости $x-y$ и подвижку по оси z .

К числу необходимых узлов БСОМ относится сенсорная схема, реагирующая на изменение расстояния между зондом и образцом. Чувствительным элементом чаще всего служит небольшая кварцевая пластинка в форме вилки с резонансной частотой, составляющей несколько десятков кГц. Вилку с приклеенным к одной из ножек острием, закрепляют на сканере. Под действием переменного напряжения, настроенного в резонанс с частотой колебаний кварца, вилка колеблется, подобно маятнику, в плоскости, перпендикулярной поверхности образца.

Сближение острия с образцом приводит к уменьшению размаха колебаний под действием тангенциальной составляющей сил атомного взаимодействия, что в свою очередь вызывает изменение пьезонапряжения, снимаемого между ножками вилки. Качественно ту же картину можно наблюдать, постепенно погружая в жидкость колеблющуюся иглу.

Силы, тормозящие колебания острия, направлены преимущественно вдоль поверхности образца, как и силы сдвига в кристалле (shear force).

Исходя из этой аналогии, термин “shear force” стали применять и для обозначения метода контроля расстояния между острием и образцом по изменению амплитуды и фазы колебаний острия, происходящих вдоль поверхности образца. Зная зависимость амплитуды колебаний от расстояния z до поверхности, можно остановить сближаемое с образцом острие на нужной высоте (обычно 1...10 нм), не допуская его столкновения с поверхностью.

Существуют и чисто оптические схемы контроля z , в которых колеблющаяся игла помещается между вспомогательными источником и приемником света. Их взаимное расположение при юстировке подбирается таким образом, чтобы достигающая детектора часть светового потока уменьшалась при росте амплитуды колебаний. Лучшего, чем в режиме “shear force” разрешения можно добиться в режиме “tapping”, когда острие периодически вплотную приближается к поверхности и удаляется от нее.

При движении острия вдоль шероховатой поверхности система обратной связи меняет напряжение на сканере каждый раз, когда амплитуда колебаний отклоняется от заранее выбранного значения. Так как скачок напряжения пропорционален изменению ширины зазора, по его величине можно судить о высоте встреченной острием неоднородности рельефа. Соответствующая информация обрабатывается в канале АСМ и может быть представлена как карта топографического рельефа сканируемого участка поверхности. Одновременно в оптическом канале или канале БСОМ отслеживается изменение интенсивности светового потока, приходящего на детектор.

Холмики или углубления на поверхности регистрируются одновременно в обоих каналах. Более интересна ситуация, когда инородное включение находится на небольшой глубине и может быть замечено только благодаря своим особым оптическим свойствам. Пусть, например, в образце содержатся частицы флуорофора, переизлучающие падающий свет с изменением частоты (рис. 1а). Возможность локализации флуоресцентных меток весьма существенна при исследовании биологических тканей, что связано со способностью различающихся по составу меток осаждаться избирательно на различных участках клеток. Запись изображения в канале АСМ позволяет обнаружить выступ на поверхности (рис. 1б), но не дает информации о том, отличается ли он по своим оптическим свойствам от соседних участков. Канал БСОМ, принимающий сигнал только на частоте флуоресценции, устраняет эту неопределенность. Более того, БСОМ, в отличие от АСМ, способен обнаружить и частицы, находящиеся на небольшой глубине под поверхностью (рис. 1в). Чтобы реализовать эти дополнительные возможности, необходимо использование сенсорных схем

АСМ (или СТМ при исследовании проводящих образцов), так как иначе было бы трудно правильно установить острие и избежать его повреждения при столкновении с поверхностью в процессе сканирования.

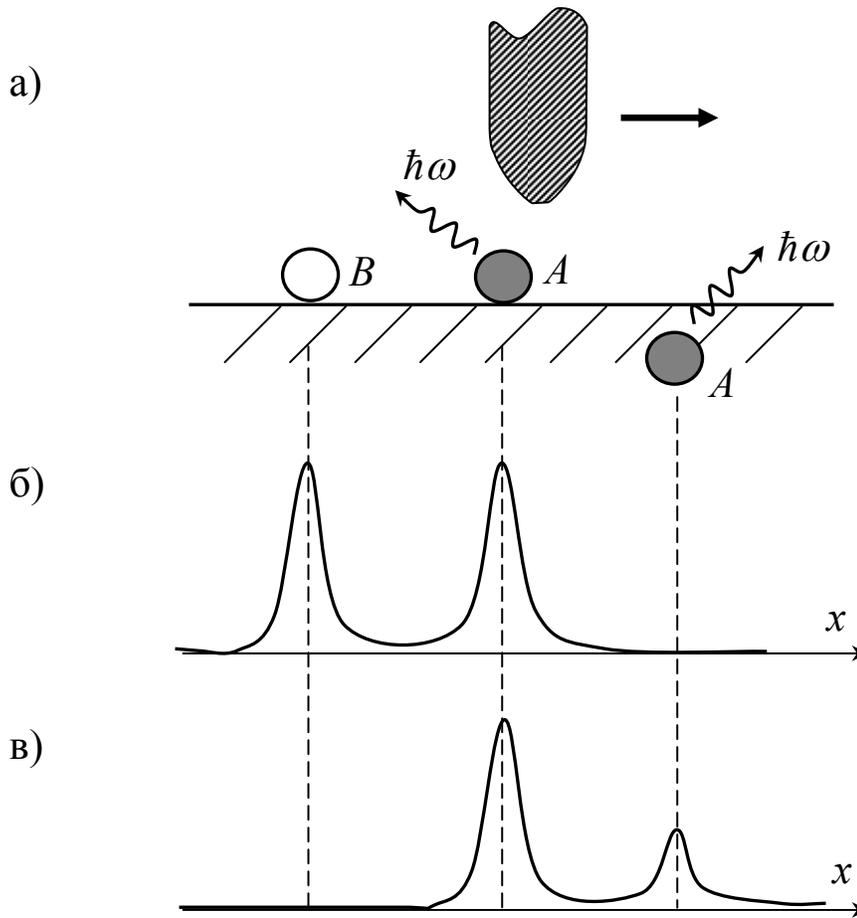


Рис.1 а) Схема сканирования зондом образца, содержащего флуоресцентные метки (А) и оптически нейтральную частицу (В); б, в) Изображения выделенного сечения образца в каналах АСМ (б) и БСОМ (в); кривая записи оптического сигнала приведена с поправкой на изменение положения острия по высоте.

Отличаясь от других зондовых микроскопов рядом характерных особенностей, многочисленные разновидности БСОМ могут существенно различаться и между собой. Известно около 20 вариантов БСОМ, предназначенных для исследования образцов в прошедшем или отраженном свете, использующих оптический зонд для освещения образца или сбора фотонов, имеющих общие узлы с СТМ, АСМ, интерферометрами и т.д. Не меньше и разнообразие оптических зондов. На первом этапе использовались в основном зонды, изготавливаемые из прозрачных материалов, например, уже упоминавшееся заостренное оптическое волокно, конец которого защищен металлическим экраном с малым отверстием (диафрагмой) на вершине (рис. 2а). Такие зонды, как и приборы, в которых они используются, получили название “апертурные” (от aperture – диафрагма). Типичный радиус диафрагмы составляет 50 нм, но в зависимости от решаемых прибором задач может варьироваться в диапазоне 10...150 нм.

Безапертурные (nonaperture) зонды чаще всего изготавливают из непрозрачных материалов. Методы химического и ионного травления позволяют получать острия с радиусом кривизны до 1...1,5 нм. В силу ряда причин, рассмотренных в разделе 3.10, электромагнитное поле концентрируется у вершины металлического острия, что приводит фактически к тому же эффекту, что и сжатие светового пучка диафрагмой в непрозрачном экране (рис. 2б).

Сравнивая типичные размеры зондов, нетрудно разделить сферы наиболее эффективного применения апертурных и безапертурных приборов. Первые предпочтительнее в тех случаях, когда необходима значительная интенсивность локального светового воздействия на образец, вторые – когда основные требования предъявляются к предельному разрешению. Необходимо также иметь в виду, что при использовании безапертурных зондов рассеянная ими мощность представляет лишь малую часть мощности излучения первичного пучка, попадающей непосредственно на образец, что создает дополнительные трудности при выделении полезного сигнала.

Основные узлы БСОМ, рассмотренные выше, можно найти на изображенной на рис. 3 схеме отражательного БСОМ с зондом в виде изогнутой вольфрамовой иглы, совершающей колебания вдоль нормали к поверхности образца [7]. Такая форма сближает зонд по его функциональным возможностям с рычагом АСМ и повышает чувствительность прибора к мельчайшим неоднородностям рельефа.

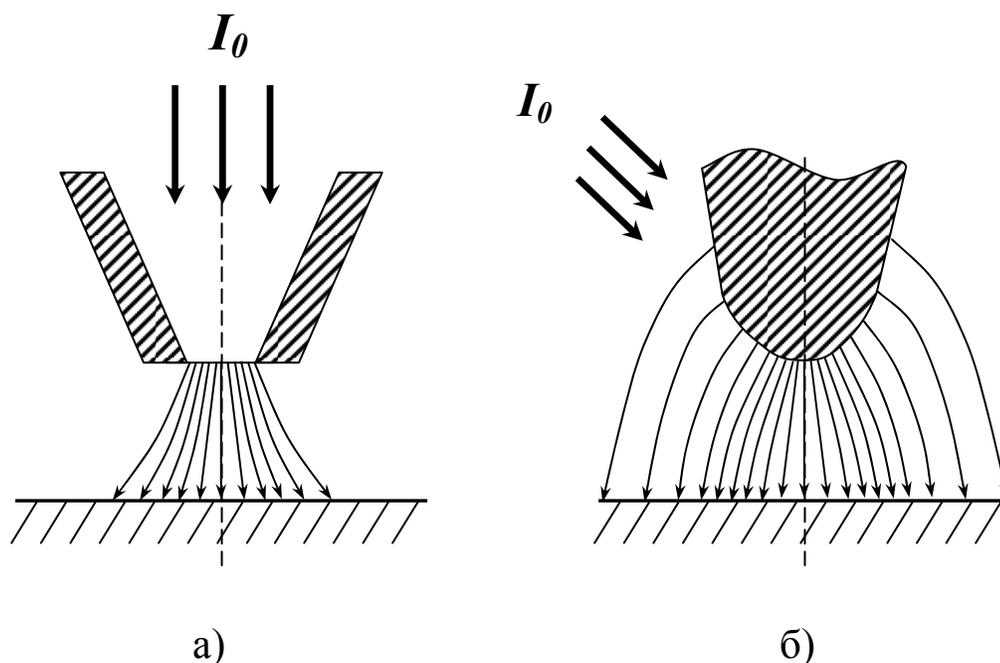


Рис.2 Схема рассеяния электромагнитной энергии вершинами апертурного а) и безапертурного б) зондов.

Во многих случаях систему прецизионного перемещения зонда и образца БСОМ устанавливают на столике обычного оптического микроскопа, что позволяет использовать весь арсенал методов классической световой микроскопии, дополняя их изображениями выделенного участка образца с высоким пространственным разрешением.

Само название БСОМ подразумевает его способность извлекать информацию о локальных оптических свойствах образца из “ближнего поля”. Особенностью ближнего поля является присутствие в нем затухающих волн (их называют также неоднородными в отличие от однородных или распространяющихся волн). Затухающие волны не способны распространяться в пространстве, а значит, не могут быть замечены удаленным наблюдателем. Необходим некий связующий элемент между ближним и дальним полями, преобразующий одну форму электромагнитной энергии в другую. Роль посредника может выполнять зонд, приближаемый вплотную к образцу.

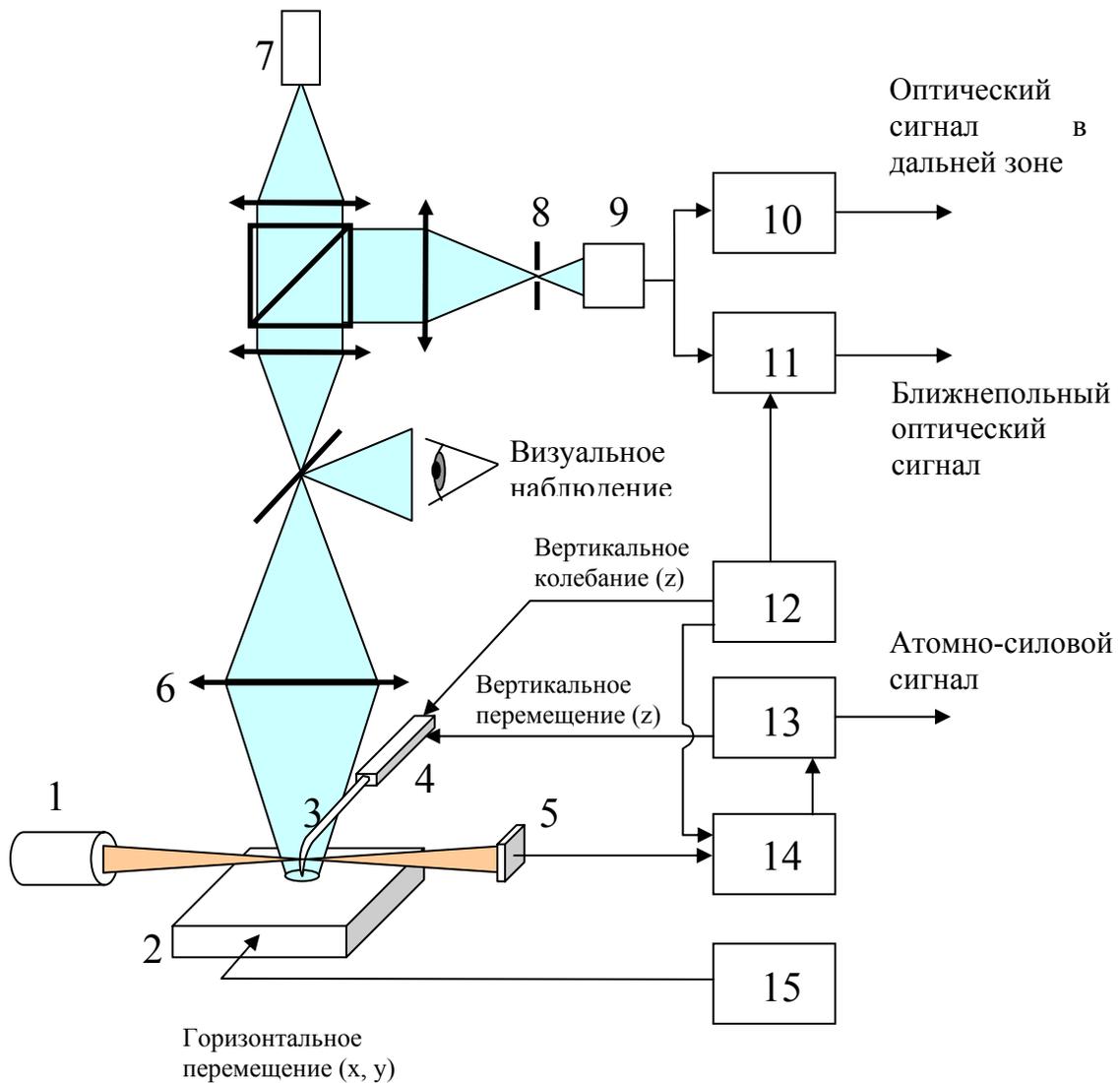


Рис.3 Принципиальная схема безапертурного БСОМ, работающего в отражательном режиме [7]: 1 – лазер для контроля зонда; 2 – образец; 3 – зонд; 4 – пьезо-сканер; 5, 9 – фотоприемники; 6 – объектив; 7 – зондирующий лазер; 8 – диафрагма; 10 – фильтр нижних частот; 11, 14 - синхронный усилитель; 12 – генератор; 13 – система обратной связи; 15 – высоковольтный усилитель.

Сосуществование распространяющихся и затухающих волн характерно для различных физических явлений, рассмотренных в следующей главе.

Два первых раздела носят вводный характер и могут быть пропущены читателями, знакомыми с основами электродинамики. Мы включили их в текст наряду с разделами, посвященными знакомству с понятиями “волновая функция” и “туннельный эффект”, чтобы улучшить связность изложения и избежать необходимости частых ссылок на курсы по теории электромагнетизма и квантовой механики.

3. Распространяющиеся и затухающие волны – две стороны физических явлений

3.1. Волновые процессы

Образование волны представляет процесс, вызванный распространением возмущения в среде. Таким возмущением может быть, например, отклонение натянутой струны из положения равновесия.

В простейшем случае гармонических колебаний зависимость отклонения от времени имеет вид $U = U_0 \sin \omega t$, где U_0 - амплитуда, ω - циклическая частота. Мы предполагаем, что затухание колебаний в среде отсутствует и отклонение U возвращается к исходной величине через промежуток времени $T = 2\pi/\omega$ с частотой $\nu = 1/T$. При распространении волны в некотором направлении, например, вдоль оси z , возмущение приходит в каждую следующую точку с опозданием на время z/v , где v – фазовая скорость. С учетом запаздывания находим

$$U = U_0 \sin \omega(t - z/v) = U_0 \sin(\omega t - kz), \quad (3.1)$$

где $k = 2\pi/\lambda$ – волновое число, λ – длина волны. Функция U обладает периодичностью как по времени, так и по координате. При фиксированном t она одинакова в точках, отстоящих одна от другой на расстояние λ , а при фиксированном z – в моменты, отличающиеся на период T . Вторые производные U по t и z отличаются множителем $(-1/v^2)$, что позволяет рассматривать саму функцию как частное решение волнового уравнения

$$\partial^2 U / \partial z^2 - (1/v^2) \partial^2 U / \partial t^2 = 0. \quad (3.2)$$

Другим частным решением этого уравнения, очевидно, является функция $U = U_0 \cos(\omega t - kz)$.

При выполнении математических операций обычно намного удобнее использовать показательные функции вместо тригонометрических и представлять волну в виде

$$U = U_0 \exp[i(\omega t \mp kz)]. \quad (3.3)$$

Выбор знака в показателе степени зависит от того, распространяется ли волна в сторону возрастания или убывания z .

В общем случае направление распространения волны задается волновым вектором \mathbf{k} , образующим углы α, β, γ с координатными осями и численно равным $|\mathbf{k}| \equiv k = 2\pi/\lambda$. Его составляющие: $k_x = k \cos\alpha, k_y = k \cos\beta, k_z = k \cos\gamma$ определяют периодичность повторения гребней волн вдоль каждой из осей. Вводя по аналогии с частотой колебаний ν пространственные частоты $f_x = k_x/2\pi, f_y = k_y/2\pi, f_z = k_z/2\pi$, приходим к следующему представлению гармонической волны

$$U = U_0 \exp[i(\omega t - \mathbf{k}\mathbf{r})] = U_0 \exp[2\pi i(\nu t - f_x x - f_y y - f_z z)]. \quad (3.4)$$

Эта функция удовлетворяет трехмерному волновому уравнению

$$\nabla^2 U - (1/v^2) \partial^2 U / \partial t^2 = 0, \quad (3.5)$$

где ∇ – векторный оператор градиента с проекциями $\partial/\partial x, \partial/\partial y, \partial/\partial z$ в декартовых координатах, ∇^2 – оператор Лапласа. Гребни волны перемещаются вдоль направления \mathbf{k} со скоростью v вместе с плоскостью $f_x x + f_y y + f_z z = \text{const}$. Такие волны называются плоскими.

Возмущение произвольного вида согласно теореме Фурье может быть представлено в виде суммы (строго говоря, содержащей бесконечное число членов) гармонических функций. Структура волнового поля в этом случае определяется суперпозицией гармонических волн, распространяющихся в разных направлениях.

3.2. Уравнения Максвелла. Электромагнитные волны

Уравнения Максвелла связывают векторы напряженностей \mathbf{E} , \mathbf{H} и индукций \mathbf{D} , \mathbf{B} электрического и магнитного полей с плотностями зарядов ρ и токов \mathbf{j} [8]. Для упрощения записи обычно оставляют по одному из пары векторов. Будем предполагать, что среда является однородной и изотропной. Тогда справедливы соотношения $\mathbf{D} = \varepsilon \mathbf{E} = \varepsilon_r \cdot \varepsilon_0 \mathbf{E}$, $\mathbf{B} = \mu \mathbf{H} = \mu_r \cdot \mu_0 \mathbf{H}$, где ε (ε_r) и μ (μ_r) – абсолютная (относительная) диэлектрическая и магнитная проницаемости среды. Выбрав за основу векторы \mathbf{E} и \mathbf{B} и учитывая, что в оптических средах $\mu_r = 1$, а постоянная μ_0 может быть представлена в виде $\mu_0 = 1/(\varepsilon_0 \cdot c^2)$, где ε_0 – диэлектрическая проницаемость вакуума и c – скорость света в вакууме, приходим к системе уравнений

$$\operatorname{div} \mathbf{E} = \rho / \varepsilon, \quad (3.6)$$

$$\operatorname{div} \mathbf{B} = 0, \quad (3.7)$$

$$\operatorname{rot} \mathbf{E} = -\partial \mathbf{B} / \partial t, \quad (3.8)$$

$$\varepsilon_0 c^2 \operatorname{rot} \mathbf{B} = \mathbf{j} + \varepsilon \partial \mathbf{E} / \partial t. \quad (3.9)$$

Входящие в уравнения (3.6) – (3.9) символы “дивергенция” и “ротор”, примененные к произвольному вектору \mathbf{Q} , в прямоугольной системе координат определяются следующим образом

$$\operatorname{div} \mathbf{Q} = \partial Q_x / \partial x + \partial Q_y / \partial y + \partial Q_z / \partial z, \quad (3.10)$$

$$\operatorname{rot}_x \mathbf{Q} = \partial Q_z / \partial y - \partial Q_y / \partial z,$$

$$\operatorname{rot}_y \mathbf{Q} = \partial Q_x / \partial z - \partial Q_z / \partial x, \quad (3.11)$$

$$\operatorname{rot}_z \mathbf{Q} = \partial Q_y / \partial x - \partial Q_x / \partial y.$$

Помимо напряженностей электрического и магнитного полей в уравнениях электродинамики присутствуют скалярный φ и векторный \mathbf{A} потенциалы. Вид первого из них непосредственно вытекает из закона Кулона

$$\varphi = (1/4\pi\varepsilon_0) \int_V (\rho/r) dV, \quad (3.12)$$

где r – расстояние от точки (x, y, z) , в которой измеряется потенциал, до точки (x', y', z') , где расположен заряд. Согласно уравнению (3.12) потенциал системы зарядов представляет сумму потенциалов, созданных элементарными зарядами ρdV , находящимися в объеме V . Векторный потенциал записывается в аналогичной форме с той разницей, что вместо плотности зарядов ρ присутствует плотность тока \mathbf{j} , а в множитель перед интегралом входит величина $\mu = \mu_0 = 1/\epsilon_0 c^2$

$$\mathbf{A} = (1/4\pi\epsilon_0 c^2) \int_V (\mathbf{j}/r) dV . \quad (3.13)$$

В формулах (3.12), (3.13) есть одна тонкость, на которой мы подробнее остановимся позднее. Она состоит в том, что потенциалы не могут сколь угодно быстро реагировать на изменения зарядов и токов и вычисляются с запаздыванием на время r/v , где v – скорость распространения возмущения.

Дополнительное условие, накладываемое на потенциалы ϕ и \mathbf{A} , имеет вид

$$\operatorname{div}\mathbf{A} = (-1/c^2)\partial\phi/\partial t . \quad (3.14)$$

Зная потенциалы, можно найти векторы \mathbf{E} , \mathbf{B} по формулам

$$\mathbf{E} = -\nabla\phi - \partial\mathbf{A}/\partial t , \quad (3.15)$$

$$\mathbf{B} = \operatorname{rot}\mathbf{A} . \quad (3.16)$$

При отсутствии переменных токов $\partial\mathbf{A}/\partial t = 0$ формула (3.15) превращается в определение электростатического потенциала.

Смысл двух первых уравнений Максвелла в основном сводится к следующему утверждению: в любой области пространства возможно присутствие разного числа элементарных электрических зарядов разного знака, но для “магнитных зарядов” такая возможность отсутствует (полюса магнита не существуют порознь один от другого). Поэтому поток электрических силовых линий через замкнутую поверхность, связанный теоремой Гаусса с расходимостью (дивергенцией) вектора, отличен от нуля, когда в области внутри поверхности находится заряд, тогда как магнитный поток через ту же поверхность при всех обстоятельствах равен нулю.

Два следующих уравнения показывают, что изменения \mathbf{E} и \mathbf{B} во времени и пространстве взаимосвязаны. Действительно, представим,

например, что электрическое поле в некоторой области, не содержащей зарядов и токов, станет исчезать ($\partial \mathbf{E} / \partial t < 0$). Согласно уравнению (3.9) и определению ротора (3.11) это приведет к распространению магнитного поля в соседнюю область, где раньше оно могло отсутствовать, а значит, к уменьшению его напряженности в исходной точке. Но изменение магнитного поля со временем согласно уравнению (3.10) в свою очередь вызовет распространение электрического поля. Возникает самоподдерживающийся процесс, в котором электрическая и магнитная составляющие поля находятся в непрерывном взаимодействии. Комбинируя уравнения Максвелла и применяя к ним простейшие операции векторного анализа, легко убедиться, что в однородной изотропной среде этот процесс представляет собой распространяющуюся волну. Его можно представить набором поперечных волн, каждая из которых характеризуется векторами \mathbf{E} и \mathbf{B} , перпендикулярными один другому и направлению движения волны.

Покажем, что такие представления, по крайней мере, не противоречат уравнениям (3.6) – (3.9). Пусть векторы \mathbf{E} и \mathbf{B} ориентированы соответственно по осям x и y : $E = E_x$, $B = B_y$. Из уравнения (3.9) при $\varepsilon = \varepsilon_r \cdot \varepsilon_0 = \text{const}$ и формул (3.11) следует

$$-c^2 \partial B / \partial z = \varepsilon_r \partial E / \partial t .$$

Дифференцируя это соотношение по времени и переставляя в левой части операторы $\partial / \partial t$ и $\partial / \partial z$, находим

$$-c^2 \partial / \partial z (\partial B / \partial t) = \varepsilon_r (\partial^2 E / \partial t^2),$$

или после подстановки $\partial B / \partial t$ из (3.8)

$$\partial^2 E / \partial z^2 - (1/v^2) \partial^2 E / \partial t^2 = 0, \quad (3.17)$$

$$\text{где } v = c / \sqrt{\varepsilon_r} \equiv c/n, \quad (3.18)$$

n – показатель преломления, показывающий, во сколько раз фазовая скорость распространения возмущения в прозрачной среде меньше скорости света в пустоте.

Мы получили волновое уравнение для вектора \mathbf{E} с заданной ориентацией. Этому уравнению и аналогичному ему уравнению для вектора \mathbf{B} удовлетворяют плоские волны вида

$$E = E_x = E_0 \exp[i(\omega t - kz)], \quad B = B_y = B_0 \exp[i(\omega t - kz)]. \quad (3.19)$$

В общем случае, когда ориентация \mathbf{E} произвольна, одномерное волновое уравнение заменяется трехмерным

$$\nabla^2 \mathbf{E} - (1/v^2) \partial^2 \mathbf{E} / \partial t^2 = 0, \quad (3.20)$$

где $\mathbf{E} = \mathbf{E}(\mathbf{r}, t)$, ∇^2 – оператор Лапласа (в декартовой системе $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$). Аналогичные уравнения справедливы и для функций \mathbf{B} , φ , \mathbf{A} . Все решения трехмерного волнового уравнения могут быть представлены в виде суперпозиции одномерных решений, которые описывают волны, бегущие вдоль осей x , y , z или в любом промежуточном направлении.

Для гармонических колебаний двойное дифференцирование \mathbf{E} по времени приводит к появлению дополнительного множителя $(-\omega^2)$. После сокращения на $\exp(i\omega t)$ уравнение (3.20) приходит к виду

$$\nabla^2 \mathbf{E} + k^2 \mathbf{E} = 0, \quad (3.21)$$

где $\mathbf{E} = \mathbf{E}(\mathbf{r})$ и $k = \omega/v$. Это уравнение называется уравнением Гельмгольца.

Если свет переходит из одной среды в другую, причем разделяющая поверхность не содержит токов, из уравнений Максвелла вытекают граничные условия

$$(E_t)_1 = (E_t)_2, \quad (B_t)_1 = (B_t)_2, \quad (3.22)$$

где индексом t отмечены составляющие векторов \mathbf{E} и \mathbf{B} , касательные к границе раздела. На границе идеального проводника

$$E_t = 0. \quad (3.23)$$

Распространение электромагнитных волн сопровождается переносом энергии. Направление и величина потока энергии определяются вектором Пойнтинга

$$\mathbf{S} = [\mathbf{E}, \mathbf{H}] = \varepsilon_0 c^2 [\mathbf{E}, \mathbf{B}]. \quad (3.24)$$

С позиций квантовой физики перенос энергии можно рассматривать как движение потока фотонов – частиц, энергия и импульс которых равны соответственно

$$W = \hbar\omega, \quad \mathbf{p} = \hbar\mathbf{k}, \quad (3.25)$$

где \hbar - постоянная Планка, \mathbf{k} – волновой вектор.

3.3. Физический смысл комплексного волнового вектора.

Распространение света в материальной среде сопровождается процессами поглощения и рассеяния. Их можно рассматривать как результат взаимодействия фотонов с составляющими среду материальными частицами. Естественно ожидать, что доля фотонов, испытавших взаимодействие в слое однородного материала, пропорциональна его толщине. Тогда уменьшение интенсивности dI параллельного пучка при прохождении через слой толщиной dz можно представить в виде $dI = -I(z)\gamma dz$, где γ – коэффициент поглощения. Отсюда следует

$$I(z) = I(0) \exp(-\gamma z). \quad (3.26)$$

Экспоненциальное ослабление интенсивности прошедшего света с толщиной z подтверждается экспериментально.

Поглощение света в твердом теле обычно предпочитают характеризовать связанным с γ коэффициентом $\chi = nk = \gamma\lambda_0/4\pi$, где λ_0 – длина волны в вакууме [9]. Учитывая квадратичную зависимость интенсивности I от напряженности поля E , из формулы (3.1) для гармонической волны, распространяющейся вдоль оси z , находим

$$E(z) = E(0) \exp(-2\pi nkz / \lambda_0) \exp[i(\omega t - kz)], \quad (3.27)$$

где $k = 2\pi/\lambda = 2\pi n/\lambda_0$ – волновое число, n – показатель преломления. Эту формулу можно представить в более удобном виде, если ввести комплексное волновое число $\hat{k} = k_1 + ik_2$, где $k_1 = k$, $k_2 = -2\pi nk/\lambda_0 = -k\chi$. Тогда

$$E(z) = E(0) \exp[i(\omega t - \hat{k}z)]. \quad (3.28)$$

Введение комплексного волнового числа автоматически приводит и к комплексному показателю преломления

$$\hat{n} = n (1 - i\kappa), \quad (3.29)$$

вещественная часть которого, как и в полностью прозрачной среде, характеризует фазовую скорость распространения волны $v = c/n$, а мнимая – скорость уменьшения амплитуды волны по мере проникновения в поглощающую среду.

Удобство представления показателя преломления в форме (3.29) состоит в том, что соотношения, полученные для прозрачных сред, сохраняют силу и с учетом поглощения при замене n на \hat{n} . Рассмотрим простейший пример. Согласно формулам Френеля, при нормальном падении света из воздуха на диэлектрик с показателем преломления n коэффициент отражения равен $R = (n - 1)^2 / (n + 1)^2$. Чтобы найти коэффициент отражения от поверхности металла, достаточно заменить n на \hat{n}

$$R = \frac{(\hat{n} - 1)(\hat{n}^* - 1)}{(\hat{n} + 1)(\hat{n}^* + 1)} = \frac{(n - 1)^2 + (n\kappa)^2}{(n + 1)^2 + (n\kappa)^2}. \quad (3.30)$$

Металлы с хорошей электропроводностью характеризуются большими значениями $n\kappa$ и малыми n в видимом диапазоне. Так, подставляя характерные для Ag значения $n\kappa = 3,64$, $n = 0,18$ при $\lambda = 600$ нм [9] в формулу (5), находим $R = 0,95$. Близкие к единице значения R , таким образом, указывают на то, что мнимая часть волнового числа по абсолютной величине существенно больше вещественной. При $R = 1$ переход света в форме распространяющейся волны через границу раздела становится невозможным, что приводит к полному внутреннему отражению (ПВО).

Условия возникновения ПВО вытекают из закона Снеллиуса, связывающего углы падения и преломления θ и θ_r при переходе света из среды 1 в среду 2

$$\sin \theta / \sin \theta_r = n_1 / n_2 \equiv n_{12}. \quad (3.31)$$

Формула (3.31) фактически является следствием необходимости сохранения величины составляющей волнового вектора, параллельной границе раздела. Действительно, как видно из рис. 4, на котором показаны последовательные положения софазных поверхностей (например, гребней

волн) в средах 1 и 2, условием сопряжения волн на границе является равенство

$$\lambda_1 / \sin \theta = \lambda_2 / \sin \theta_r \quad \text{или} \quad k_{1x} = k_{2x}. \quad (3.32)$$

Из равенства $k_{2z}^2 + k_{2x}^2 = k_2^2$ и формулы (3.32) следует

$$k_{2z}^2 = k_2^2 - k_{1x}^2 = (2\pi / \lambda_2)^2 (1 - n_{12}^2 \sin^2 \theta). \quad (3.33)$$

В оптически менее плотной среде преломленный луч исчезает при углах падения, превышающих предельный угол $\theta_c = \arcsin (1/n_{12})$. В этих условиях k_{2z} является мнимой величиной: $k_{2z} = i\kappa_z$, где κ_z – действительное положительное число, равное

$$\kappa_z = (2\pi/\lambda_2) (n_{12}^2 \sin^2 \theta - 1)^{1/2}. \quad (3.34)$$

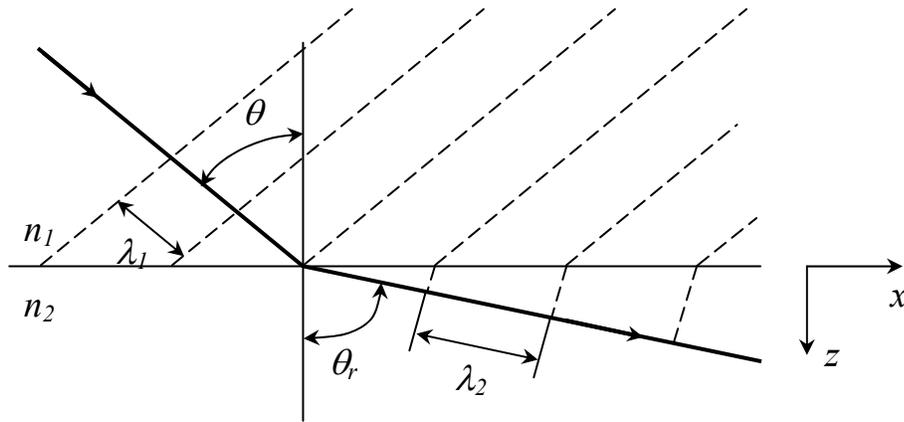


Рис. 4. Преломление светового луча при переходе из среды 1 в среду 2; $n_1 > n_2$. Штриховыми линиями показаны мгновенные положения гребней волны.

Во многих случаях вторая среда является воздухом. Тогда, опуская индексы показателя преломления и производя замену $k_2 = (2\pi/\lambda_2) = \omega/c$, находим

$$\kappa_z = (\omega/c) (n^2 \sin^2 \theta - 1)^{1/2}. \quad (3.35)$$

Выбор знака “плюс” перед квадратным корнем в (3.34), (3.35) обеспечивает физически приемлемое решение – амплитуда волны спадает с ростом расстояния z от границы раздела по закону: $U(z) = U(0) \exp(-\kappa_z z)$. На расстоянии $\delta = 1/(2\kappa_z)$ от границы интенсивность поля убывает в e раз. Величину δ называют глубиной проникновения поля. В идеальном диэлектрике свет распространяется без потерь на любое расстояние. В металлах $n\kappa \geq 1,5$ для видимого диапазона, и свет проходит на глубину $\delta \leq 0,1\lambda_0$. При ПВО в широком диапазоне углов θ поле проникает в запрещенную область на расстояние того же порядка, но при $\theta \rightarrow \theta_c$ δ неограниченно возрастает.

3.4. Корпускулярно-волновой дуализм. Уравнение Шредингера. Волновая функция

Способность фотонов проявлять себя в зависимости от постановки опыта как волна или как частица (“корпускулярно-волновой дуализм”) характерна и для других объектов микромира. Согласно гипотезе де-Бройля, частице с импульсом $\mathbf{p} = m\mathbf{v}$ (m – масса частицы, \mathbf{v} – ее скорость) может быть приписана длина волны $\lambda = 2\pi p/\hbar$ или волновой вектор $\mathbf{k} = \mathbf{p}/\hbar$, где \hbar – постоянная Планка (то же название имеет и постоянная $h = 2\pi\hbar$).

В квантовой механике [10] поведение малой частицы описывается волновой функцией $\Psi(\mathbf{r}, t)$, подчиняющейся уравнению Шредингера

$$-(\hbar^2 / 2m)\nabla^2\Psi + V\Psi = i\hbar(\partial\Psi / \partial t), \quad (3.36)$$

где $V = V(\mathbf{r}, t)$ – потенциал поля, в котором находится частица, m – ее масса, t – время. Уравнение Шредингера, как и уравнения Максвелла, не может быть строго выведено на основе общих принципов. Его правильность подтверждается согласием предсказанных результатов с опытами. Прямой подстановкой легко убедиться, что, если поле консервативно, уравнению (3.36) удовлетворяет функция вида $\Psi(\mathbf{r}, t) = \psi(\mathbf{r}) \exp(-iWt)$, где $\psi(\mathbf{r})$ является решением уравнения

$$[-(\hbar^2 / 2m)\nabla^2 + V(\mathbf{r})]\psi(\mathbf{r}) = W\psi(\mathbf{r}). \quad (3.37)$$

Постоянная W характеризует полную энергию частицы. При решении широкого круга задач достаточно рассматривать уравнение Шредингера для стационарного состояния (3.37), подразумевая, что волновая функция содержит еще и множитель, зависящий от t . Если частица локализована в некоторой области, возможные значения энергии W в (3.37) квантованы и называются собственными значениями уравнения, а соответствующие им решения – собственными функциями. Поскольку разность $W - V$ представляет кинетическую энергию частицы, ее можно записать в виде $W - V = mv^2/2 = p^2/2m = \hbar^2 k^2/2m$.

Рассмотрим простейший случай, когда частица движется в поле с постоянным потенциалом и волновая функция зависит только от одной координаты, например, z . Тогда одним из частных решений уравнения (3.37) при $W - V > 0$ является функция $\psi = \psi_0 \exp(ikz)$. Добавляя множитель $\exp(-iWt/\hbar)$, приходим к волновой функции

$$\Psi = \psi_0 \exp[i(kz - \omega t)], \quad (3.38)$$

где $\omega = W/\hbar$. За исключением знака в показателе степени, формула (3.38) эквивалентна уравнению волны, бегущей вдоль оси z . Выбор знака не играет существенной роли, так как в квантовой механике физический смысл приписывается не самой функции Ψ , а квадрату ее модуля $\Psi\Psi^* = |\Psi|^2$, который пропорционален плотности вероятности нахождения частицы в данном месте в данный момент времени. Связь $|\Psi|^2$ с вероятностью события приводит к условию нормировки – равенству единице интеграла этой функции по всему объему, в котором может находиться частица. Помимо этого к функции Ψ предъявляют требования непрерывности и однозначности, а также непрерывности ее первых производных по координатам (в одномерном случае – непрерывности $\delta\Psi/\delta z$). Если направление движения частицы не совпадает с направлением одной из координатных осей, но по-прежнему $V = \text{const}$, уравнению Шредингера удовлетворяет функция

$$\Psi = \psi_0 \exp[i(\mathbf{kr} - \omega t)]. \quad (3.39)$$

Уравнение (3.36), как и его решения, во многом напоминает волновое уравнение, но содержит первую производную по времени вместо второй. Это отличие носит принципиальный характер. В то время как волновому уравнению удовлетворяют как гармонические, так и показательные

комплексные функции, решение зависящего от времени уравнения Шредингера нельзя представить в виде линейной комбинации синусов и косинусов. Необходимость выбора комплексного решения означает, что функцию Ψ (в отличие от $|\Psi|^2$) нельзя отождествлять с какой-либо величиной, имеющей физический смысл. Волновая функция, таким образом, играет роль некой вспомогательной конструкции, средства для достижения цели. Напрашивается аналогия с ролью векторов **E** и **H**, например, при приеме излучения антенны, когда для сравнения расчетов с опытом представляют интерес не сами по себе напряженности полей, но их произведение, определяющее регистрируемый детектором поток мощности.

3.5. Туннельный эффект

Рассмотрим поведение частицы, например, электрона в поле со ступенчатым потенциалом: $V = 0$ в области 1 ($z \leq 0$) и $V = V_0 > 0$ в области 2 ($z > 0$) (рис. 5). Будем предполагать, что электрон с энергией W первоначально находится в области 1 и перемещается в положительном направлении оси z . Область 2 может простирается на бесконечное расстояние или иметь конечную протяженность d . Ситуация существенно различна в трех случаях, характеризующихся различными значениями W/V_0 и d .

$$1) \quad W > V_0, \quad d \rightarrow \infty$$

Зависящие от координаты z части ψ_1 и ψ_2 волновых функций электрона в областях 1 и 2 удовлетворяют уравнениям Гельмгольца, представляющим более компактный вариант записи одномерного уравнения Шредингера

$$d^2\psi_1 / dz^2 + k_1^2\psi_1 = 0, \quad (3.40)$$

$$d^2\psi_2 / dz^2 + k_2^2\psi_2 = 0, \quad (3.41)$$

$$\text{где } k_1 = (2mW)^{1/2} / \hbar = 2\pi / \lambda_1, \quad (3.42)$$

$$k_2 = [2m(W - V_0)]^{1/2} / \hbar = 2\pi / \lambda_2, \quad (3.43)$$

где λ_1, λ_2 – длины волн де-Бройля.

Решениями уравнений (3.40), (3.41) являются линейные комбинации плоских волн де-Бройля

$$\psi_1 = a_1 \exp(ik_1 z) + b_1 \exp(-ik_1 z), \quad (3.44)$$

$$\psi_2 = a_2 \exp(ik_2 z) + b_2 \exp(-ik_2 z), \quad (3.45)$$

Для сокращения записи примем $a_1 = 1$. Тогда функция ψ_1 представляет суперпозицию прямой волны с единичной амплитудой и отраженной от ступени потенциала волны с амплитудой b_1 . В области 2 отраженной волны нет и, следовательно, $b_2 = 0$. Чтобы найти b_1 , нужно учесть требование непрерывности ψ_1 и $d\psi_1/dz$ на границе. Приравнявая волновые функции (3.44), (3.45) и их производные по z при $z = 0$, приходим к равенствам

$$1 + b_1 = a_2, \quad k_1(1 - b_1) = a_2 k_2 \quad (3.46)$$

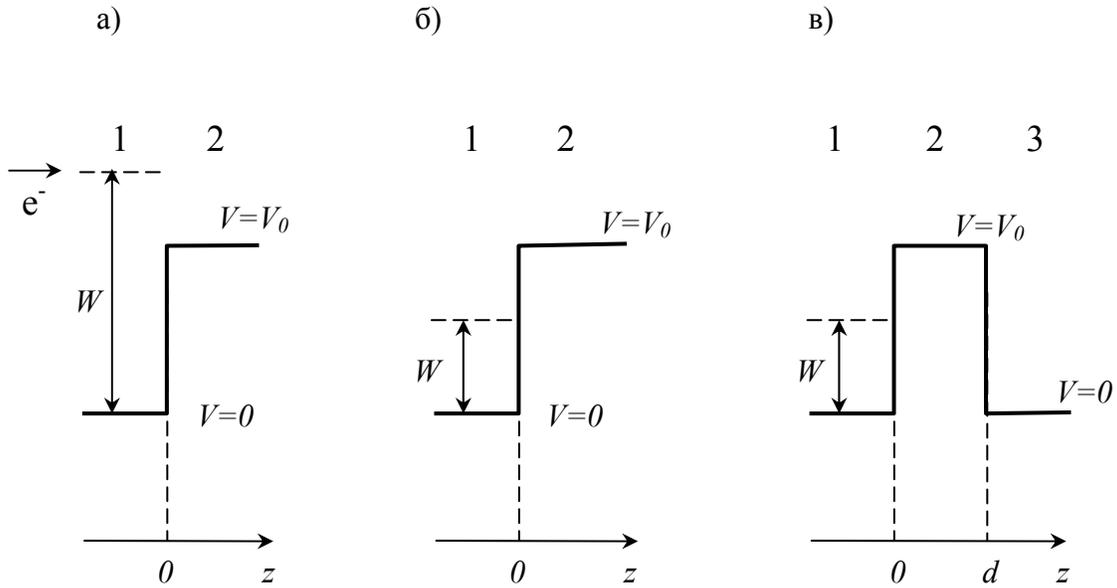


Рис. 5. К объяснению взаимодействия электрона с энергией W с прямоугольным потенциальным барьером:

а) $W/V_0 > 1, d \rightarrow \infty$;

б) $W/V_0 < 1, d \rightarrow \infty$;

в) $W/V_0 < 1, d < \infty$.

Исключая из этой системы a_2 , находим b_1 и коэффициент отражения R , характеризующий относительную интенсивность отраженной волны

$$R = b_1^2 = \left[\frac{k_1/k_2 - 1}{k_1/k_2 + 1} \right]^2. \quad (3.47)$$

Формула (3.47) аналогична соотношению, связывающему коэффициент отражения света от границы раздела двух сред с вещественными показателями преломления в условиях, когда волны могут распространяться по обе стороны от границы (см. разд.3.3). Коэффициент прозрачности границы раздела для электронов $D = 1 - R > 0$.

2) $W < V_0, d \rightarrow \infty$

Уравнение Шредингера для волновой функции в среде 2 принимает вид

$$d^2\psi_2 / dz^2 - \kappa^2\psi_2 = 0, \quad (3.48)$$

$$\text{где } \kappa \equiv [2m(V_0 - W)]^{1/2} / \hbar. \quad (3.49)$$

Как и в других подобных задачах, мы считаем $\kappa > 0$, т.е. выбираем знак “плюс” перед квадратным корнем. Только такой выбор приводит к решению, имеющему физический смысл.

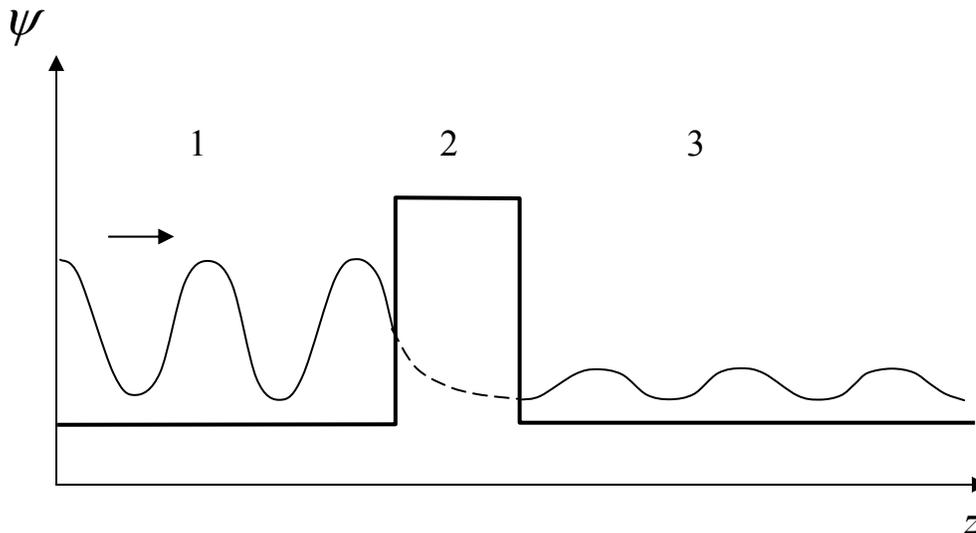


Рис. 6. Прохождение волны через барьер.

Уравнение (3.48) называют модифицированным уравнением Гельмгольца. Коэффициент отражения по-прежнему характеризуется формулой (3.47), но с заменой k_2 на ik

$$R = \left| \frac{k_1 - ik}{k_1 + ik} \right|^2 = \frac{(k_1 - ik)(k_1 + ik)}{(k_1 + ik)(k_1 - ik)} = 1. \quad (3.50)$$

Соответственно, $D = 1 - R = 0$.

Как и следовало ожидать, бесконечно широкий потенциальный барьер полностью отражает электронную волну.

3) $W < V_0$, $d < \infty$

Конечная ширина барьера осложняет поиск решения, требующего теперь согласования волновых функций и их производных на двух границах раздела. В результате система (3.46) заменяется системой четырех уравнений с четырьмя неизвестными коэффициентами. Но принцип построения решения остается тем же: функция, описывающая распространяющуюся в области 1 плоскую синусоидальную волну, сшивается на левой границе барьера с убывающей экспонентой, “хвост” которой в свою очередь переходит в синусоиду с меньшей амплитудой в среде 3 (рис. 6).

Коэффициент прозрачности барьера пропорционален величине $|\psi_2|^2$ в плоскости $z = d$ на правой границе барьера и, следовательно, экспоненциально уменьшается с ростом d

$$D = D_0 \exp(-2kd), \quad (3.51)$$

где $D_0 \approx 1$, а k определяется формулой (3.49).

Прохождение частицы сквозь барьер называется туннельным эффектом. Это чисто квантовый эффект, который с позиций классической физики эквивалентен утверждению возможности проникновения частицы сквозь сплошную стену. С учетом волновой природы частиц такое утверждение перестает казаться абсурдным. Волне, наталкивающейся на барьер, требуется некоторый конечный отрезок, чтобы сбросить амплитуду до нуля и повернуть в обратную сторону. Если барьер достаточно тонок, переход энергии прямой волны в отраженную энергию не сможет реализоваться полностью, и коэффициент прозрачности D будет отличен от нуля. Для электронов в металлах типичные значения работы выхода,

определяемой разностью $V_0 - W$, близки к 5 эВ. Расчет по формулам (3.49), (3.51) показывает, что прозрачность воздушного барьера для электронов убывает от $D \sim 10^{-2}$ при $d = 0,2$ нм до $D \sim 10^{-12}$ при $d = 1$ нм. Таким образом, чтобы обнаружить туннельный эффект и использовать его на практике необходимо подносить приемник тока практически вплотную к источнику электронов. Прикладывая напряжение между электродами, можно менять форму барьера, а значит, и его прозрачность.

На этом принципе основано действие СТМ, в котором роль источника электронов может выполнять металлическое острие, сканирующее проводящий, чаще всего металлический образец. При расстоянии порядка 1 нм между острием и образцом возникает туннельный ток, величина которого зависит от ширины зазора d , а также от ускоряющего напряжения, влияющего на эффективную высоту барьера. В режиме постоянного тока острие воспроизводит картину распределения плотности электронных состояний (плотности поверхностного заряда) в образце.

Изобретение СТМ позволило наглядно представить атомную структуру поверхности широкого круга материалов.

3.6. Туннелирование фотонов. Поверхностные электромагнитные волны

Результаты, изложенные в предыдущем разделе, показывают, что движение электронов в зоне проводимости металла должно иметь много общего с распространением света. При определенных условиях стационарное уравнение Шредингера переходит в уравнение Гельмгольца и волновая функция электрона приобретает вид, характерный для плоской волны. Когда электрон наталкивается на потенциальный барьер, препятствующий его распространению, волновая функция электрона не исчезает полностью в области барьера и подчиняется модифицированному уравнению Гельмгольца. При достаточной ширине барьера он становится непрозрачным для электронов и коэффициент отражения $R = 1$.

Все указанные особенности характерны и для света при полном внутреннем отражении (ПВО). Роль барьера выполняет среда, в которой при углах падения, превышающих критический угол, распространение света становится невозможным.

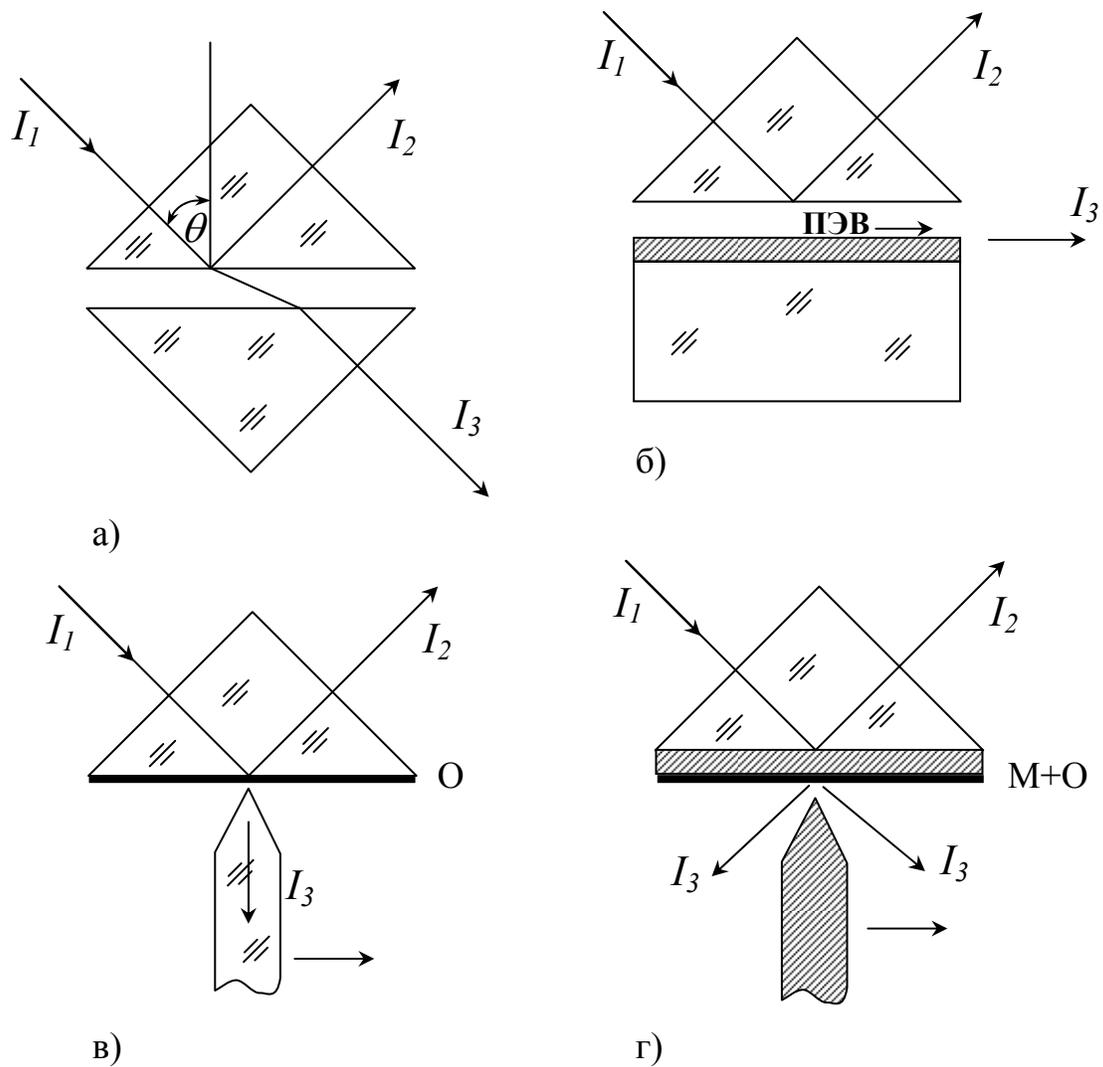


Рис. 7. Принципиальные схемы опытов с нарушенным полным внутренним отражением:

а) переход света через воздушную прослойку между двумя призмами; $\theta > \theta_c = \arcsin(1/n)$

б) возбуждение поверхностной электромагнитной волны (ПЭВ) по методу Кречмана;

в) и г) преобразование затухающего поля в излучение в ФСТМ (в) и СПОМ (г); О – образец, М – металлический слой

Исходя из этой аналогии, мы вправе ожидать, что, создавая условия “нарушенного полного внутреннего отражения” (НПВО), например, путем уменьшения ширины барьера, можно наблюдать оптические эффекты, подобные туннелированию электронов. Впервые на один из таких эффектов обратил внимание Ньютон. Он заметил, что если с гранью призмы в схеме ПВО соприкасается слегка закругленная грань другой призмы, то свет переходит в нее не только через область контакта, но и через часть воздушного промежутка. Позднее подобные опыты стали ставить и с двумя призмами, имеющими плоские грани, сближая их до расстояния порядка длины волны (рис. 7а). Было показано, что пропускание света прослойкой зависит от ее показателя преломления, а также от угла падения и поляризации света.

Если на грань одной из призм нанести слой металла или определенную комбинацию слоев металла и диэлектрика, можно заметить другое интересное явление – часть энергии падающего луча переходит в энергию поверхностной электромагнитной волны (ПЭВ) [11], распространяющейся вдоль границы раздела (рис. 7б). Особенностью ПЭВ является экспоненциальный спад интенсивности поля с расстоянием z по обе стороны от границы: $I(z) = I(0) \exp(\pm 2\kappa_i z)$, где $i = 1, 2$, а знак в показателе степени противоположен знаку z ; предполагается, что граница раздела совпадает с плоскостью $z = 0$.

Коэффициенты затухания κ_1 и κ_2 соответственно в металле и диэлектрике существенно различны ($\kappa_1 \ll \kappa_2$), так что ПЭВ фактически распространяется в открытом металлическом волноводе, лишь слегка проникая в стенку. Интенсивность ПЭВ убывает и вдоль направления распространения из-за поглощения в металле и рассеяния на поверхностных неоднородностях. Из-за затухания ПЭВ в обеих контактирующих средах их нельзя возбудить объемной волной, если граница раздела является гладкой и однородной. Приходится применять специальные методы, в том числе с использованием различных схем НПВО. В свою очередь, ПЭВ на гладкой границе раздела не может преобразоваться в излучение.

Условия возбуждения ПЭВ чрезвычайно чувствительны к состоянию поверхности. Иногда достаточно присутствия монослоя адсорбированных молекул, чтобы вызвать заметное изменение интенсивности отраженного или прошедшего света. Этот эффект используется в “ПЭВ-микроскопии” (микроскопии поверхностных плазмонов), позволяющей различать участки неоднородной поверхности по их способности поддерживать ПЭВ.

По аналогии с туннелированием электронов переход света через среду, в которой при выбранных условиях возможно существование только затухающих волн, называют туннелированием фотонов или оптическим

туннелированием. Оптическим аналогом СТМ является фотонный сканирующий туннельный микроскоп (ФСТМ) [12], в котором заостренный волоконный зонд скользит над гранью призмы или прозрачной плоской пластины с закрепленным на ней тонким образцом (рис.7в).

При освещении образца под углом $\theta > \theta_c$ интенсивность улавливаемого острием света меняется с расстоянием z от поверхности по закону

$$I(z) = I(0) \exp(-2kz) , \quad (3.52)$$

где k определяется формулой (3.35) (здесь и ниже мы опускаем нижний индекс z при k , введенный в разделе 3.3 с целью различить коэффициенты затухания волны и поглощения света в металлах). Поскольку k зависит от показателя преломления, сигнал меняется при подходе острия к участкам, отличающихся по оптическим свойствам от остальных участков поверхности. Необходимо иметь в виду, что хотя соотношение (3.52) по виду аналогично зависимости интенсивности затухающей волны от расстояния до плоской границы (например, в каждом из слоев, несущих ПЭВ), входящая в него величина z не поддается однозначному определению. Здесь, как и при рассмотрении туннельного тока, существенен выбор точки, от которой ведется отсчет расстояния до образца. Анализ показывает, что такой точкой приближенно можно считать центр кривизны вершины острия.

При одинаковом для СТМ и ФСТМ характере зависимости тока детектора (или интенсивности затухающего поля) от z , два прибора существенно отличаются по величине коэффициента затухания k или по глубине проникновения δ волны в область барьера. Для электронов в СТМ при типичных ускоряющих напряжениях $\delta = \delta_e \sim 0,1$ нм, а для фотонов видимого диапазона $\delta = \delta_{ph} \sim 10^2$ нм. Это различие дает ФСТМ преимущество при исследованиях, требующих конечной глубины анализа, но в то же время затрудняет обнаружение с помощью оптических методов малых относительных смещений поверхностных атомов, доступное для СТМ.

Помимо ФСТМ, оптическое туннелирование играет принципиальную, хотя и не столь очевидную роль и в других вариантах БСОМ.

При возбуждении ПЭВ по методу Кречмана, когда металлический слой нанесен непосредственно на грань призмы ПВО [11], приближение металлической иглы к поверхности образца меняет локальные условия распространения ПЭВ и вызывает перекачку части энергии в излучение, что проявляется в изменении интенсивности отраженного луча (рис. 7г). На этом эффекте основано действие сканирующих плазмонных оптических

микроскопов [13], сочетающих в себе высокое пространственное разрешение СЗМ с высоким контрастом ПЭВ-микроскопов.

3.7. Структура поля в волноводах. Отсечка

Волны, распространяющиеся в волноводе, должны удовлетворять граничным условиям, которые при отсутствии поверхностных токов сводятся к требованию непрерывности тангенциальных составляющих электрического и магнитного векторов. Каждому типу волны (моде) соответствует характерная структура электромагнитного поля и может быть приписана определенная граничная частота (частота отсечки). Последовательная отсечка мод при уменьшении сечения волновода является основной причиной значительных потерь мощности в оптических зондах.

При рассмотрении отсечки начнем с анализа простейшей картины электрического поля, возникающего в полном прямоугольном волноводе с идеально проводящими стенками. Направим ось z по оси волновода, а оси x и y соответственно вдоль горизонтальной и вертикальной стенок с поперечными размерами $2a$ и $2b$. Среди различных структур электромагнитного поля могут существовать такие, в которых вектор \mathbf{E} внутри волновода сохраняет неизменную ориентацию, например, параллелен оси y . В этом случае автоматически обеспечивается равенство нулю тангенциальной составляющей E_t на горизонтальных стенках. Аналогичному требованию на вертикальных стенках удовлетворяет функция

$$E = E_y \cos(\pi x / 2a) \exp[i(\omega t - k_z z)]. \quad (3.53)$$

В то же время, она является и одним из возможных решений волнового уравнения (3.20). Действительно, подставляя вторые производные E по x , z и t в (3.20) при $k_x = \pi/2a$, $k_y = 0$ (E не зависит от y), приходим к тождеству: $k_x^2 + k_z^2 = (\omega/c)^2 = k_0^2$. Отсюда следует

$$k_z = \sqrt{(\omega/c)^2 - (\pi/2a)^2}. \quad (3.54)$$

Волновое число k_z определяет фазовую скорость $v_z = \omega/k_z$ и длину волны вдоль оси z

$$\lambda_z = 2\pi/k_z = \lambda_0 / \sqrt{1 - (\lambda_0/4a)^2}, \quad (3.55)$$

где $\lambda_0 = 2\pi/k_0$ – длина волны в вакууме.

В достаточно широком волноводе при $a \gg \lambda_0$ свет распространяется почти как в свободном пространстве и $\lambda_z \approx \lambda_0$. Однако когда поперечные размеры волновода становятся соизмеримыми с длиной волны света, λ_z существенно растет и при $\lambda_0 \rightarrow 4a$, $\lambda_z \rightarrow \infty$. Такая волна, очевидно, не может распространяться по волноводу, так как расстояние между двумя гребнями волны превышает длину любого отрезка трубы.

К аналогичному выводу можно прийти и из формулы (3.54). При уменьшении ω или a k_z из действительной величины становится мнимой. Критическая частота ω_c определяется равенством

$$\omega_c = \pi c / 2a . \quad (3.56)$$

В области частот $\omega < \omega_c$ волновое число k_z можно представить в виде $k_z = ik$, где k – действительная положительная величина

$$k = \sqrt{(\pi/2a)^2 - (\omega/c)^2} . \quad (3.57)$$

Очевидно, возможные решения отнюдь не исчерпываются рассмотренным типом волн. Требованию $E_t = 0$ на идеально проводящих стенках, помимо (3.53), удовлетворяют и функции, получающиеся при замене $\cos(\pi x / 2a)$ на $\cos(\pi l x / 2a)$, где l – любое целое число. Такая замена эквивалентна увеличению числа полуволн, укладываемых на отрезке $2a$. С равным основанием мы могли бы выбрать за направление \mathbf{E} и ось x . Тогда все формулы останутся в силе, если поменять a на b , а вместо l ввести m – целое число полуволн вдоль вертикальной стенки. Поля с вектором \mathbf{E} , направленным перпендикулярно оси волновода, называются поперечно-электрическими и обозначаются TE_{lm} . Помимо них в волноводе могут распространяться поперечно-магнитные моды TM_{lm} , название которых указывает на поперечную ориентацию магнитного вектора, а также моды смешанного типа [8].

Для оптики наибольший интерес представляют волноводы с круглым сечением, обычно со стеклянной или кварцевой сердцевиной. В диэлектрике с показателем преломления n длина волны и фазовая скорость равны соответственно $\lambda = \lambda_0 / n$, $v = c/n$ и, следовательно, в полученных для полого волновода формулах (3.54) - (3.57) необходимо заменить λ_0 на λ_0/n и c на c/n . Переход от прямоугольной формы сечения к круглой в данном случае, как и

в других подобных задачах, приводит к замене тригонометрических функций бесселевыми. Для тех и других характерны осцилляции величины функции при увеличении аргумента, причем функция попеременно принимает положительные и отрицательные значения.

Возможные типы ТМ и ТЕ-волн в круглом волноводе радиусом a с идеально проводящими стенками определяются соответственно корнями уравнений [8]

$$J_l(u) = 0, \quad \partial J_l / \partial u = 0, \quad (3.58)$$

где $u = a(k^2 - k_z^2)^{1/2}$, $k = (\omega/c)n$. Каждому индексу l функций Бесселя отвечает, вообще говоря, бесчисленное число решений уравнений (3.58): $u = u_{lm}$, где m – порядковый номер решения. На практике обычно приходится иметь дело с несколькими первыми индексами l и m . Расположенные в порядке возрастания значения u_{lm} для ТМ-волн составляют

$$u_{01}=2,405, \quad u_{11}=3,832, \quad u_{21}=5,136 \dots \quad (3.59)$$

а для ТЕ-волн

$$u_{11}=1,841, \quad u_{01}=3,832, \quad u_{12}=5,331 \dots \quad (3.60)$$

Как и в случае с прямоугольным волноводом, уменьшение частоты колебаний или поперечных размеров трубки приводит к отсечке – распространяющиеся волны сменяются затухающими с мнимой постоянной распространения $k_z = ik$, где k – действительная положительная величина, связанная с u_{lm} соотношением

$$k = \sqrt{(u_{lm} / a)^2 - (n\omega / c)^2}. \quad (3.61)$$

При заданном радиусе a скорость затухания волны тем меньше, чем меньше u_{lm} . Два минимальных значения u_{lm} , согласно равенствам (3.59), (3.60), отвечают ТЕ₁₁ и ТМ₀₁ модам. В оптическом зонде, конец которого представляет собой отрезок конического волновода, эти (точнее, родственные им смешанные) моды распространяются дальше других и определяют интенсивность поля вблизи диафрагмы.

Формулу (3.61) можно представить в другом виде

$$\kappa = (2\pi/\lambda)\sqrt{(\lambda/\lambda_{lm})^2 - 1}, \quad (3.62)$$

где $\lambda_{lm} = 2\pi a/u_{lm}$ – длина волны в направлении, перпендикулярном оси волновода. Условие $\kappa = 0$ с учетом (3.59), (3.60) дает предельные длины волн, еще способных распространяться по волноводу: они составляют $3,5a$ и $2,6a$ для ТЕ и ТМ волн соответственно. Поскольку минимальный радиус прозрачной сердцевины апертурных зондов у вершины конуса (радиус диафрагмы) обычно не превышает $0,1$ мкм, примыкающий к диафрагме отрезок зонда задерживает значительную часть спектра видимого диапазона (с длинами волн $\lambda_0 > 3,5 na$) и интенсивность поля на выходе определяется в основном глубиной проникновения затухающих мод.

3.8. Электромагнитное поле осциллятора

В теории электромагнетизма рассматриваются три последовательно возрастающие по сложности задачи. В первой, решаемой в курсе школьной физики, из закона Кулона выводится формула для потенциала помещенного в вакуум точечного заряда q на расстоянии r от него

$$\varphi(r) = (1/4\pi\epsilon_0)q/r. \quad (3.63)$$

Вторая задача состоит в определении потенциала поля, созданного двумя разноименными точечными зарядами $+q$ и $-q$, разделенными расстоянием l . При $l \ll r$ эти заряды образуют диполь. Направим ось z по оси диполя и поместим его центр в начало координат. Простые вычисления показывают, что суммарный потенциал в точке $\mathbf{r} = (x, y, z)$ равен

$$\varphi(\mathbf{r}) = (1/4\pi\epsilon_0)\frac{\mathbf{p} \cdot \mathbf{r}}{r^3}, \quad (3.64)$$

где $\mathbf{p} = ql$ – дипольный момент, $\mathbf{p} \cdot \mathbf{r} = pr \cos\theta$ – скалярное произведение векторов \mathbf{p} и \mathbf{r} , θ – угол между ними.

В третьей, существенно более сложной задаче требуется найти поле осциллятора – диполя с изменяющимся во времени дипольным моментом. Простейшим осциллятором является система двух зарядов, из которых один неподвижен, а другой совершает относительно него гармонические колебания с частотой ω , так что $\mathbf{p} = \mathbf{p}_0 \cos\omega t$ (или в комплексной форме $\mathbf{p} = \mathbf{p}_0$

$\exp(i\omega t)$). Как и раньше, мы предполагаем, что диполь находится в вакууме. Особенностью задачи является запаздывание потенциалов \mathbf{A} и φ на время r/c , затрачиваемое на распространение возмущения от источника к точке наблюдения, удаленной на расстояние r . Из формулы (3.13), учитывая, что $\mathbf{j} = q\mathbf{v} = q(\partial\mathbf{l}/\partial t) = \partial\mathbf{p}/\partial t$, находим

$$\mathbf{A}(\mathbf{r}, t) = (1/4\pi\epsilon_0 c^2 r) \frac{\partial \mathbf{p}(t')}{\partial t}, \quad (3.65)$$

где $t' \equiv t - (r/c)$. Чтобы определить $\varphi(\mathbf{r}, t)$, используем соотношение (3.14), связывающее два потенциала. Так как вектор \mathbf{A} , как и \mathbf{p} , имеет отличную от нуля составляющую только по оси z , справедливо равенство

$$\text{div}\mathbf{A} = \partial A / \partial z = \frac{1}{4\pi\epsilon_0 c^2 r} \left[-\frac{z}{r^3} \frac{\partial p(t')}{\partial t} - \frac{z}{cr^2} \frac{\partial^2 p(t')}{\partial t^2} \right]. \quad (3.66)$$

Первый член в квадратных скобках получен дифференцированием по z множителя $(1/r) = (x^2 + y^2 + z^2)^{-1/2}$. При вычислении второго члена используются следующие преобразования

$$\begin{aligned} \frac{\partial}{\partial z} \frac{\partial}{\partial t} \exp[i\omega(t - \frac{r}{c})] &= \frac{\partial}{\partial t'} \frac{\partial t'}{\partial r} \frac{\partial r}{\partial z} \exp(i\omega t) = \\ (i\omega)(-i\omega/c) \frac{z}{r} \exp(i\omega t') &= -\frac{z}{rc} \frac{\partial^2}{\partial t^2} \exp(i\omega t'). \end{aligned}$$

Из (3.14) и равенства $p z = \mathbf{p} \cdot \mathbf{r}$ следует

$$\varphi(\mathbf{r}, t) = \int (\text{div}\mathbf{A}) dt + \text{const} = \frac{1}{4\pi\epsilon_0 r^3} \left[\mathbf{p}(t') \cdot \mathbf{r} + \frac{r}{c} \frac{\partial \mathbf{p}(t')}{\partial t} \cdot \mathbf{r} \right]. \quad (3.67)$$

Постоянная интегрирования принята равной нулю, так как наличие статического поля не сказывается на излучающих свойствах диполя и можно заранее допустить, что это поле отсутствует.

Поскольку для гармонического осциллятора дифференцирование p по времени эквивалентно умножению на $i\omega$, второй член в (3.67) отличается по абсолютной величине от первого в $(r\omega/c) = rk = 2\pi r/\lambda$ раз. При $r \ll \lambda/2\pi$ (эта

область называется ближней зоной источника) t' мало отличается от t , так что запаздыванием потенциала можно пренебречь, и величина φ определяется первым слагаемым. В этом случае формула (3.67) имеет тот же вид, что и формула (3.64) для потенциала статического диполя. В другом предельном случае, когда $r \gg \lambda/2\pi$ (в дальней или волновой зоне источника) основной вклад в $\varphi(\mathbf{r},t)$ дает второй член, вызванный наличием переменного тока.

Формул (3.65), (3.67) наряду с (3.15), (3.16) достаточно для строгого вычисления напряженностей электрического и магнитного полей. Конечный результат нагляднее всего может быть представлен в сферических координатах [8]

$$\begin{aligned} E_r &= \frac{p(t')}{2\pi\epsilon_0} \left(\frac{1}{r^3} - \frac{ik}{r^2} \right) \cos \theta, \\ E_\theta &= \frac{p(t')}{4\pi\epsilon_0} \left(\frac{1}{r^3} - \frac{ik}{r^2} - \frac{k^2}{r} \right) \sin \theta, \\ H_\varphi &= \frac{B_\varphi}{\epsilon_0 c^2} = -\frac{i\omega}{4\pi} p(t') \left(\frac{1}{r^2} - \frac{ik}{r} \right) \sin \theta. \end{aligned} \quad (3.68)$$

В непосредственной близости от источника при $kr = 2\pi r/\lambda \ll 1$ преобладают “статический” и “индукционный” члены, убывающие при увеличении r как r^{-3} и r^{-2} . На больших расстояниях остаются только ответственные за излучение члены, пропорциональные r^{-1}

$$\begin{aligned} E &= E_\theta = -\frac{k^2}{4\pi\epsilon_0 r} p(t') \sin \theta, \\ H &= H_\varphi = -\frac{ck^2}{4\pi r} p(t') \sin \theta. \end{aligned} \quad (3.69)$$

Взаимное расположение векторов \mathbf{E} , \mathbf{H} , \mathbf{r} в волновой зоне показано на рис. 8. Эти векторы образуют правовинтовую систему, причем силовые линии электрического поля лежат в меридиональных плоскостях, а магнитного – представляют концентрические окружности вокруг оси диполя.

Из формул (3.24), (3.69) следует, что величина потока электромагнитной энергии, определяемая абсолютной величиной вектора Пойнтинга, обратно пропорциональна квадрату расстояния от источника

$$S = \llbracket \mathbf{E}, \mathbf{H} \rrbracket = \frac{c}{16\pi^2 \varepsilon_0} \frac{k^4 p^2(t')}{r^2} \sin^2 \theta. \quad (3.70)$$

Значит, полная энергия, переносимая через концентрические сферы (площадь которых растет как r^2) не зависит от расстояния. Этот вывод имеет простой физический смысл: в пространстве, лишенном зарядов и токов (например, в вакууме вдали от диполя), отсутствуют центры поглощения, которые могли бы вызвать уменьшение излучаемой энергии.

Из формул (3.24), (3.69) следует, что величина потока электромагнитной энергии, определяемая абсолютной величиной вектора Пойнтинга, обратно пропорциональна квадрату расстояния от источника

$$S = \llbracket \mathbf{E}, \mathbf{H} \rrbracket = \frac{c}{16\pi^2 \varepsilon_0} \frac{k^4 p^2(t')}{r^2} \sin^2 \theta. \quad (3.70)$$

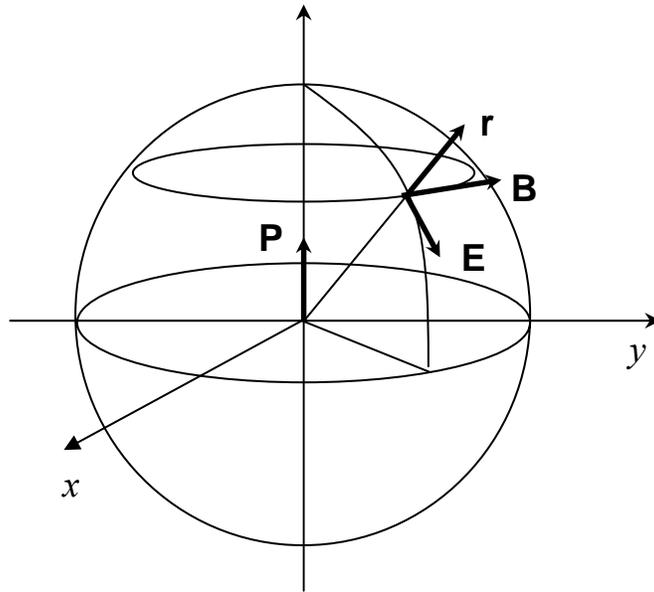


Рис. 8. Взаимное расположение векторов \mathbf{r} , \mathbf{E} , \mathbf{B} в дальней зоне осциллятора.

При оценке потока энергии мы не учитывали входящих в выражения (3.68) членов, меняющихся с r как r^{-3} и r^{-2} . Наличие этих составляющих

поля никак не сказывается на мощности излучения. В то же время, определяемая ими часть вектора Пойнтинга на малых расстояниях отлична от нуля, причем направление потока может не совпадать с направлением радиуса-вектора, проведенного из центра сферы, что указывает на присутствие радиальной составляющей поля. Очевидно, в ближней зоне происходит циркуляция энергии, но без выноса за пределы зоны.

По мере приближения к осциллятору все больший вес приобретают члены, пропорциональные r^{-3} , так что электрическое поле начинает преобладать над магнитным и по своей структуре напоминает поле статического диполя. Разница лишь в том, что напряженность поля не остается постоянной, а, как и дипольный момент, меняется со временем по синусоидальному закону.

Формулы (3.68), (3.69) представляют принципиальный интерес для электродинамики вообще и для понимания теории ближнего поля в частности. Суммируем основные следствия этих формул.

1. Источник излучения, в простейшем случае электрический диполь, непрерывно теряет энергию, перенос которой в дальней зоне ($kr \gg 1$) осуществляется поперечными электромагнитными волнами. В распространяющейся волне векторы **E**, **H** совершают совпадающие по фазе колебания в двух взаимно перпендикулярных плоскостях.

2. В ближней зоне источника ($kr \ll 1$) циркуляция основной части энергии не сопровождается ее выбросом в окружающее пространство. Можно сказать, что диполь, создав некоторый поток мощности, вновь поглощает его, непрерывно подпитывая сам себя. Косвенной аналогией могут быть протуберанцы на Солнце – всплески сгустков раскаленных газов, которые возвращаются к поверхности полем тяготения.

3. Плотность электромагнитной энергии в ближней зоне из-за растущего на малых расстояниях вклада составляющих поля, меняющихся с расстоянием r как r^{-3} и r^{-2} , а также из-за малой протяженности зоны, существенно превышает плотность энергии вдали от источника. Однако убедиться в наличии этих “резервов энергии” можно, только введя детектор непосредственно в ближнюю зону.

На данном этапе нельзя сказать ничего определенного о величине напряженностей электрического и магнитного полей при $r \rightarrow 0$, так как сама постановка задачи расчета поля осциллятора потребовала допущения $r \gg l$, где l – длина диполя.

3.9. Прохождение света через малую диафрагму

При расчете оптических систем часто приходится принимать во внимание ослабление светового потока диафрагмой – отверстием в непрозрачном экране. Прохождение света через диафрагму принято рассматривать на основе принципа Гюйгенса-Френеля, позднее получившего более строгую формулировку в теории Кирхгофа. Предполагается, что вырезаемая диафрагмой часть волнового фронта является источником вторичных волн, причем ввод диафрагмы не сказывается на структуре электромагнитного поля на выделенном участке фронта (направления и абсолютные величины вектора \mathbf{E} и его производной по нормали $\partial\mathbf{E}/\partial n$ остаются неизменными при переходе через плоскость диафрагмы). Очевидно, что это допущение может быть в той или иной степени оправдано лишь при условии $a \gg \lambda$ (или $ak \gg 1$), где a – радиус диафрагмы, λ – длина волны света, k – волновое число. В противном случае необходимо считаться с искажениями поля, вносимыми краями диафрагмы. В теории Кирхгофа сравниваются интенсивности I и I_0 поля в точках, находящихся по разным сторонам от экрана на значительном удалении от диафрагмы радиусом $a \gg \lambda$. В одну точку помещен источник излучения с интенсивностью I_0 , в другую – приемник. Искомая зависимость I от a и k имеет вид

$$I = I_0 A_1(\theta_1) k^2 a^4 \quad \text{при } ak \gg 1, \quad (3.71)$$

где $A_1(\theta_1)$ – функция угла θ_1 между нормалью к экрану и лучом, направленным к наблюдателю. Наличие множителя a^4 вызвано тем, что площадь диафрагмы участвует в расчете дважды: она определяет долю полезной (пропущенной экраном или дошедшей до наблюдателя) мощности как с одной, так и с другой стороны экрана.

В случае $ak \ll 1$ приходится искать другие пути решения. Оригинальный подход был предложен Бете. В электродинамике встречаются задачи (их называют “прямая” и “дополнительная”), решения которых отличаются лишь тем, что место электрического вектора занимает магнитный и наоборот. В частности, одинаковую структуру имеют поля, возникающие при рассеянии света на проводящем диске ($\mathbf{E}_d, \mathbf{H}_d$) и при прохождении света через диафрагму в непрозрачном экране ($\mathbf{E}_a, \mathbf{H}_a$). По принципу Бабине в его уточненной формулировке переход от одного поля к другому осуществляется путем замены $\mathbf{E}_a \rightarrow \pm \mathbf{H}_d, -\mathbf{H}_a \rightarrow \pm \mathbf{E}_d$ [14]. В рассеянии света диском участвуют заряды и токи, индуцированные падающей волной. Внутри области, ограниченной диафрагмой, зарядов нет, но, исходя из

аналогии в характере решения двух задач, можно ввести фиктивные магнитные заряды и по их распределению искать решение по той же схеме, что и в задаче с диском.

Применив этот прием, Бете нашел ослабление интенсивности света, проходящего через малую диафрагму в бесконечно тонком идеально проводящем экране. Если I_0 – интенсивность поля падающей плоской волны, распространяющейся по нормали к экрану, то интенсивность поля I за экраном на фиксированном расстоянии $r \gg a$ от диафрагмы может быть представлена в виде

$$I = I_0 A_2(\theta_1) k^4 a^6 \quad \text{при } ak \ll 1, \quad (3.72)$$

где θ_1 – угол, определенный выше. Помимо изменения характера угловой зависимости интенсивности прошедшего света ($A_2(\theta_1) \neq A_1(\theta_1)$), формула (3.72) отличается от формулы (3.71) наличием дополнительного множителя $(ak)^2$. Малая диафрагма пропускает существенно меньше света, чем можно ожидать из чисто геометрических соображений, экстраполируя формулу Кирхгофа в область $ak \ll 1$.

Для практических целей наибольший интерес представляет расчет коэффициента пропускания η_a диафрагмы, определяемого отношением мощностей прошедшего и падающего на диафрагму световых пучков. Из формулы (3.72) следует, что при уменьшении радиуса диафрагмы прошедшая мощность ослабевает пропорционально $(ka)^4$ (дополнительный множитель a^2 исчезает при учете площади диафрагмы). В случае, когда неравенство $ak \ll 1$ перестает выполняться, но остается в силе неравенство $ak < 1$, прибегают к разложению η_a по степеням $(ak)^2$.

$$\eta_a \approx \frac{64}{27\pi^2} (ak)^4 [1 + 0,9(ak)^2 + \dots]. \quad (3.73)$$

Эту формулу с погрешностью порядка 10% можно использовать при $ak \leq 0,3$, сохраняя лишь два первых члена разложения. Так, при $\lambda = 600$ нм и $a = 30$ нм находим $t_a \approx 3 \cdot 10^{-3}$. Считая, как принято при расчете стандартных оптических систем, $\eta_a = 1$, мы допустили бы ошибку, превышающую два порядка.

Известно [8], что дипольный момент p малых частиц пропорционален их объему; так, для сферы радиуса a справедливо соотношение $p \sim a^3$. С учетом этой зависимости и формулы (3.70) поток энергии, рассеянный малой

сферой, можно представить в виде $S = S_0 k^4 a^6$, где для сокращения записи в коэффициент S_0 включены все множители, не зависящие от a и k .

Аналогия с зависимостью, предсказываемой теорией Бете, не является случайной. Анализ [14] показывает, что излучение, пропущенное диафрагмой, по мощности и угловому распределению эквивалентно излучению двух диполей – электрического и магнитного, центры которых совпадают с центром диафрагмы, а моменты равны соответственно

$$\mathbf{p}_e = (a^3/3\pi)\mathbf{E}_0, \quad \mathbf{p}_m = (-2a^3/3\pi)\mathbf{H}_0, \quad (3.74)$$

где $\mathbf{E}_0, \mathbf{H}_0$ – поле падающей волны. Первый диполь ориентирован вдоль оси диафрагмы, а второй – параллельно плоскости экрана. Диафрагма рассеивает в прямом направлении примерно такую же мощность, как равновеликая сфера из материала с $\epsilon_r = 2$.

Взаимодействие электромагнитного поля с малыми ($a \leq \lambda/10$) частицами известно как рэлеевское рассеяние. Его характерной особенностью является зависимость интенсивности рассеянного света от частоты вида $I \sim \omega^4$, приводящая к преобладанию коротковолновой части рассеянного белого спектра (классический пример – голубой цвет неба). Тот же закон выполняется и при прохождении света через малую диафрагму.

Причиной ослабления света диафрагмой является взаимодействие падающей волны с квазисвободными электронами в металлическом экране. Колебания электронов и циркуляция токов в металле способствуют гашению прямой волны за счет усиления отраженной волны. Очевидно, эти эффекты будут ослаблены при удалении части материала экрана или при образовании сквозного отверстия. Однако при малом размере отверстия их нельзя устранить полностью, так как индуцированные поля проникают в область, ограниченную диафрагмой. По своим оптическим свойствам эта область близка к металлу с несколько меньшим коэффициентом поглощения, чем материал экрана.

Сходство диафрагмы с двумя рассеивающими диполями отчетливо проявляется вдали от нее, но исчезает по мере приближения к экрану. Структура поля в непосредственной близости от диафрагмы, т.е. в плоскости $z = +0$ (при координате экрана $z = 0$), описывается формулой Бете-Боукампа [14], которую правильнее было бы называть совокупностью формул для различных составляющих векторов \mathbf{E} и \mathbf{B} . Она является по существу единственным в ближнепольной оптике соотношением, полученным на основе последовательного аналитического подхода. Повышенный интерес к

формуле Бете-Боукампа вызван еще и тем, что многие предсказанные ей результаты получили экспериментальные подтверждения.

Согласно теории Бете-Боукампа составляющие электрического вектора \mathbf{E} в плоскости $z = +0$ в цилиндрических координатах имеют вид

$$\begin{aligned} E_\rho &= -E_0 \frac{8ik_0 a}{3\pi} \frac{1 - 0,5\xi^2}{(1 - \xi)^{1/2}} \cos \varphi, \\ E_\varphi &= E_0 \frac{8ik_0 a}{3\pi} (1 - \xi^2)^{1/2} \sin \varphi, \\ E_z &= 0. \end{aligned} \quad (3.75)$$

где ρ – радиальное расстояние от центра диафрагмы, φ – азимутальный угол, $\xi \equiv a/\rho$, k_0 – волновое число в вакууме, E_0 – амплитуда плоской волны, распространяющейся в вакууме (воздухе) по нормали к экрану и поляризованной в плоскости xz .

Хотя соотношения (3.75) получены при расчете поля вблизи диафрагмы в идеальном плоском экране, они, как показывает эксперимент, правильно описывают основные особенности структуры поля вблизи вершины реального апертурного зонда. Остановимся на некоторых следствиях формул (3.75).

1. Вопреки основному допущению теории Кирхгофа, используемой при расчетах стандартных оптических систем, электрическое поле \mathbf{E} непосредственно за малой диафрагмой отличается от поля падающей волны \mathbf{E}_0 . Более того, однородное поле при переходе через плоскость диафрагмы приобретает сложную структуру, чувствительную к направлению поляризации.

2. При $\xi \rightarrow 1$ $E_\varphi \rightarrow 0$, а значит, выполняется необходимое для идеального металла требование исчезновения тангенциальной составляющей электрического вектора на границе. В то же время для направлений, отличных от $\varphi = \pi/2$, радиальная составляющая E_ρ вектора \mathbf{E} при приближении к краю диафрагмы бесконечно растет. Сингулярность решения ($E_\rho \rightarrow \infty$) характерна для задач, в которых рассматривается дифракция на экране с резкой границей. В реальной ситуации резкость перехода сглаживается из-за конечной проводимости материала экрана, но сохраняются характерные всплески напряженности поля вблизи краев диафрагмы вдоль направления поляризации падающей волны.

3. В плоскости zy ($\varphi = \pi/2$), перпендикулярной плоскости поляризации, $E_\rho = 0$ и остается только азимутальная составляющая E_φ , но на

границе обращается в ноль и она. Поле, таким образом, оказывается сконцентрированным вблизи оси, а световой пучок в этом сечении сжимается до размеров, меньших диаметра диафрагмы.

3.10. Усиление электромагнитного поля у вершины зонда

Помимо сжатия светового пучка малой диафрагмой существует и другая возможность концентрации электромагнитного поля в области субволновых размеров. Она основана на использовании известного в электростатике “эффекта громоотвода” – сгущения силовых линий у заостренного конца отрезка металлического провода. Этот эффект может многократно усиливаться в динамическом поле, частота которого настроена в резонанс с колебаниями электронной плазмы в металле. Оценку коэффициента усиления поля проще всего получить, аппроксимируя форму кончика иглы сферой или эллипсоидом вращения.

Рассмотрим металлический эллипсоид, большая и малая полуоси которого равны соответственно a и b , помещенный в однородное внешнее поле \mathbf{E}_0 , направление которого совпадает с направлением оси a . Зависимость диэлектрической проницаемости металла $\varepsilon = \varepsilon' + i\varepsilon''$ от частоты ω по теории Друде определяется соотношением

$$\varepsilon(\omega) \approx 1 + \omega_p^2 / (i\gamma\omega - \omega^2), \quad (3.76)$$

где ω_p – частота колебаний электронной плазмы и γ – коэффициент затухания колебаний, учитывающий взаимодействие электронов с фононами и дефектами решетки. Если $a, b \ll \lambda$, можно пренебречь запаздыванием потенциала. При этом волновое уравнение переходит в уравнение Лапласа, решение которого дает локальное поле у вершины эллипсоида

$$\mathbf{E} = \varepsilon \mathbf{E}_0 / [1 + (\varepsilon - 1)A], \quad (3.77)$$

где A – коэффициент деполяризации, зависящий от a и b . Эта зависимость имеет наиболее простой вид в двух случаях

$$A = 1/3 \quad \text{при } a = b, \quad (3.78a)$$

$$A = (b/a)^2 \ln(b/a) \quad \text{при } a \gg b. \quad (3.78b)$$

Преобразуя формулу (3.76) к виду

$$\varepsilon - 1 = \frac{\omega_p^2}{\omega^2} \frac{i\gamma\omega + \omega^2}{\gamma^2 + \omega^2}, \quad (3.79)$$

и выделяя вещественную и мнимую части выражения справа, легко убедиться, что резонансу соответствует частота

$$\omega_r = \sqrt{\omega_p^2 A - \gamma^2}. \quad (3.80)$$

Напряженность поля на этой частоте при $A \ll 1$ на основании формул (3.77), (3.79), (3.80) определяется соотношениями

$$\mathbf{E} = -i \frac{\varepsilon \mathbf{E}_0}{\varepsilon''(\omega_c) A} \approx \frac{1}{A} \left(1 + i \frac{\omega_r}{\gamma} \right) \mathbf{E}_0. \quad (3.81)$$

Резонанс наиболее ярко выражен в металлах с высокой концентрацией свободных электронов. Для массивных образцов Ag $\omega_p/\gamma \approx 30$ и, если бы это отношение оставалось неизменным и в малых частицах, то при использовании очень тонкой иглы с $a/b = 10$ можно было бы рассчитывать на усиление напряженности локального поля в 10^3 раз. Реальный эффект примерно на порядок ниже вследствие того, что при уменьшении размеров частицы коэффициент затухания γ растет из-за рассеяния электронных волн на поверхности. Оптимальный размер частиц Ag близок к 30 нм. При адсорбции молекул на шероховатой пленке Ag усиление локального поля на микровыступах приводит к резкому увеличению интенсивности комбинационного рассеяния света. На этом основан эффект, получивший название “гигантское комбинационное рассеяние” (в иностранной литературе используется термин “SERS” – усиленное поверхностью рамановское рассеяние).

При сближении острия с образцом эффективная диэлектрическая проницаемость среды меняется. Соответственно и условия резонанса становятся чувствительными не только к параметрам острия, но и к состоянию всей системы в целом, в том числе к ширине зазора острие-образец. Под величиной ε в формулах (3.79), (3.81), строго говоря, следует понимать некоторую усредненную величину, зависящую от диэлектрических проницаемостей острия и образца, а также от расстояния между ними. Не

вдаваясь в особенности этой зависимости, которая может быть достаточно сложной, отметим лишь, что она представляет собой типичное проявление ближнепольного эффекта [15].

На практике приходится учитывать то обстоятельство, что лазерный пучок даже при оптимальной фокусировке взаимодействует с образцом на площади, существенно превышающей размеры участка, находящегося непосредственно под острием. Чтобы отделить относительно слабый эффект действия острия на поверхность, проще всего промодулировать сигнал колебаниями острия на частоте ~ 1 кГц и затем выделить полезную составляющую сигнала системой синхронного детектирования. Указанный метод широко используется в отражательных БСОМ с металлическими или кремниевыми остриями и позволяет получать информацию об оптических свойствах участков размером порядка радиуса кривизны острия.

В случае, когда локализация поля острием используется с целью модификации образца, например, для нанесения лазерных меток на поверхность, условия облучения следует подбирать так, чтобы интенсивность поля была достаточной для изменения состояния поверхности только на участках непосредственно под острием. Освещая острие и образец импульсным лазером, можно добиться образования ямок или холмиков размером до нескольких десятков и даже единиц нанометров [5]. В модификации поверхности могут принимать участие различные процессы, в частности, связанные с локальным плавлением и испарением материалов, а также с переносом вещества с острия на образец.

3.11. Определение термина “ближнее поле”. Механизм переноса энергии в системе источник – зонд - детектор

Мы убедились на различных примерах, что затухающие волны представляют собой столь же универсальное явление, как и распространяющиеся. Излучающий или рассеивающий объект с радиусом $a \ll \lambda$ создает вокруг себя циркуляцию электромагнитной энергии. Эта энергия сконцентрирована в области, размер которой близок к размеру излучателя. В случае, когда распространяющиеся волны наталкиваются на непрозрачный барьер – среду, в которой при данных условиях их существование невозможно, поле проникает в барьер на расстояние порядка длины волны.

Ближним полем источника (или совокупности элементарных источников, если речь идет о границе раздела двух сред) называется область существования затухающих волн. Протяженность δ этой области различна для разных объектов. Каждой пространственной частоте f в спектре рельефа неоднородной поверхности соответствует своя глубина δ_f проникновения затухающего поля.

Понятия “ближнее поле” и “ближняя зона” не являются тождественными. Ближней зоной осциллятора длиной l принято считать область, точки которой удалены от осциллятора на расстояние r , определяемое неравенствами $l \ll r \ll \lambda/2\pi$ (см. разд. 3.8.). При таком определении как внешняя, так и внутренняя границы области оказываются существенно размытыми.

Фактически всю часть пространства, находящуюся за пределами ближнего поля и, следовательно, содержащую только распространяющиеся волны, можно называть “дальним полем” или “дальней зоной”. Границу дальнего поля со стороны малых r обычно определяют неравенством $r > \lambda$. С внешней стороны дальнее поле ограничено лишь чувствительностью приемника излучения.

Центральным для БПО является вопрос: каким образом наблюдатель, находящийся в дальнем поле, может получить информацию о структуре ближнего поля, которое не излучает энергию и, следовательно, представляется ему неким подобием “черной дыры”? Ответ на этот вопрос содержится в способности электромагнитных волн переходить из одной формы в другую.

Любой малый объект, помещенный в световой поток, создает вокруг себя затухающее поле, превращая тем самым часть энергии в скрытую форму. Таким объектом может быть молекула люминофора, диафрагма в

непрозрачном экране, малая рассеивающая свет частица и т.д. Исходя из обратимости оптических явлений, можно ожидать, что если малый объект ввести в затухающее поле, он преобразует часть его энергии в излучение. Доказательство способности малых (радиусом $a \leq \lambda/2$) тел частично преобразовывать электромагнитную энергию из одной формы в другую содержится в теореме Вольфа – Нието Весперинеса [16], сформулированной вскоре после изобретения БСОМ.

Затухающее поле может существовать и вблизи протяженных объектов с шероховатой поверхностью, спектр пространственных частот которой содержит высокочастотные компоненты. Из дальнего поля мелкие неровности не видны. Дифракционная решетка с 10 000 штрихов/мм в отраженном свете выглядит, как гладкая поверхность. Но зонд, перемещаемый в непосредственной близости от решетки, способен обнаружить ее периодичность, проявляющуюся в структуре ближнего поля

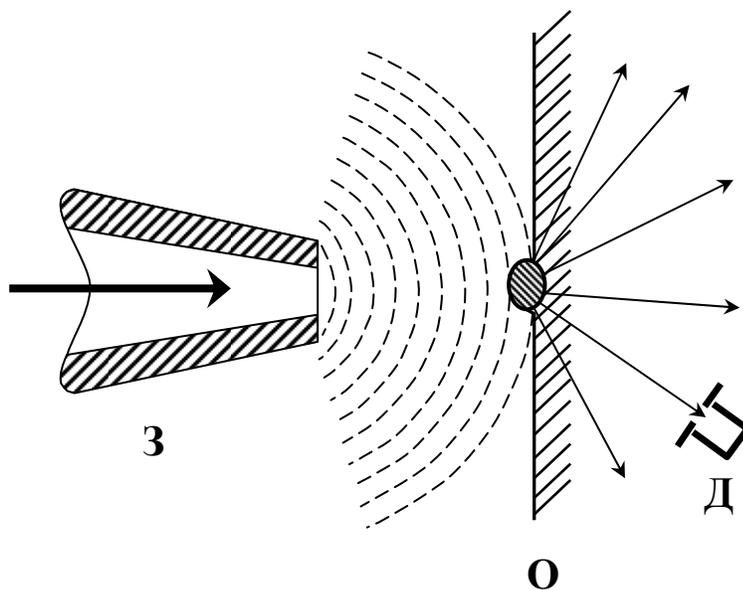


Рис. 9. Схема переноса излучения в системе зонд-образец-детектор. Затухающие волны вблизи вершины зонда преобразуются в распространяющиеся малой частицей на поверхности образца.

Малая частица, находящаяся вблизи поверхности образца и освещаемая внешним источником, создает вокруг себя затухающее поле. Приближение

другой малой частицы (вершины зонда) высвобождает часть скрытой энергии, преобразуя ее в излучение, регистрируемое детектором. Схема переноса энергии симметрична и процесс ее преобразования может быть направлен в другую сторону: внешний источник возбуждает затухающее поле зонда, а малый объект преобразует часть энергии этого поля в излучение (Рис. 9).

Обратимость характерна и для туннелирования частиц через барьер симметричной формы. Перенос энергии от источника к детектору можно рассматривать как результат туннелирования фотонов через промежуток острие-образец. Устанавливая детектор под различными углами к оси прибора, можно менять эффективную высоту барьера.

До сих пор мы не учитывали того обстоятельства, что рассеивающая свет малая частица излучает электромагнитные волны в окружающее пространство и при отсутствии зонда, находящегося в непосредственной близости от нее. По интенсивности этих волн наблюдатель в дальнем поле, в принципе, может судить о размере частицы (или отверстия в экране). Но если в световой луч попадает несколько частиц, каждая из них на изображении, построенном по классической схеме, будет размыта в кружок с минимальным размером $\lambda/2$. По виду кружка рассеяния трудно судить о количестве элементарных источников, находящихся рядом друг с другом. Точно так же, измеряя при отсутствии образца излучение, пропущенное находящейся на вершине зонда диафрагмой, нельзя сказать ничего определенного об изображающих свойствах апертурного зонда, которые, помимо размера и формы диафрагмы, зависят от ряда других факторов, в том числе, от толщины покрытия, формы микровыступов на торце зонда и т.д.

При образовании изображения в БСОМ попадание на детектор прямого (без промежуточного перехода энергии из одной формы в другую) излучения, проходящего через образец или отраженного от образца, создает фон, уменьшающий контраст мелких деталей. Полезный сигнал, содержащий информацию о структуре ближнего поля образца, несет лишь излучение, возникшее при возмущении этого поля зондом. Симметрия схемы переноса излучения от источника к детектору требует обязательного участия в процессе преобразования энергии двух партнеров: образца и зонда.

Приближение зонда вплотную к образцу не может не сказываться на локальной структуре электромагнитного поля. В связи с этим изображение следует рассматривать как картину распределения интенсивности поля в системе зонд-образец в целом. Каждая смена зонда требует введения корректив в алгоритм расшифровки принимаемого сигнала.