

## 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: быстродействием, 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  $\lambda$ ) 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  $xy$  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 вспомогательным 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.

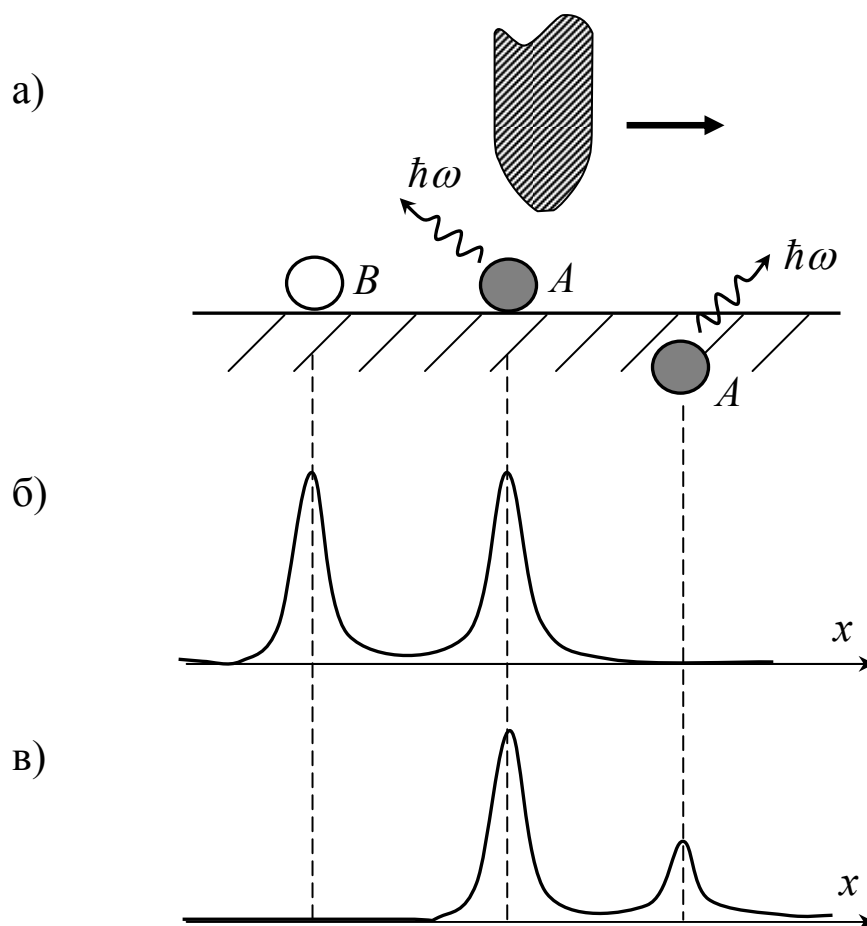


Рис.1 а) Схема сканирования зондом образца, содержащего флуоресцентные метки (А) и оптически нейтральную частицу (В); б, в) Изображения выделенного сечения образца в каналах АСМ (б) и БСОМ (в); кривая записи оптического сигнала приведена с поправкой на изменение положения острия по высоте.

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.

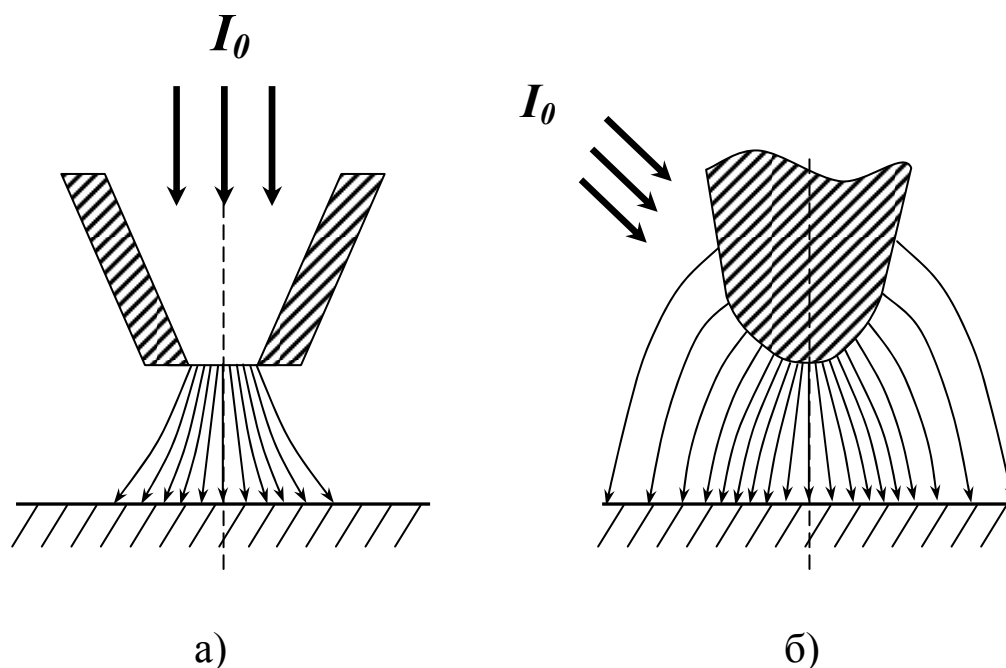


Рис.2 Схема рассеяния электромагнитной энергии  
вершинами апертурного а) и безапертурного б) зондов.

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.



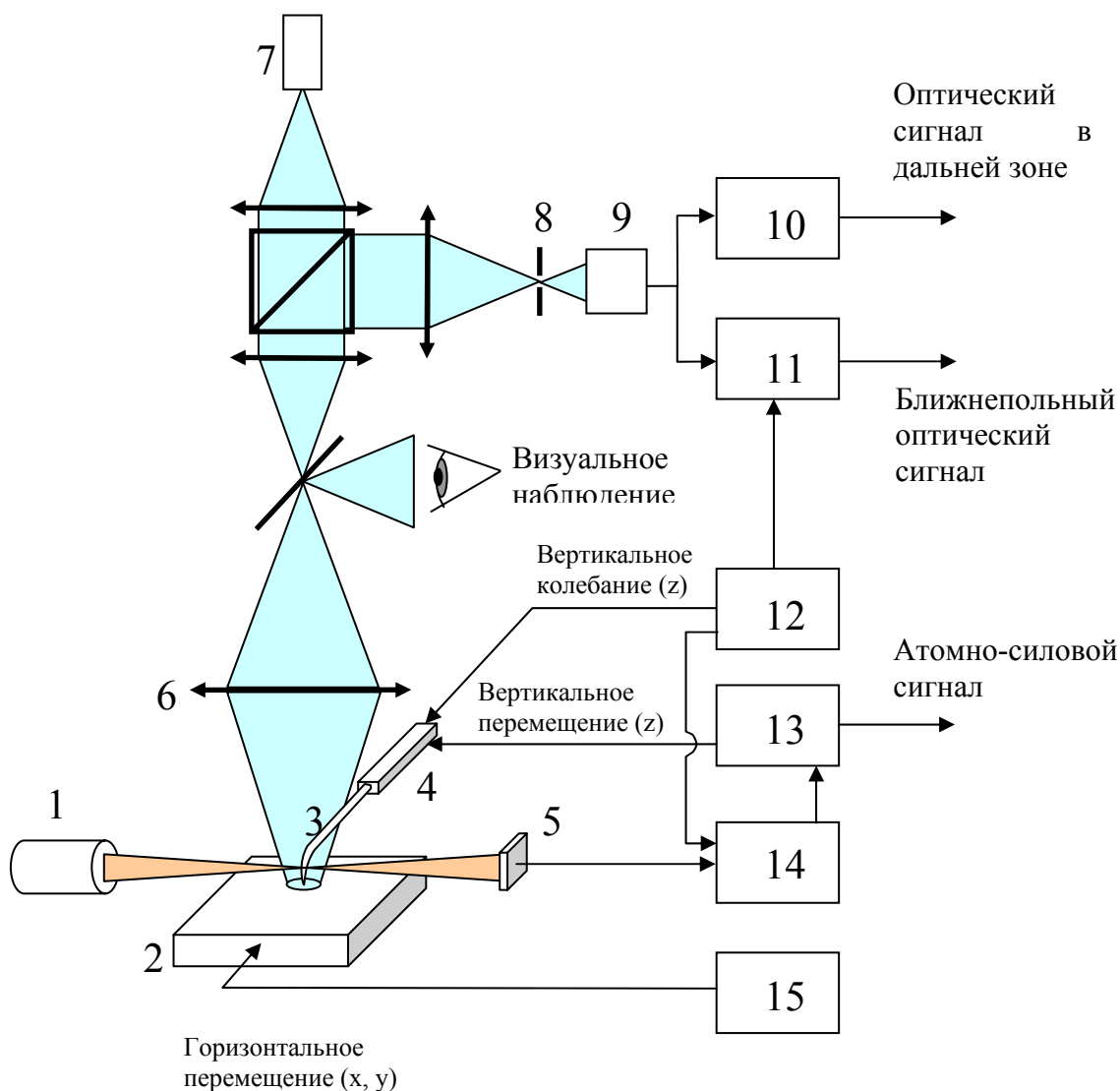


Рис.3 Принципиальная схема безапертурного БСОМ, работающего в отражательном режиме [7]: 1 – лазер для контроля зонда; 2 – образец; 3 – зонд; 4 – пьезо-сканер; 5, 9 – фотоприемники; 6 – объектив; 7 – зондирующий лазер; 8 – диафрагма; 10 – фильтр нижних частот; 11, 14 – синхронный усилитель; 12 – генератор;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

Существование 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,  $\omega$  - 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,  $\lambda$  – 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  $\lambda$ , 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  $\alpha, \beta, \gamma$  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  $\nu$  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  $\nabla$  – векторный оператор градиента с проекциями  $\partial/\partial x$ ,  $\partial/\partial y$ ,  $\partial/\partial z$  в декартовых координатах,  $\nabla^2$  – 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 **E**, **H** и индукций **D**, **B** электрического и магнитного полей с плотностями зарядов  $\rho$  и токов **j** [8]. Для упрощения записи обычно оставляют по одному из пары векторов. Будем предполагать, что среда является однородной и изотропной. Тогда справедливы соотношения **D** =  $\epsilon$ **E** =  $\epsilon_r \cdot \epsilon_0$ **E**, **B** =  $\mu$ **H** =  $\mu_r \cdot \mu_0$ **H**, где  $\epsilon$  ( $\epsilon_r$ ) и  $\mu$  ( $\mu_r$ ) – абсолютная (относительная) диэлектрическая и магнитная проницаемости среды. Выбрав за основу векторы **E** и **B** и учитывая, что в оптических средах  $\mu_r = 1$ , а постоянная  $\mu_0$  может быть представлена в виде  $\mu_0 = 1/(\epsilon_0 \cdot c^2)$ , где  $\epsilon_0$  – диэлектрическая проницаемость вакуума и  $c$  – скорость света в вакууме, приходим к системе уравнений

$$\text{div} \mathbf{E} = \rho / \epsilon, \quad (3.6)$$

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

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

$$\epsilon_0 c^2 \text{rot} \mathbf{B} = \mathbf{j} + \epsilon \partial \mathbf{E} / \partial t. \quad (3.9)$$

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

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

$$\begin{aligned} \text{rot}_x \mathbf{Q} &= \partial Q_z / \partial y - \partial Q_y / \partial z, \\ \text{rot}_y \mathbf{Q} &= \partial Q_x / \partial z - \partial Q_z / \partial x, \\ \text{rot}_z \mathbf{Q} &= \partial Q_y / \partial x - \partial Q_x / \partial y. \end{aligned} \quad (3.11)$$

Помимо напряженностей электрического и магнитного полей в уравнениях электродинамики присутствуют скалярный  $\phi$  и векторный **A** потенциалы. Вид первого из них непосредственно вытекает из закона Кулона

$$\phi = (1/4\pi\epsilon_0) \int_V (\rho/r) dV, \quad (3.12)$$

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  $\rho dV$ , находящимися в объеме  $V$ . Векторный потенциал записывается в аналогичной форме с той разницей, что вместо плотности зарядов  $\rho$  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  $\phi$  и  $\mathbf{A}$ , имеет вид

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

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

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

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

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  $\mathbf{E} = \mathbf{E}(\mathbf{r}, t)$ ,  $\nabla^2$  – 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  $\mathbf{B}$ ,  $\varphi$ ,  $\mathbf{A}$ . 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  $\gamma$  – 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  $\gamma$  коэффициентом  $\chi = n\kappa = \gamma\lambda_0/4\pi$ , где  $\lambda_0$  – 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

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

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  $k_1 = k$ ,  $k_2 = -2\pi n\kappa/\lambda_0 = -\kappa k$ . 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  $\theta$  и  $\theta_r$  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  $\kappa_z$  – действительное положительное число, равное

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

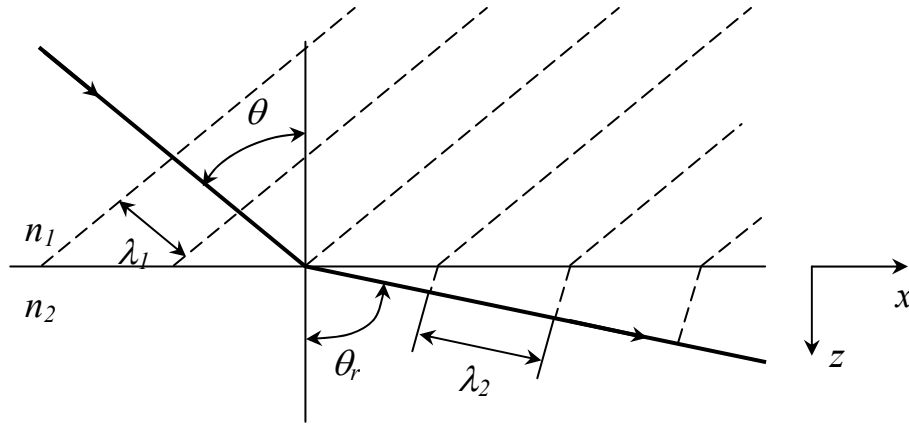


Рис. 4. Преломление светового луча при переходе из среды 1 в среду 2;  $n_1 > n_2$ . Штриховыми линиями показаны мгновенные положения гребней волны.

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  $\delta$  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  $\theta$  поле проникает в запрещенную область на расстояние того же порядка, но при  $\theta \rightarrow \theta_c$   $\delta$  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  $\Psi$ , 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  $\Psi$  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  $\Psi$  (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  $d$ .

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

Зависящие от координаты  $z$  части  $\psi_1$  и  $\psi_2$  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  $\lambda_1, \lambda_2$  – 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  $\psi_1$  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  $\psi_1$  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

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

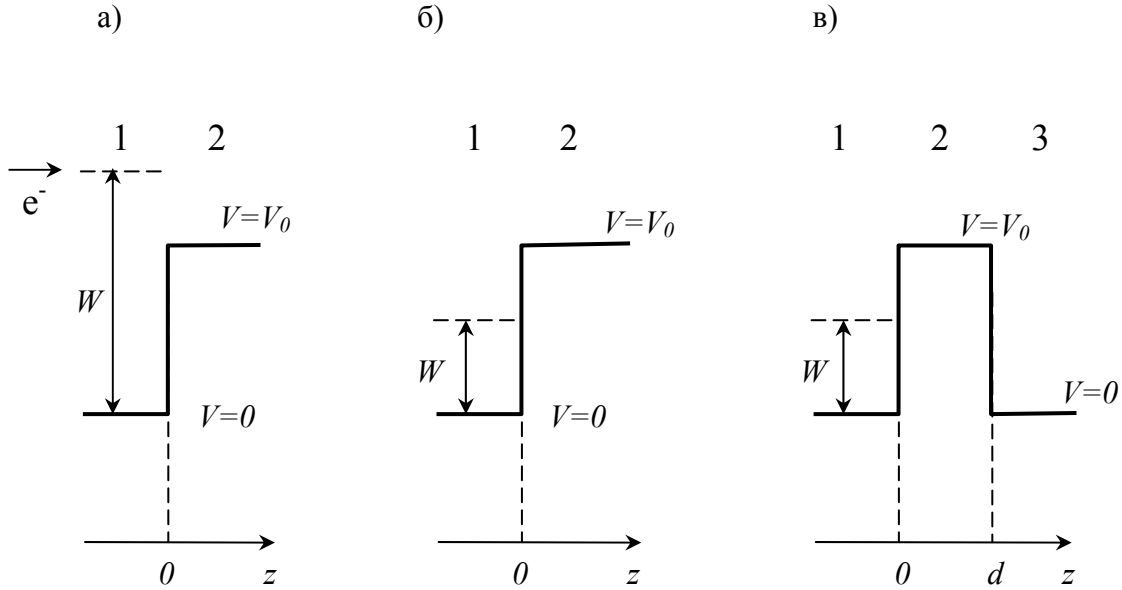


Рис. 5. К объяснению взаимодействия электрона с энергией  $W$  с прямоугольным потенциальным барьером:

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.

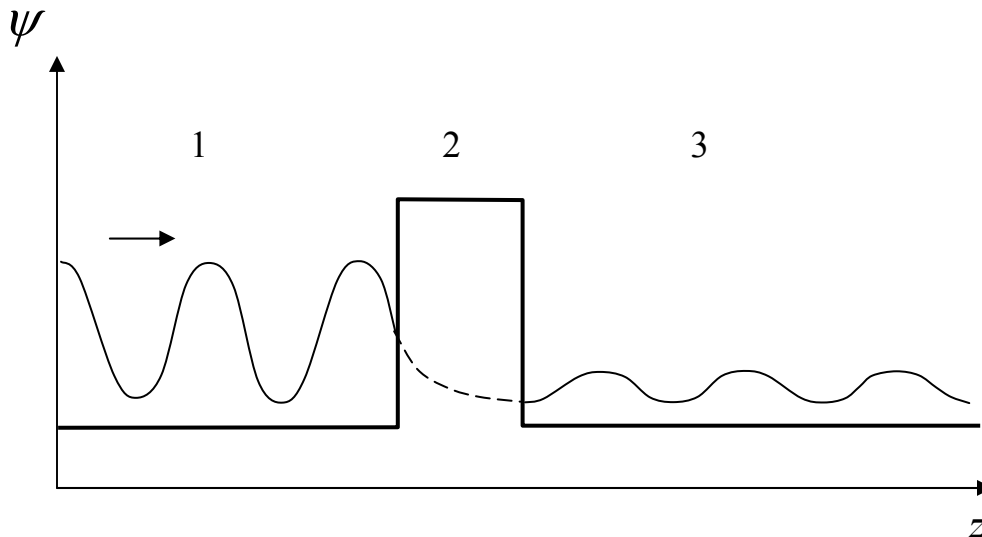


Рис. 6. Прохождение волны через барьер.

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

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

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  $d$

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

где  $D_0 \approx 1$ , а  $\kappa$  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.

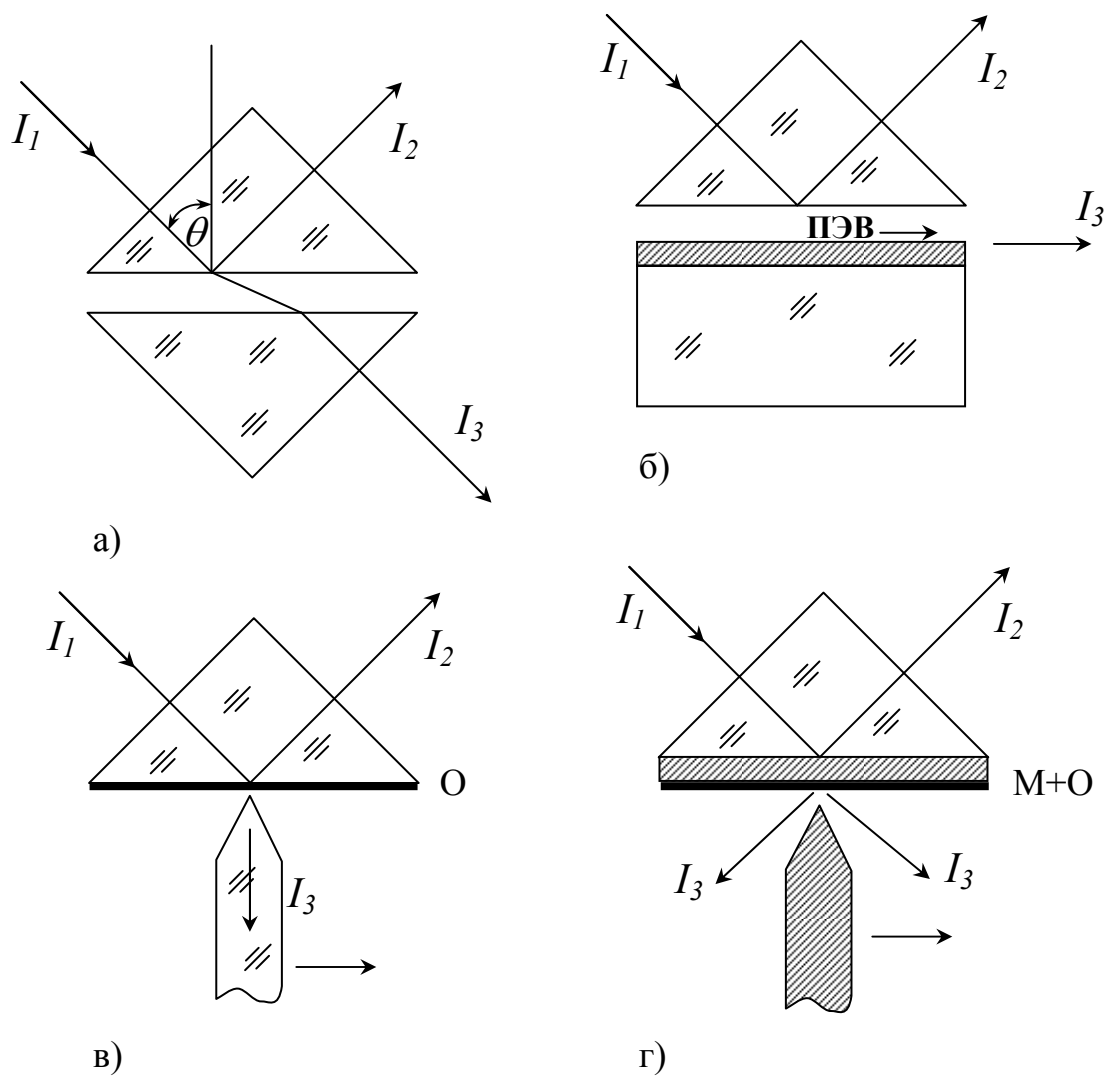


Рис. 7. Принципиальные схемы опытов с нарушенным полным внутренним отражением:

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  $\kappa_1$  и  $\kappa_2$  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  $\delta$  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 (моду) 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,  $\lambda_z$  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 придти и из формулы (3.54). При уменьшении  $\omega$  или  $a$   $k_z$  из действительной величины становится мнимой. Критическая частота  $\omega_c$  определяется равенством

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

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

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

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, укладывающихся на отрезке  $2a$ . С равным основанием мы могли бы выбрать за направление **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 **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  $\lambda_0$  на  $\lambda_0/n$  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  $k_z = i\kappa$ , где  $\kappa$  – действительная положительная величина, связанная с  $u_{lm}$  соотношением

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

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 ТЕ<sub>11</sub> и ТМ<sub>01</sub> 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 сердцевинной 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  $\mathbf{r}$ ,  $\theta$  – 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  $\omega$ , 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  $\varphi$  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  $pz = \mathbf{p} \cdot \mathbf{r}$  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  $\varphi$  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

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

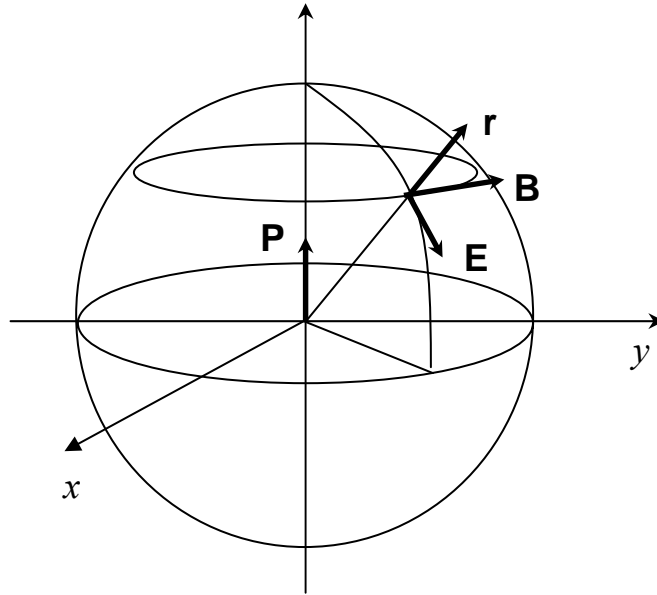


Рис. 8. Взаимное расположение векторов  $\mathbf{r}$ ,  $\mathbf{E}$ ,  $\mathbf{B}$  в дальней зоне осциллятора.

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,  $\lambda$  – 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  $\theta_1$  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  $\theta_1$  – 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  $\eta_a$  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  $\eta_a$  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  $\varepsilon_r = 2$ .

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 **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  $\rho$  – радиальное расстояние от центра диафрагмы,  $\varphi$  – 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 **E** непосредственно за малой диафрагмой отличается от поля падающей волны **E**<sub>0</sub>. 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  $E_\rho$  вектора **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  $E_\varphi$ , 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  $\omega$  по теории Друде определяется соотношением

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

где  $\omega_p$  – частота колебаний электронной плазмы и  $\gamma$  – 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  $\gamma$  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  $\varepsilon$  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  $\sim 1$  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  $\delta$  этой области различна для разных объектов. Каждой пространственной частоте  $f$  в спектре рельефа неоднородной поверхности соответствует своя глубина  $\delta_f$  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

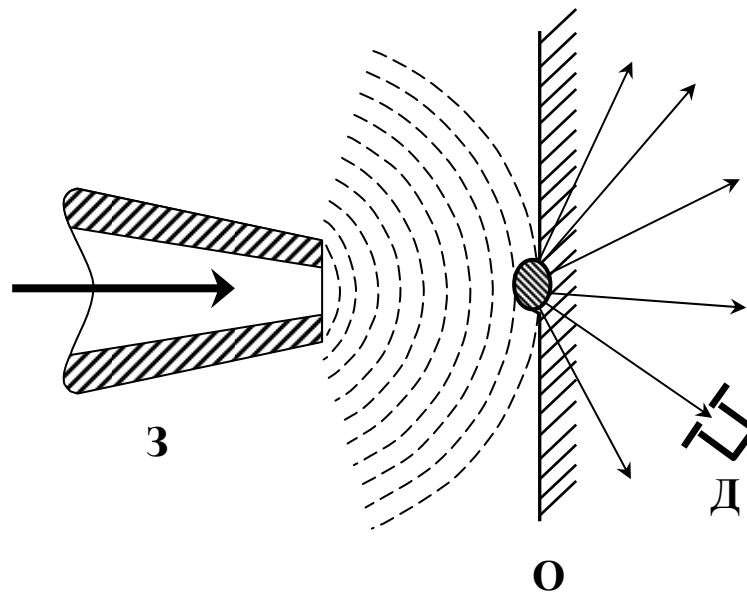


Рис. 9. Схема переноса излучения в системе зонд-образец-детектор. Затухающие волны вблизи вершины зонда преобразуются в распространяющиеся малой частицей на поверхности образца.

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.

## 4. Элементы фурье-оптики ближнего поля

### 4.1. Метод разложения по плоским волнам

Согласно теореме Фурье периодическая функция  $f(x) = f(x+2\pi)$ , удовлетворяющая определенным условиям (которым соответствует практически любая функция, встречающаяся в физических исследованиях), может быть представлена тригонометрическим рядом

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx), \quad (4.1)$$

с коэффициентами

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx, \quad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx dx. \quad (4.2)$$

Обычно бывает удобнее заменить тригонометрический ряд эквивалентным комплексным рядом

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \exp(inx), \quad (4.3)$$

где

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\alpha) \exp(-in\alpha) d\alpha. \quad (4.4)$$

При помощи замены переменных разложение Фурье может быть видоизменено так, чтобы изображать функцию  $f(x)$  в области  $-l \leq x \leq l$

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \exp\left(\frac{in\pi}{l} x\right), \quad (4.5)$$

$$\text{где} \quad c_n = \frac{1}{2l} \int_{-l}^l f(\alpha) \exp\left(-\frac{in\pi}{l} \alpha\right) d\alpha. \quad (4.6)$$

Переход в формулах (4.5), (4.6) к пределу при  $l \rightarrow \infty$  позволяет представить непериодическую функцию  $f(x)$  двойным интегралом

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} f(\alpha) \exp[iu(x - \alpha)] d\alpha. \quad (4.7)$$

Это выражение можно разбить на два симметричных равенства

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(u) \exp(iux) du \equiv F^{-1}\{g(u)\}, \quad (4.8)$$

$$g(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) \exp(-iux) dx \equiv F\{f(x)\}, \quad (4.9)$$

где символы  $F$ ,  $F^{-1}$  обозначают соответственно прямое и обратное преобразования Фурье. Функцию  $g(u)$  называют спектральной плотностью функции  $f(x)$ . Вспомним (см. разд. 3.1), что распространение плоской волны вдоль оси  $x$  создает возмущение, которое в каждый момент времени может быть представлено в виде  $U = U_0 \exp(\pm i k_x x)$ . Формулы (4.8), (4.9) содержат члены такого же вида. Следовательно, их можно рассматривать как представление произвольного возмущения  $f(x)$  набором плоских волн с амплитудами  $g(k_x) dk_x / \sqrt{2\pi}$  для волновых чисел в интервале  $(k_x, k_x + dk_x)$ . Такое представление возможно для линейных систем, реакция которых на одновременное воздействие нескольких источников возмущения равна сумме реакций на действие отдельных источников.

Аналогичным образом определяется двумерное преобразование Фурье, отличающееся от (4.8), (4.9) лишь тем, что интегрирование проводится по двум координатам  $x, y$  и двум составляющим волнового вектора  $k_x, k_y$

$$g(x, y) = \frac{1}{2\pi} \iint G^\vee(k_x, k_y) \exp[i(k_x x + k_y y)] dk_x dk_y \equiv F^{-1}\{G^\vee(k_x, k_y)\}, \quad (4.10)$$

$$G^\vee(k_x, k_y) = \frac{1}{2\pi} \iint g(x, y) \exp[-i(k_x x + k_y y)] dk_x dk_y \equiv F\{g(x, y)\}. \quad (4.11)$$

Во многих случаях удобнее вместо  $k_x, k_y$  использовать переменные  $f_x = k_x/2\pi$ ,  $f_y = k_y/2\pi$  [17]. Эта замена приводит к соотношениям



$$g(x, y) = \iint G(f_x, f_y) \exp[2\pi i(f_x x + f_y y)] df_x df_y, \quad (4.12)$$

$$G(f_x, f_y) = \iint g(x, y) \exp[-2\pi i(f_x x + f_y y)] dx dy, \quad (4.13)$$

причем функции  $G(f_x, f_y)$  и  $G^\vee(k_x, k_y)$  связаны между собой следующим образом:  $G(f_x, f_y) = 2\pi G^\vee(k_x, k_y)$ .

Двумерные преобразования Фурье можно представить в более компактной векторной форме. Так, формулы (4.10), (4.11) приводятся к виду

$$g(\mathbf{r}^\diamond) = \frac{1}{2\pi} \iint G^\vee(\mathbf{k}^\diamond) \exp i(\mathbf{r}^\diamond \cdot \mathbf{k}^\diamond) d\mathbf{k}^\diamond, \quad (4.14)$$

$$G^\vee(\mathbf{k}^\diamond) = \frac{1}{2\pi} \iint g(\mathbf{r}^\diamond) \exp[-i(\mathbf{r}^\diamond \cdot \mathbf{k}^\diamond)] d\mathbf{r}^\diamond, \quad (4.15)$$

где  $\mathbf{r}^\diamond, \mathbf{k}^\diamond$  – соответственно проекции радиуса вектора  $\mathbf{r}$  и волнового вектора  $\mathbf{k}$  на плоскость  $xy$ :  $\mathbf{r} = \mathbf{r}^\diamond + \mathbf{e}_z z$ ,  $\mathbf{k} = \mathbf{k}^\diamond + \mathbf{e}_z k_z$  ( $\mathbf{e}_z$  – единичный вектор вдоль оси  $z$ ).

Используя разложение Фурье, можно по известной структуре поля в одной плоскости, например, в плоскости  $z = +0$  непосредственно за диафрагмой восстановить его структуру во всем полупространстве. Пусть возмущение в плоскости  $z = +0$  определяется функцией  $U_0(\mathbf{r}^\diamond)$ , имеющей двумерный фурье-образ  $A_0(\mathbf{k}^\diamond)$

$$U_0(\mathbf{r}^\diamond) = \frac{1}{2\pi} \iint A_0(\mathbf{k}^\diamond) \exp i(\mathbf{r}^\diamond \cdot \mathbf{k}^\diamond) d\mathbf{k}^\diamond. \quad (4.16)$$

Нас интересует вид функции  $U_z(\mathbf{k}^\diamond)$ , представляющей возмущение в плоскости на расстоянии  $z$  от исходной. Задача решается подстановкой в подинтегральное выражение в (4.16) множителя  $\exp(-ik_z z)$ , определяющего изменение амплитуды и фазы каждой из бесконечной последовательности плоских волн

$$U_z(\mathbf{r}^\diamond) = \frac{1}{2\pi} \iint A_z(\mathbf{k}^\diamond) \exp i(\mathbf{r}^\diamond \cdot \mathbf{k}^\diamond) d\mathbf{k}^\diamond, \quad (4.17)$$

$$\text{где } A_z(\mathbf{k}^\diamond) \equiv A_0(\mathbf{k}^\diamond) \exp(-ik_z z). \quad (4.18)$$

Характер изменения  $A_z(\mathbf{k}^\diamond)$  с расстоянием от исходной плоскости определяется соотношением модулей волнового вектора  $|\mathbf{k}| = k = n\omega/c$  и его горизонтальной составляющей  $|\mathbf{k}^\diamond| = k^\diamond$ . При  $k > k^\diamond$   $k_z^2 = k^2 - (k^\diamond)^2 > 0$  и,

следовательно,  $k_z$  является действительной величиной. В этом случае каждая плоская волна сохраняет неизменную амплитуду и меняется с расстоянием лишь по фазе. Такая ситуация рассматривается в классической фурье-оптике.

Однако учет только распространяющихся волн не позволяет полностью охарактеризовать возмущение. Необходимо суммировать вклады бесконечной совокупности плоских волн, в том числе таких, волновые векторы которых удовлетворяют неравенству  $k < k^\diamond$ . Произведя замену  $ik_z = \kappa = \sqrt{(k^\diamond)^2 - k^2}$  и выбирая, как и в других подобных задачах, положительный знак перед квадратным корнем, находим

$$A_z(k^\diamond) = A_0(k^\diamond) \exp(-\kappa z) \quad \text{при } k < k^\diamond. \quad (4.19)$$

Вследствие зависимости  $\kappa$  от  $k^\diamond$  промежуток зонд-образец в БСОМ осуществляет фильтрацию пространственных частот. При  $k^\diamond \gg k$  составляющая сигнала с пространственной частотой  $f = k^\diamond/2\pi$  ослабевает по интенсивности в  $e$  раз на расстоянии  $z = 1/(2\kappa) \approx 1/(4\pi f)$ . Отсюда следует, что для уверенного различения структуры с периодом  $d = 1/f$  зонд должен находиться на расстоянии  $z \approx 0,1d$  от образца.

Фурье-оптика ближнего поля решает задачу восстановления структуры электромагнитного поля по спектру пропускаемых зондом пространственных частот в диапазоне, включающем распространяющиеся и затухающие волны.

## 4.2. Частотный спектр шероховатой поверхности

К числу практических задач, решаемых методом Фурье, относится анализ поверхностного рельефа. Проиллюстрируем возможности метода на простейшем примере. Пусть функция  $f(x)$ , характеризующая отклонение точек рельефа от уровня гладкой поверхности, задана следующим образом:  $f(x) = |x|$  в промежутке  $[-\pi, \pi]$  (выбор любого другого симметричного интервала приводит лишь к необходимости замены переменной, как в формулах (4.5), (4.6)). Продолжив  $f(x)$  с периодом  $2\pi$  в обе стороны оси  $x$ , приходим к профилю треугольной решетки (Рис. 10). При разложении четной функции в тригонометрический ряд (4.1) коэффициенты  $b_n$  равны нулю. Используя при вычислении  $a_n$  интегрирование по частям, приходим к представлению  $f(x)$  в виде ряда

$$f(x) = \frac{1}{2}\pi - \frac{1}{\pi} \left( \frac{\cos x}{1^2} + \frac{\cos 3x}{3^2} + \frac{\cos 5x}{5^2} + \dots \right). \quad (4.20)$$

Поскольку амплитуды гармоник быстро убывают с ростом частоты, функцию  $f(x)$  с удовлетворительной точностью характеризует сумма лишь нескольких первых членов ряда. Как видно из Рис. 10, включение всего лишь двух гармоник позволяет правильно представить период и глубину модуляции профиля решетки. Добавление каждого следующего члена разложения повышает заостренность зубцов.

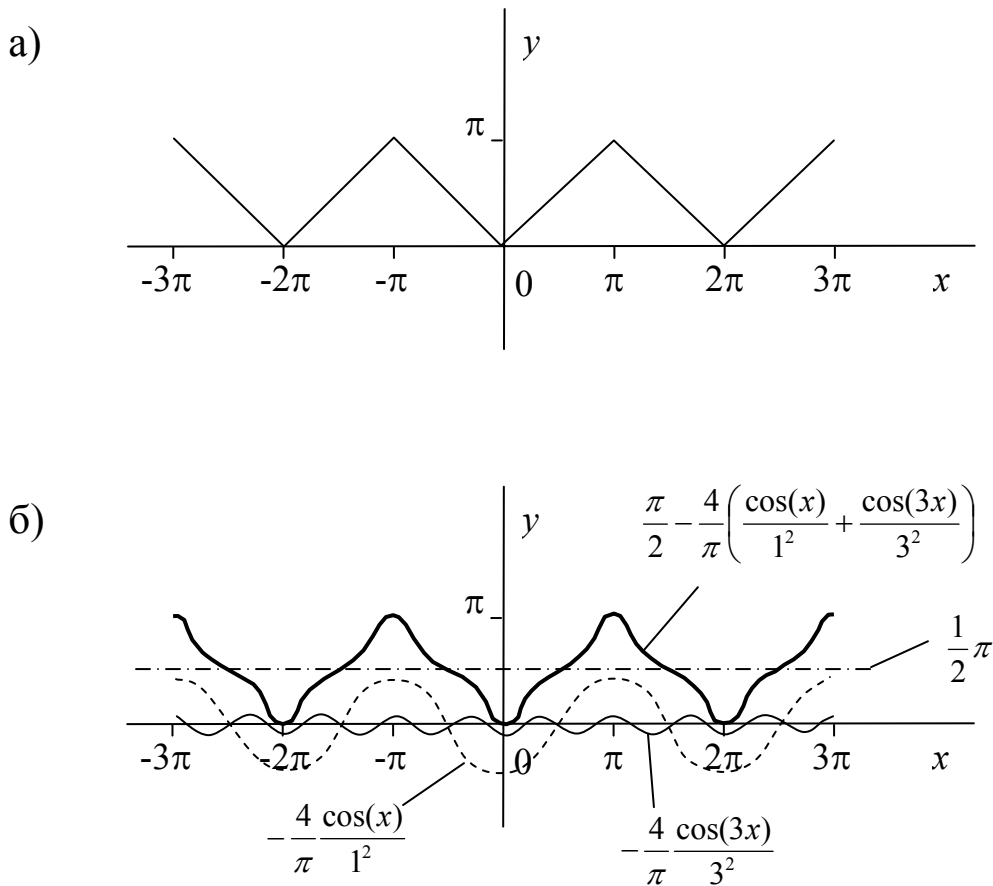


Рис. 10. Профиль решетки (а) и его представление суммой трех первых членов разложения Фурье (формула 4.20) (б).

Представим теперь, что профиль решетки анализируется предельно тонким зондом, перемещающимся поперек канавок. Из геометрических соображений очевидно, что восстановить на изображении любую из пространственных частот можно лишь в том случае, если размах колебаний зонда вдоль нормали к поверхности будет сопоставим с амплитудой соответствующей гармоники. С ростом порядкового номера гармоники амплитуда убывает и становится все сложнее отличить полезный сигнал от шумов.

Затухающие волны, переносящие высокочастотные составляющие спектра, помимо меньшей амплитуды характеризуются и меньшей глубиной проникновения в зазор образец-острие. Увеличение ширины зазора фактически эквивалентно ограничению числа членов разложения  $f(x)$ .

#### 4.3. Пары преобразований Фурье, часто встречающихся в прикладной оптике

Анализ преобразований Фурье упрощается, если  $f(x)$  является действительной четной функцией. В этом случае, используя формулу Эйлера  $\exp(ix) = \cos x + i \sin x$  и отбрасывая мнимые части экспоненциальных членов в формулах (4.8), (4.9), приходим к соотношениям

$$f(x) = \sqrt{2/\pi} \int_0^{\infty} g(u) \cos(ux) du, \quad (4.21)$$

$$g(u) = \sqrt{2/\pi} \int_0^{\infty} f(x) \cos(ux) dx. \quad (4.22)$$

Задача определения  $f(x)$  по известной  $g(u)$  и обратная задача легко решаются с помощью формул (4.21), (4.22), если функцию и ее фурье-образ удастся представить в виде комбинации некоторых элементарных функций, часто используемых при аппроксимации экспериментальных кривых.

1. Гауссова функция ошибок:  $f(x) = \exp(-a^2 x^2/2)$ .

Подстановка  $f(x)$  в формулу (4.22) приводит к табличному интегралу

$$g(u) = F\{f(x)\} = \frac{1}{a} \exp\left(-\frac{u^2}{2a^2}\right). \quad (4.23)$$

Аналогичным образом связаны функции двух переменных: если  $g(x, y) = \exp[-a^2(x^2 + y^2)/2]$ , то

$$G(u, v) = F\{g(x, y)\} = \frac{1}{a^2} \exp\left[-\frac{1}{2a^2}(u^2 + v^2)\right]. \quad (4.24)$$

## 2. Дельта-функция.

Она представляет собой бесконечно узкий импульс:  $\delta(x) = \lim_{a \rightarrow 0} s(a, x)$ , где  $s(a, x) = \frac{1}{a\sqrt{2\pi}} \exp\left(-\frac{x^2}{2a^2}\right)$ . Выбор множителя  $1/\sqrt{2\pi}$  диктуется условием нормировки:  $\int_{-\infty}^{\infty} s(a, x) dx = 1$ . С учетом (4.23)

$$g(u) = F\{\delta(x)\} = \frac{1}{\sqrt{2\pi}} \lim_{a \rightarrow 0} \left[ \exp\left(-\frac{a^2 u^2}{2}\right) \right] = \frac{1}{\sqrt{2\pi}}. \quad (4.25)$$

В двумерном пространстве  $\delta(x, y) = \lim_{a \rightarrow 0} \left[ \frac{1}{2\pi a^2} \exp\left(-\frac{x^2 + y^2}{2a^2}\right) \right]$ :

$$g(u, v) = F\{\delta(x, y)\} = \frac{1}{2\pi}. \quad (4.26)$$

Переходя к пространственным частотам, находим

$$G(f_x, f_y) = F\{\delta(x, y)\} = 1. \quad (4.27)$$

Таким образом, в частотном спектре возмущения в форме бесконечно узкого импульса, изображаемого  $\delta$ -функцией, в равной степени представлены все пространственные частоты в интервале  $(-\infty, +\infty)$ .

## 3. Прямоугольная функция

$$f(x) = \text{rect}(x) = 1 \text{ при } |x| \leq 1/2, \\ = 0 \text{ в остальных случаях.}$$

По формуле (4.22)  $g(u) = (1/\sqrt{2\pi})[\sin(u/2)/(u/2)]$  или

$$G(f_x) = \frac{\sin(\pi f_x)}{\pi f_x}.$$

Фурье-образ функции двух переменных  $g(x,y) = \text{rect}(x) \cdot \text{rect}(y)$  можно представить в виде

$$G(f_x, f_y) = F\{g(x, y)\} = \text{sinc}(\pi f_x) \cdot \text{sinc}(\pi f_y), \quad (4.28)$$

где  $\text{sinc}(x) \equiv (\sin x)/x$ .

#### 4. Круговая функция

$$\begin{aligned} \text{circ}(\rho) &= 1 \text{ при } \rho \leq 1, \\ &= 0 \text{ в остальных случаях.} \end{aligned}$$

График функции  $\text{circ}(\rho)$  можно построить, вращая прямоугольник вокруг оси, совмещенной с одной из сторон. Поэтому естественно ожидать, что и фурье-образы функций  $\text{circ}(\rho)$  и  $\text{rect}(x) \cdot \text{rect}(y)$  являются родственными функциями. Переход от прямоугольной симметрии к круговой, как обычно, требует замены тригонометрических функций бесселевыми. Эта замена приводит к соотношению

$$G(f) = F\{\text{circ}(\rho)\} = \frac{J_1(2\pi f)}{f},$$

где  $\rho, f$  – модули радиусов-векторов в двумерных пространствах соответственно координат и частот.

Рассмотренные пары преобразований Фурье для переменных  $x, y$  и  $f_x, f_y$  (или  $\rho$  и  $f$  в случае круговой симметрии) сведены в схему, в которой прямое и обратное преобразования изображены стрелками, направленными в противоположные стороны

$$\exp\left[-\frac{a^2}{2}(x^2 + y^2)\right] \leftrightarrow \frac{2\pi}{a^2} \exp\left[-\frac{2\pi^2}{a^2}(f_x^2 + f_y^2)\right], \quad (4.29)$$

$$\delta(x, y) \leftrightarrow 1, \quad (4.30)$$

$$\text{rect}(x) \cdot \text{rect}(y) \leftrightarrow \text{sinc}(\pi f_x) \cdot \text{sinc}(\pi f_y), \quad (4.31)$$

$$circ(\rho) \leftrightarrow \frac{J_1(2\pi f)}{f}. \quad (4.32)$$

Графики функций  $sinc(x)$  и  $J_1(x)/x$  изображены на Рис. 11.

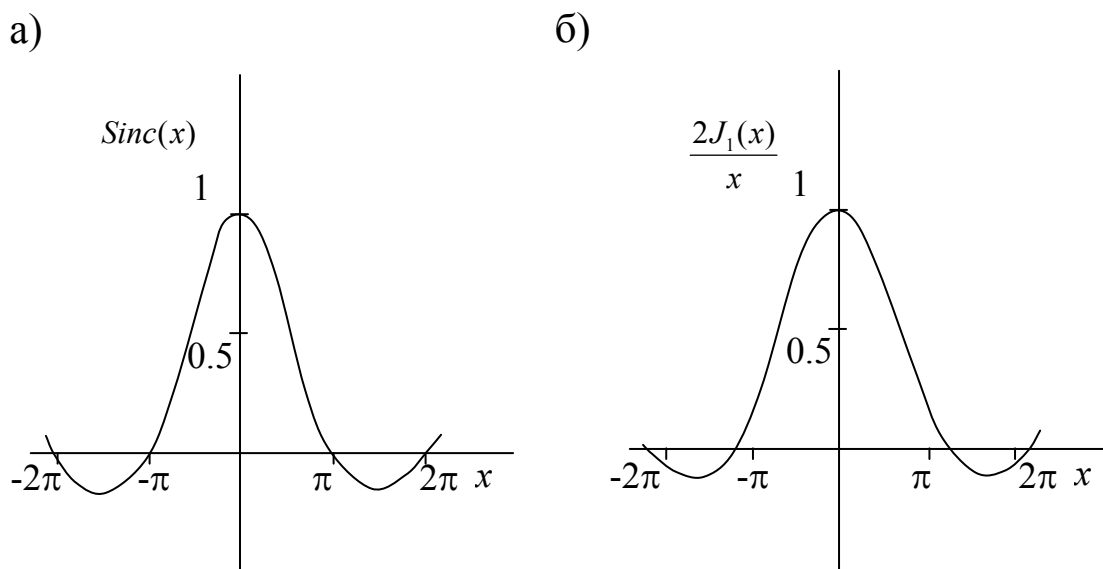


Рис. 11. Графики функций  $Sinc(x)=\sin(x)/x$  (а) и  $2J_1(x)/x$ , определяющих спектры пространственных частот волнового поля соответственно за квадратной и круглой диафрагмами.

В дальнейшем нам потребуется одно из свойств преобразований Фурье, утверждаемое теоремой подобия [17]: если  $G(f_x, f_y) = F\{g(x, y)\}$ , то

$$F\{g(ax, by)\} = \frac{1}{|ab|} G\left(\frac{f_x}{a}, \frac{f_y}{b}\right). \quad (4.33)$$

В справедливости этого утверждения легко убедиться, заменив  $x$  на  $ax$ ,  $y$  на  $by$  в формулах (4.12), (4.13). Для функций, обладающих круговой симметрией, по теореме подобия

$$F\{circ(\rho/a)\} = a^2 \frac{J_1(2\pi af)}{af} = \frac{a}{f} J_1(2\pi af). \quad (4.34)$$

## 5. Пространственное разрешение оптических приборов. Преодоление дифракционного предела разрешения

### 5.1. Теория Аббе и критерий Рэлея

Пространственное разрешение характеризует способность прибора различать близко расположенные предметы или элементы периодической структуры. Компоненты разрешения вдоль каждой из трех координатных осей в общем случае могут быть различными. Под пространственным разрешением оптических приборов обычно подразумевают его горизонтальную (параллельную плоскости образца или перпендикулярную оси прибора) составляющую. Говоря о пространственном разрешении БСОМ, мы будем иметь в виду горизонтальную проекцию наименьшего расстояния между двумя различимыми на изображении элементами структуры образца. Что касается разрешения БСОМ по вертикали, заметим лишь, что оно во многом определяется уровнем шумов, в частности, механических вибраций, дестабилизирующих положение зонда при записи каждого элемента раstra; обычно оно составляет  $\sim 0,1$  нм.

Вопрос о предельном разрешении оптических приборов при когерентном освещении был рассмотрен Аббе. При освещении параллельным пучком нанесенной на прозрачную основу дифракционной решетки за ней возникает система лучей, рассеянных под разными углами. Дифрагированные лучи, собранные линзой, усиливают друг друга, если разность хода между ними составляет целое число длин волн. Этому требованию при нормальном падении света удовлетворяет следующее соотношение между периодом  $d$  решетки, углом дифракции  $\varphi$  и длиной волны  $\lambda$ :  $d \sin \varphi = m\lambda$ , где  $m = 0, 1, 2, \dots$ . При косом освещении, зеркальном относительно дифрагированного луча, (рис. 12) разность хода возрастает вдвое и направления дифракционных максимумов задаются равенствами

$$2d \sin \varphi = m\lambda, \quad (5.1)$$



где  $m$  – целое число.

Для того, чтобы на изображении проявилась периодичность решетки, необходимо попадание в апертуру линзы наряду с прямым ( $m = 0$ ) хотя бы одного дифрагированного пучка ( $m = 1$ ). В пределе при  $\varphi \rightarrow \pi/2$  из (5.1) следует

$$\Delta r_{\min} = d_{\min} = \lambda/2, \quad (5.2)$$

где  $\Delta r_{\min}$  – предельное разрешение, отождествляемое с минимальным различимым на изображении периодом решетки  $d_{\min}$ .

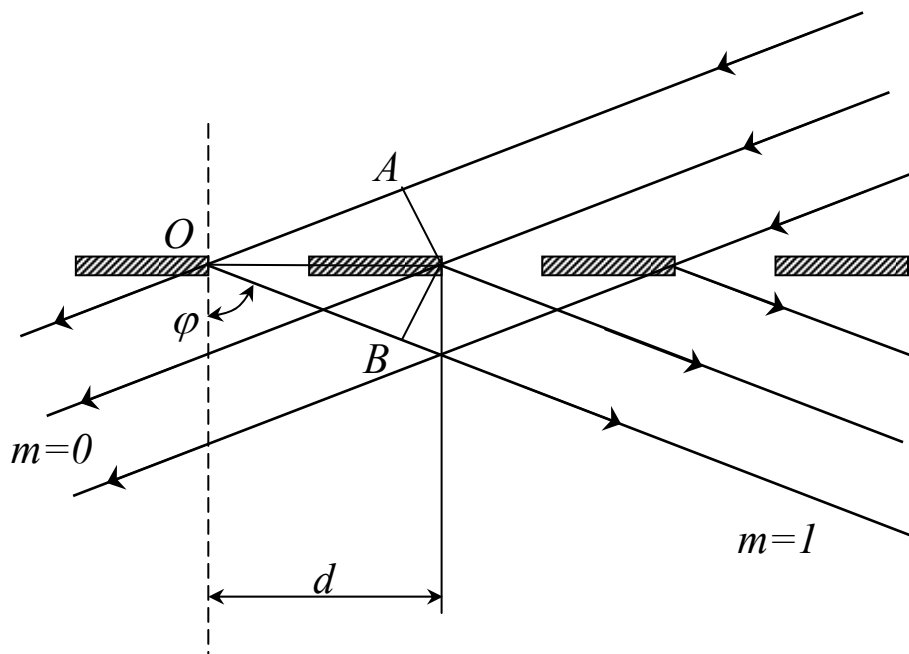


Рис. 12. Рассеяние света решеткой при косом освещении:  $2(OA + OB) = 2d \sin \varphi$  – разность хода между лучами, рассеянными соседними узлами решетки.

Рэлей предложил считать изображения двух светящихся точек раздельными, если максимум дифракционной картины Эйри,

представляющей изображение одной из точек, совпадает с первым минимумом на изображении другой точки. При этом суммарная интенсивность в провале между двумя пиками составляет примерно 80% максимальной (рис. 13), что обеспечивает приемлемый контраст изображения. В безаберрационной линзовой системе при некогерентном освещении распределение интенсивности на изображении точечного источника задается “функцией размытия точки” (point spread function), имеющей следующий вид

$$PSF = \left[ 2 \frac{J_1(\alpha)}{\alpha} \right]^2, \quad (5.3)$$

$$\text{где} \quad \alpha \equiv \frac{2\pi}{\lambda_0} \rho n \sin \theta, \quad (5.4)$$

$J_1(\alpha)$  – функция Бесселя первого порядка,  $\rho$  – расстояние от центра изображения точки, пересчитанное к плоскости образца,  $\lambda_0$  – длина волны света в вакууме,  $n$  – показатель преломления среды, в которую помещен объект,  $\theta$  – апертурный угол, измеряемый между осью линзы и направлением на край диафрагмы. Учитывая (см. формулу (3.59)), что положение первого минимума PSF задается равенством  $\alpha = 3,83$ , из (5.3) находим разрешение по Рэлю

$$\Delta r = 0,61 \frac{\lambda_0}{n \sin \theta} = 0,61 \frac{\lambda}{\sin \theta}. \quad (5.5)$$

При  $\theta \rightarrow \pi/2$   $\Delta r \rightarrow \Delta r_{\min} = 0,61\lambda$ , что лишь слегка расходится с результатом, предсказываемым теорией Аббе. Это расхождение вызвано тем, что в теории Аббе рассматриваются лучи, рассеянные в плоскости падения, а в теории Рэля – распространяющиеся внутри конуса.

К формуле (5.5) можно прийти, используя метод Фурье. Спектральная плотность точечного источника характеризуется функцией  $G(f_x, f_y) = 1$  (или  $G(f) = 1$ ), а значит, в спектре в равной степени представлены все пространственные частоты в диапазоне  $0 \leq f < \infty$ . Зрачок линзы ограничивает пропущенные в плоскость изображения пространственные частоты величиной  $f_{\max} = (\sin \theta)/\lambda$ .

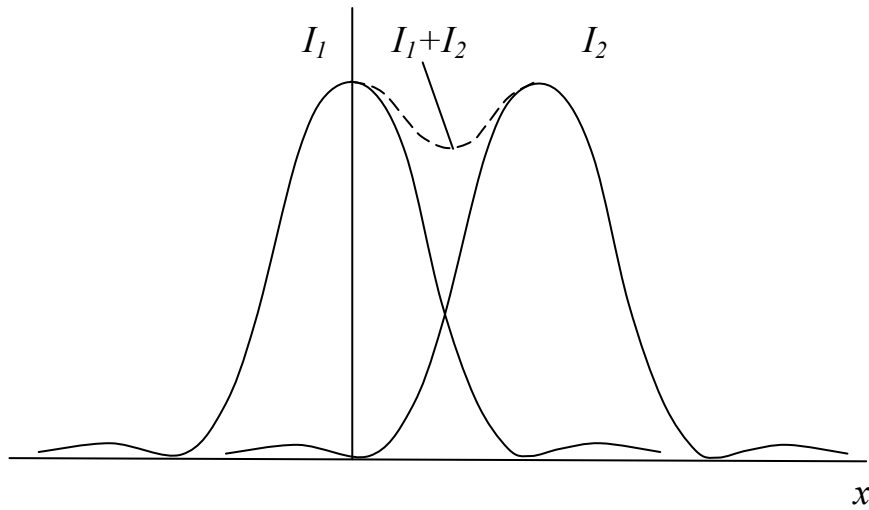


Рис. 13. Иллюстрация критерия Рэлея. Изображения двух точечных источников считаются разрешенными, если максимум одного распределения совпадает с первым минимумом другого.

При отсутствии aberrаций частотный спектр в пространстве за линзой может быть представлен в виде

$$G(f) = \text{circ}(bf), \quad (5.6)$$

где  $b = f_{\max}^{-1} = \lambda/\sin\theta$ . Меняя местами  $\rho$  и  $f$  в левой и правой частях формулы (4.32) и используя теорему подобия (формула (4.34)), приходим к соотношению

$$F^{-1}\{\text{circ}(bf)\} = \frac{1}{b^2} \frac{J_1(2\pi\rho/b)}{\rho/b}. \quad (5.7)$$

Положение первого нуля функции  $J_1(2\pi\rho/b)$  или полуширина центрального максимума определяются равенством  $(2\pi\rho_0/b) = 3,83$ , что соответствует разрешению по Рэлею:  $\Delta r = \rho_0 = 0,61\lambda/\sin\theta$ .

## 5.2. Передаточная функция БСОМ

Размытие точечного источника света в стандартной (линзовой) оптической системе вызвано тем, что зрачок линзы пропускает лишь часть

спектра пространственных частот, причем предельная частота задается числовой апертурой линзы. В БПО ограничение разрешения накладывается конечным размером формируемого диафрагмой светового пятна. Исходя из обратимости световых лучей, изображение точечного источника, сканируемого оптическим зондом, можно рассматривать как эквивалент изображения вершины зонда, сканируемой точечным источником. Поэтому естественно ожидать, что минимальный размер точки, рассматриваемой в апертурном БСОМ, должен быть близок к диаметру диафрагмы.

Более строгий подход к определению пространственного разрешения БСОМ требует анализа спектра пропущенных диафрагмой пространственных частот. Для определенности будем рассматривать БСОМ, работающий в режиме освещения, когда свет приходит на образец по апертурному зонду. Простейшей моделью зонда является диафрагма в непрозрачном экране. Пусть на диафрагму радиусом  $a$  падает плоская волна с единичной амплитудой. Распределение напряженности поля в световом пятне непосредственно за диафрагмой определяется амплитудным коэффициентом пропускания  $t(x,y)$ . Согласно формуле Бете-Боукампа (3.75) это распределение является существенно неравномерным и различным в плоскостях  $xz$  и  $xy$ , соответственно параллельной и перпендикулярной направлению поляризации падающей волны. Зная составляющие  $E_x$ ,  $E_y$  (или  $E_\rho$ ,  $E_\phi$  в цилиндрических координатах) электрического вектора и используя метод Фурье, можно найти частотный спектр поля за диафрагмой. Одна из трудностей прямого решения задачи связана с тем, что  $E_\rho$  неограниченно возрастает у края диафрагмы. Чтобы обойти расходимость решения и избежать громоздких математических выкладок, будем исходить из распределения поля в плоскости на малом, но конечном расстоянии  $z$  от экрана. Анализ показывает, что при росте  $z$  пики  $E_\rho$  у краев диафрагмы сглаживаются, и зависимости квадрата амплитуды электрического вектора от расстояния до оси в двух взаимно перпендикулярных плоскостях приобретают вид, напоминающий графики гауссовой и прямоугольной функций (рис.14).

Несмотря на очевидное различие этих функций, их фурье-образы при надлежащем выборе параметров, характеризующих распределения амплитуд электрического вектора, близки, а значит, должны быть близки и предельные пространственные частоты  $(f_x)_{\max}$ ,  $(f_y)_{\max}$ , пропущенные диафрагмой соответственно вдоль и поперек направления поляризации.

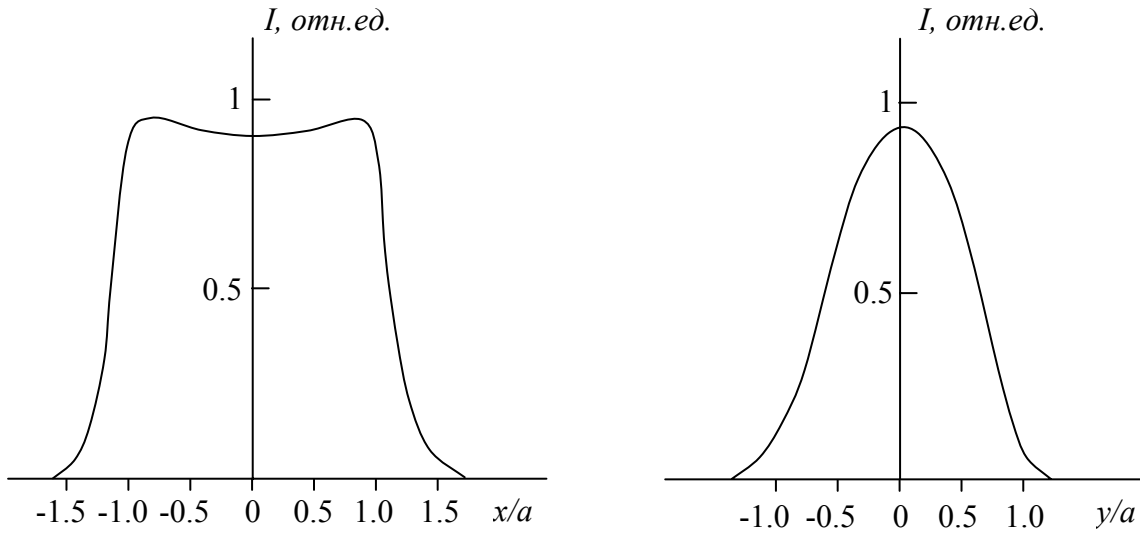


Рис. 14. Зависимость интенсивности электромагнитного поля от расстояния до центра диафрагмы вдоль (а) и поперек (б) направления поляризации падающей волны. Расчет для плоскости, отстоящей на расстоянии  $z=0.2a$  от бесконечно тонкого идеально проводящего экрана. Радиус диафрагмы  $a=0.1\lambda$ .

Чтобы убедиться в этом, сравним спектральные плотности  $G_1(f_x)$  и  $G_2(f_y)$  функций  $g_1(x) = \text{rect}(x/2a)$ ,  $g_2(y) = \exp(-y^2/2\sigma^2)$ :

$$G_1(f_x) = F\{g_1(x)\} = 2a \text{sinc}(2\pi a f_x), \quad (5.8)$$

$$G_2(f_y) = F\{g_2(y)\} = \sqrt{2\pi}\sigma \exp[-2(\pi f_y \sigma)^2]. \quad (5.9)$$

Интенсивности каждой из гармоник на частотах  $f_x, f_y$  в волновом поле за диафрагмой, характеризующиеся квадратами амплитуд  $G_1, G_2$  при  $\sigma = 0.8a$  представлены кривыми 1, 2 на рис. 15.

Определим границу полосы пропускания пространственных частот условием  $[G_i(f)/G_i(0)]^2 = s$ , где  $i=1,2$ ,  $f=f_x, f_y$  и  $s \ll 1$ . Из графиков видно, что при любом выборе допустимого коэффициента ослабления сигнала  $s$  в разумном интервале  $0.01 \leq s \leq 0.1$  граничные частоты  $(f_x)_{\max}, (f_y)_{\max}$  одинаковы с погрешностью порядка 10%. При  $s = 0.01$   $f_{\max} \leq 1/2a$ .

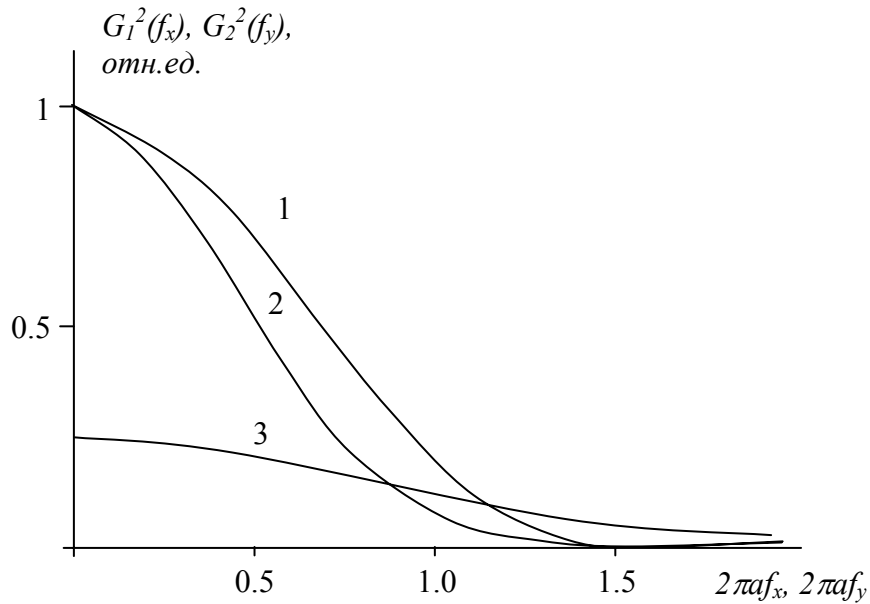


Рис. 15. Зависимость относительной интенсивности сигнала вдоль (кривая 1) и поперек (кривые 2, 3) направления поляризации от пространственной частоты. Спектральные плотности  $G_1^2(f_x)$ ,  $G_2^2(f_y)$  определяются формулами (5.8), (5.9) при  $\delta=0.8a$  (2) и  $\delta=0.4a$  (3).

К аналогичному результату можно прийти, считая распределение поля за диафрагмой осесимметричным и равномерным. Пусть  $t(x,y) = \text{circ}(\rho/a)$ . Тогда спектральная плотность поля за диафрагмой

$$G(f) = F\{\text{circ}(\rho/a)\} = a \frac{J_1(2\pi a f)}{f}. \quad (5.10)$$

Формула (5.10) эквивалентна распределению амплитуд в картине Эйри с той лишь разницей, что координаты и частоты поменялись местами. Определяя по аналогии с критерием Рэля предельную пространственную частоту условием  $J_1(2\pi a f_{\max}) = 0$ , находим

$$f_{\max} = 0,61/a ,$$

что после замены  $d_{\min} = 1/f_{\max}$  дает пространственное разрешение

$$\Delta r = d_{\min} \approx 2a.$$

Мы получили соотношение, имеющее простой физический смысл: разрешение оптического прибора, в котором построение изображения осуществляется путем сканирования образца диафрагмой с коэффициентом пропускания  $t(x,y) = \text{circ}(\rho/a)$ , ограничено диаметром диафрагмы  $2a$ . Две светящиеся точки, например, две молекулы люминофора, находящиеся одна от другой на расстоянии, меньшем размера создаваемого диафрагмой светового пятна, на изображении сливаются в одну размытую точку.

Степень размытия точечного источника характеризует реакцию системы на входной сигнал в виде  $\delta$ -функции. Эту характеристику называют функцией импульсного отклика  $h(x,y)$ , а ее фурье-образ – передаточной функцией

$$H(f_x, f_y) = F\{h(x,y)\} . \quad (5.11)$$

Как показано выше, разумную аппроксимацию формы светового пятна за диафрагмой дает комбинация гауссовой и прямоугольной функций. При этом частотный спектр волнового поля близок к спектру, создаваемому источником с равномерным распределением интенсивности и изображаемому функцией  $G(f) = F\{\text{circ}(\rho/a)\}$ . По мере удаления от диафрагмы световое пятно расширяется, что сопровождается последовательным затуханием высокочастотных компонентов волнового поля.

Если образец поместить на расстоянии  $z$  от диафрагмы, амплитуды распространяющихся волн останутся неизменными, а амплитуды затухающих волн уменьшатся как  $\exp(-kz)$ , причем коэффициент затухания  $\kappa$  является функцией частоты  $f$  и волнового числа  $k$ :  $\kappa = \sqrt{(2\pi f)^2 - k^2}$ . Следовательно, передаточную функцию БСОМ можно представить в виде

$$H(f;z) = H(f;0) \quad \text{при } 2\pi f \leq k , \quad (5.12)$$

$$H(f;z) = H(f;0)\exp(-\kappa z) \quad \text{при } 2\pi f > k , \quad (5.13)$$

где  $H(f;0) = a[J_1(2\pi af)]/f$ ; второй аргумент функции  $H(f;z)$  указывает расстояние от диафрагмы.

Зависимости  $H(f; z)$  от  $af$  при  $\lambda = 600$  нм,  $a = 50$  нм и различных  $z$  приведены на рис. 16. Определяя граничную пространственную частоту условием  $[H(f_{\max}; z)/H(0; z)]^2 = s$  и принимая  $s = 0,01$ , приходим к простой зависимости  $f_{\max}$  от  $a$  и  $z$ , выполняющейся с точностью  $\sim 10\%$  при  $0 \leq z \leq a$

$$f_{\max} \approx 1/(2a + z). \quad (5.14)$$

Соответственно, пространственное разрешение определяется приближенной формулой

$$\Delta r \approx (1/f_{\max}) \approx 2a + z. \quad (5.15)$$

Сравнивая спектры пространственных частот, пропущенных диафрагмой в двух взаимно перпендикулярных плоскостях, мы заметили, что функции  $G_1(f_x)$ ,  $G_2(f_y)$ , определяемые формулами (5.8), (5.9), мало отличаются по величине одна от другой при  $\sigma = 0,8 a$ . Такой выбор параметра  $\sigma$  является в значительной степени произвольным. В непосредственной близости от диафрагмы в бесконечно тонком идеально проводящем экране полуширина светового пятна в плоскости  $yz$ , перпендикулярной направлению поляризации, близка к радиусу  $a$  диафрагмы [18], что соответствует  $\sigma \approx 0,4 a$ .

Уменьшение  $\sigma$  сопровождается снижением как максимальной интенсивности сигнала на малых пространственных частотах (при  $f_y \rightarrow 0$ ), так и скорости спада  $G_2(f_y)$  с частотой (кривая 3, рис. 15). Хотя при  $s = 0,1$  величины  $(f_y)_{\max}$ , определяемые точками пересечения прямой  $G_2^2(f_y)/G_2^2(0) = 0,1$  с кривыми 2,3, практически совпадают, выбор меньших значений допустимого коэффициента ослабления сигнала  $s$  приводит к расхождению предельных частот, причем двукратное уменьшение  $\sigma$  сопровождается пропорциональным увеличением  $(f_y)_{\max}$ . При малых  $s$  принципиально возможно увеличение и предельной частоты  $(f_x)_{\max}$  в плоскости поляризации за счет участия побочных максимумов функции  $\text{sinc}(2\pi af_x)$ . Поскольку выбор  $s$  в конечном итоге определяется способностью приемной схемы различать слабый сигнал на фоне шумов, предельная пространственная частота должна рассчитываться, исходя из индивидуальных особенностей конкретного прибора.



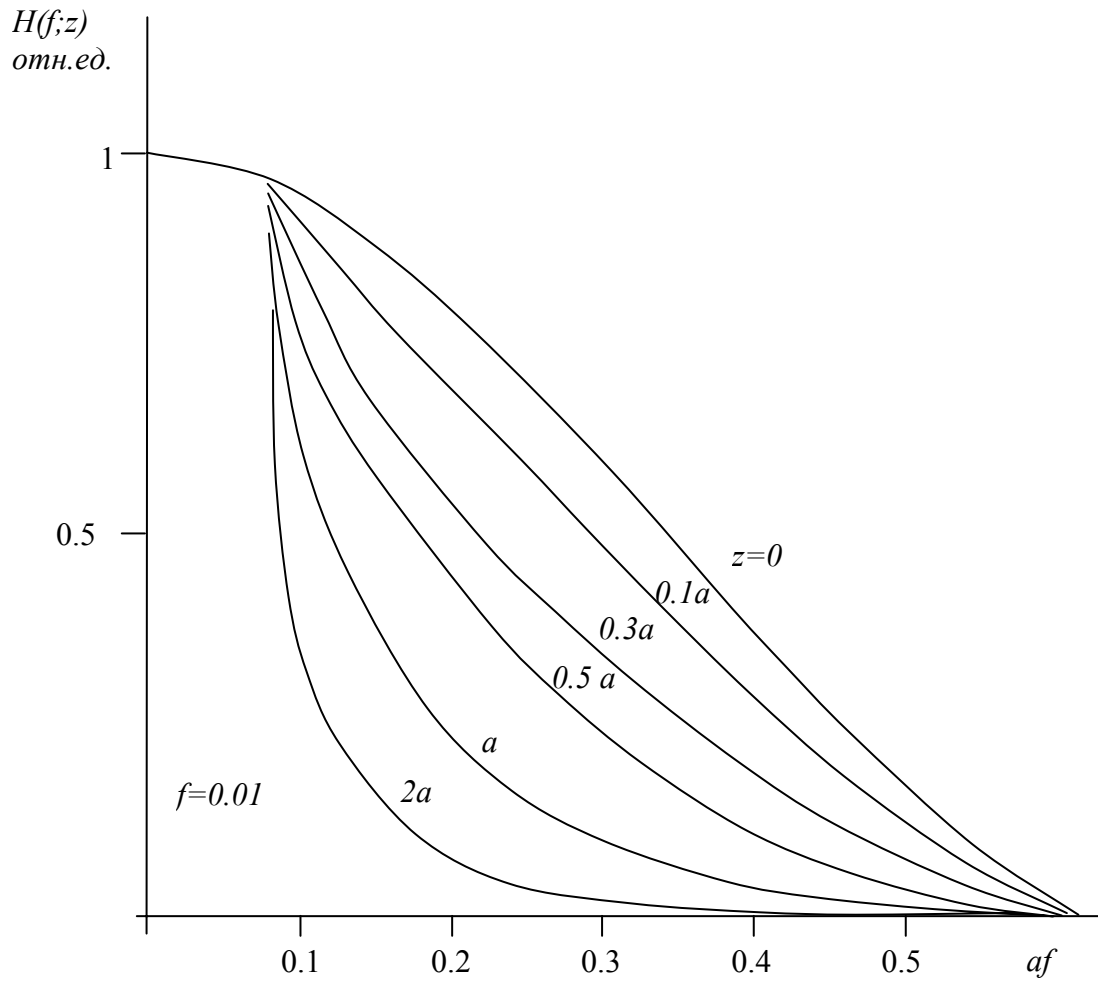


Рис. 16. Изменение передаточной функции апертурного БСОМ с расстоянием  $z$  от диафрагмы  $\lambda=600$  нм,  $a=50$  нм.

С учетом указанных обстоятельств формулу (5.15) можно представить в более общем виде

$$\Delta r \approx \zeta a / [\lg(1/s)] + z, \quad (5.16)$$

где  $\zeta \approx 3$  при  $0.01 \leq s \leq 0.1$ . Последняя формула указывает на возможность реализации разрешения, несколько меньшего диаметра диафрагмы.

### 5.3. Предел разрешения и принцип неопределенности

Критерий Рэлея часто используют в качестве иллюстрации принципа неопределенности Гейзенберга. Определим видимый размер  $\Delta x$  точечного источника света полушириной главного максимума PSF. Учитывая связь волнового вектора  $\mathbf{k}$  и импульса  $\mathbf{p}$  фотона:  $\mathbf{p} = \hbar \mathbf{k}$ , а также возможность отклонения рассеянного на угол  $\theta$  фотона в сторону положительного и отрицательного направлений оси  $x$ , находим

$$|\Delta p_x| = 2p \sin \theta = \frac{4\pi\hbar}{\lambda} \sin \theta. \quad (5.17)$$

Подстановка в это неравенство  $\lambda/\sin\theta$  из формулы (5.5) при  $|\Delta x| = \Delta r$  дает  $|\Delta x| \cdot |\Delta p_x| \approx 1,2h$  в согласии с соотношением неопределенности

$$|\Delta x| |\Delta p_x| \geq h, \quad (5.18)$$

где  $h = 2\pi\hbar$  – одна из форм представления постоянной Планка.

Погрешность  $\Delta p_x$  определения составляющей импульса зависит от апертурного угла  $\theta$ . При  $\theta \rightarrow \pi/2$   $|\Delta p_x| \rightarrow 2p = 2h/\lambda$  и, следовательно, минимальная неопределенность координаты объекта или предельное разрешение оптического прибора заданы величиной  $\Delta r_{\min} \approx h/|\Delta p_x| = \lambda/2$ .

Возможность получить с помощью БСОМ отдельные изображения точек, отстоящих одна от другой на расстояние  $\Delta r \ll \lambda$ , казалось бы, противоречит одному из основных физических принципов.

Однако следует иметь в виду, что принцип неопределенности, ограничивая одну из составляющих импульса, оставляет открытым вопрос о значениях двух других составляющих. Условие  $p_x \leq p$  или  $k_x \leq k$ , вытекающее из представления о конусе световых лучей с конечной и действительной апертурой  $\theta$ , применимо к распространяющимся волнам с действительными компонентами волнового вектора  $k_x, k_y, k_z$ . В общем случае составляющая  $k_z$  волнового вектора вдоль оси прибора может принимать не только действительные, но и мнимые значения. Тогда, вводя  $\kappa = ik_z$  и совмещая плоскость падения луча с плоскостью  $xz$ , приходим к неравенству

$$|k_x| = \sqrt{k^2 - k_z^2} = \sqrt{k^2 + \kappa^2} > k. \quad (5.19)$$

При  $k \rightarrow \infty$  область допустимых значений  $k_x$  неограниченно возрастает, а значит,  $|\Delta x|$  может быть сколь угодно малым.

Хотя волны с  $k_x > k$  не могут быть непосредственно зарегистрированы удаленным детектором, они обладают вполне реальным физическим содержанием и при определенных условиях могут быть преобразованы в излучение. Известно [16], что значительную часть спектра рассеяния света малым объектом радиусом  $a \ll \lambda$  составляют затухающие волны. По теореме обратимости, тот же объект, помещенный в затухающее поле, должен преобразовать часть электромагнитной энергии в излучение. В оптике, как и в радиофизике, система пропускает неискаженный сигнал, если ее полоса пропускания шире, чем частотный спектр сигнала. Из формулы (5.14) следует, что ширина полосы пропускания спектра пространственных частот малой диафрагмой при  $z \rightarrow 0$  обратно пропорциональна ее радиусу. Это соотношение характерно и для других малых объектов, например, рассеивающих свет сферических частиц. Поэтому получить неискаженное изображение такой частицы можно лишь с помощью зонда еще меньшего размера. Проблема реализации предельного разрешения упирается, таким образом, в техническую возможность изготовления миниатюрных зондов и приема слабых сигналов. Критическим параметром становится не длина волны  $\lambda$ , а радиус  $a$  диафрагмы или вершины зонда.

При распространении волны вдоль направления, составляющего угол  $(\pi/2) - \theta$  с осью  $x$ , гребни волны вдоль оси  $x$  повторяются и интервалом  $\lambda_{xp} = \lambda/\sin\theta = h/p\sin\theta$  (см. рис. 4). Зависимость разрешения от угла  $\theta$  можно рассматривать как следствие изменения  $\lambda_{xp}$  с  $\theta$ . Для затухающих волн наименьшее расстояние между гребнями волн вдоль оси  $x$  определяется равенством  $\lambda_{xe} = 1/f_{\max} \approx 2a$  и соотношение неопределенности  $|\Delta x| |\Delta p_x| \geq h$  при  $|\Delta p_x| = 2p$  принимает вид:  $|\Delta x| |\Delta p_x| = 2h |\Delta x| / \lambda_{xe} \geq h$ , т.е.  $|\Delta x| \geq a$ . Таким образом, возможность разрешения элементов структуры, разделенных расстоянием  $a \ll \lambda$  не противоречит принципу неопределенности [19].

#### 5.4. Пространственное разрешение при туннелировании

Оптический сигнал, регистрируемый при сканировании образца зондом БСОМ, несет информацию о распределении интенсивности электромагнитного поля вблизи поверхности. Вследствие ряда причин, например, из-за неизбежного искажения структуры поля образца зондом, изображение не является точной копией исследуемой поверхности. Для корректного определения пространственного разрешения прибора

необходимо сравнение полученного изображения с “идеальным”, которое можно было бы получить при отсутствии искажений. Однако точно предсказать вид идеального изображения, как правило, невозможно. Строгий расчет или моделирование электромагнитного поля в зазоре образец-острие является весьма сложной задачей, решение которой требует индивидуального подхода к каждой конкретной системе.

На практике при оценке достижимого разрешения желательно использовать упрощенную модель, учитывающую основные особенности образования изображения и в то же время допускающую аналитическое решение задачи. Такая модель для СТМ была предложена в середине 80-х годов и в дальнейшем получила широкое распространение при интерпретации электронных изображений проводящих поверхностей [20]. В основе модели лежит представление туннельного тока  $J$  в виде функции матричных элементов  $M_{\mu\nu}$  переходов электронов между разрешенными состояниями образца и острия, характеризуемыми соответственно волновыми функциями  $\psi_\nu$  и  $\psi_\mu$ . Величина матричного элемента определяется характером перекрытия волновых функций  $\psi_\nu$  и  $\psi_\mu$  в туннельном промежутке. В общем случае расчет  $M_{\mu\nu}$  сложен, но при некоторых разумных приближениях можно исключить  $\psi_\mu$ , из выражения для туннельного тока  $J$  и представить ток как функцию только  $\psi_\nu$ , тем самым, связывая его с распределением заряда вблизи образца при отсутствии зонда. Так, для сферического острия

$$J = A_e |\psi_\nu(\mathbf{r}_0; W_F)|^2 = A_e \rho(\mathbf{r}_0; W_F), \quad (5.20)$$

где  $A_e = \text{const}$  при заданном напряжении,  $\rho(r_0; W_F)$  – плотность поверхностных состояний электронов с энергией Ферми  $W_F$ ,  $\mathbf{r}_0 = (x_0, y_0, z_0)$  – координаты центра сферы. Формула (5.20) означает, что при записи изображения в режиме постоянного тока острие перемещается вдоль контура, повторяющего распределение поверхностного заряда вблизи образца на расстоянии от него, определяемом положением точки  $\mathbf{r}_0$ .

Принцип построения изображения иллюстрирует рис. 17. Распределение плотности электронов вблизи поверхности металла повторяет атомную структуру поверхности. Из-за малой глубины  $\delta$  проникновения волновой функции электронов в область потенциального барьера осцилляции плотности заметно сглаживаются уже на расстоянии  $z \leq 1$  нм от поверхности, а плотность заряда при удалении от поверхности на 1 нм уменьшается на несколько порядков. Центр острия с радиусом кривизны  $a$  не может

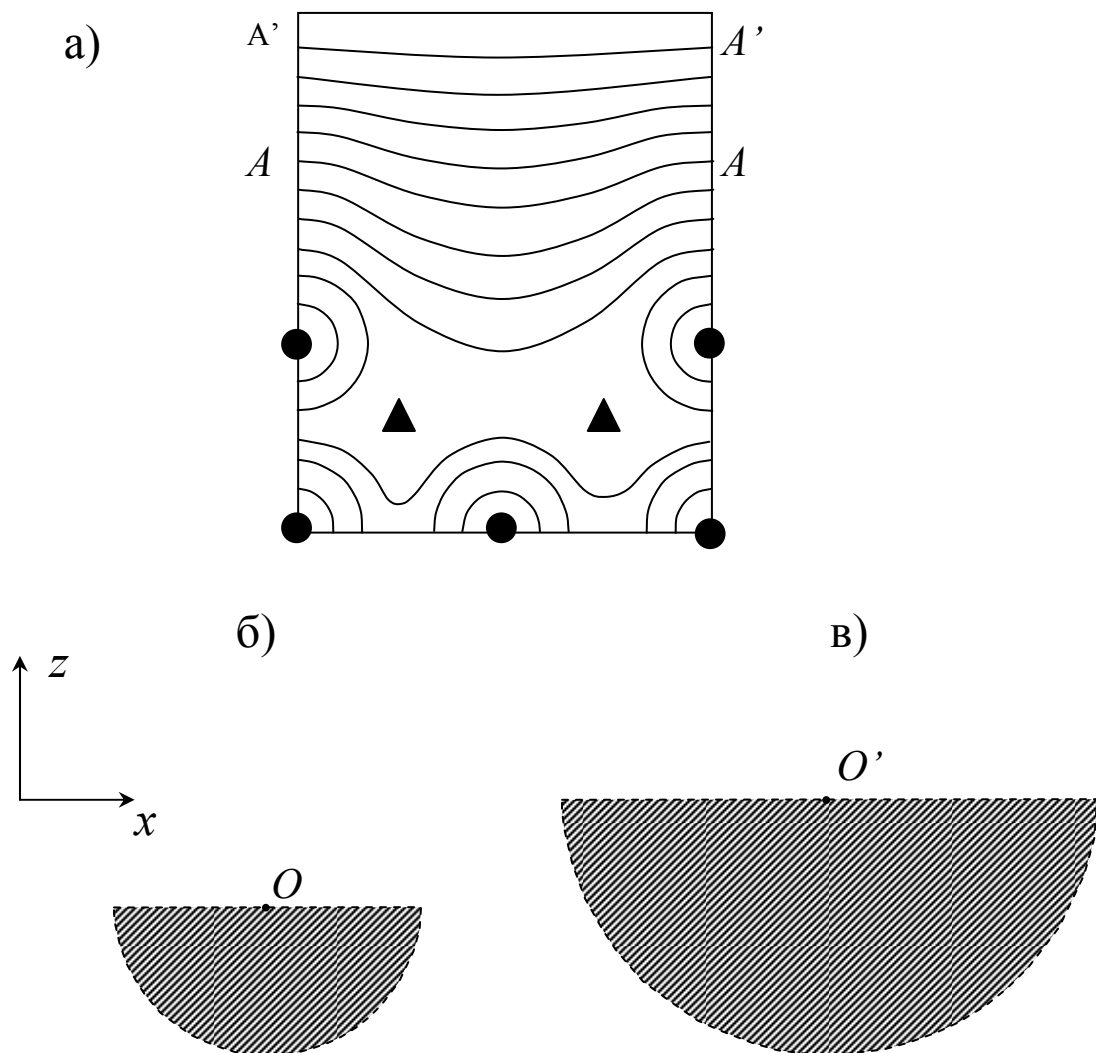


Рис. 17. Иллюстрация принципа построения изображения поверхности металла при регистрации туннельного тока  $J$  в режиме  $J=const$ .

а) контуры постоянной электронной плотности на разных расстояниях от поверхности металла;  $\bullet$ ,  $\blacktriangle$  – положение ядер атомов соответственно в плоскости и вне плоскости рисунка.

б), в) – схематическое изображение максимально приближенных к поверхности сферических острий с разными радиусами кривизны вершины. В случае б) в формировании туннельного тока участвует в основном один атом острия, выступающий над другими атомами вершины; в случае в) – несколько атомов. Эффективные центры  $O$ ,  $O'$  острий при движении вдоль оси  $x$  перемещаются по контурам  $AA$  и  $A'A'$ . По мере удаления от поверхности контуры все в меньшей степени воспроизводят структуру металла.

приблизиться к поверхности на расстояние  $z < a$ . Перемещаясь вдоль контура  $AA$ , точка, совмещенная с центром острия, воспроизводит распределение зарядов, но с существенно меньшей, чем непосредственно у поверхности (при  $z \rightarrow 0$ ) плотностью. Высокое пространственное разрешение СТМ в значительной степени вызвано тем, что в формировании нити туннельного тока может активно участвовать лишь один атом, слегка выступающий над соседними атомами вершины острия. При этом минимальное расстояние от регистрируемого контура  $AA$  до поверхности фактически эквивалентно постоянной решетки.

Механизмы образования изображения в СТМ и БСОМ имеют много общего, что оправдывает единый подход к интерпретации изображения. В рамках такого подхода запись оптического сигнала детектором БСОМ можно рассматривать как регистрацию распределения интенсивности электромагнитного поля вдоль траектории движения виртуального точечного острия [21]. В случае, когда распределение напряженности электромагнитного поля у вершины зонда характеризуется сферической симметрией, положение виртуального острия совпадает с центром кривизны вершины. При нарушении сферической симметрии виртуальное острие смещается в направлении к поверхности или от нее в зависимости от того, вытянуты ли контуры постоянной интенсивности поля вдоль оси или в перпендикулярном направлении. Минимальное расстояние между виртуальным острием (или “эффективным центром острия”) и образцом является основным параметром, определяющим предельное разрешение прибора.

Отметим два принципиальных отличия процессов туннелирования электронов в СТМ и фотонов в БСОМ.

1. В отличие от электронов в металле, положение верхних энергетических уровней которых задано энергией Ферми, смещение уровней фотонов относительно вершины барьера меняется в широких пределах в зависимости от угла  $\theta$ , определяющего направление распространения светового луча в образце. При  $\theta < \theta_c = \arcsin(1/n)$  (см. разд. 3.3) электромагнитная энергия переносится распространяющимися волнами, которые не несут информации о пространственном распределении субволновых элементов структуры образца.

Эти волны можно отсеять, например, ограничивая апертуру детектора углами  $\theta > \theta_c$ . Мы будем предполагать, что только затухающие фотоны ответственны за ток детектора, который в этом случае можно отождествлять с туннельным током.

2. Основные потери мощности на пути от источника излучения к детектору вызваны поглощением или рассеянием фотонов в оптическом зонде. Эти потери многократно увеличивают ослабление сигнала, вызванное затуханием электромагнитных волн в зазоре образец-острие. Отрезок зонда вблизи вершины, следовательно, можно рассматривать как эквивалент дополнительного туннельного промежутка.

По аналогии с формулой (5.20) представим туннельную составляющую тока, регистрируемого детектором при сканировании поверхности зондом со сферически симметричным распределением электромагнитного поля вблизи вершины в виде

$$J = A_{ph} I(\mathbf{r}_0; \theta). \quad (5.21)$$

где  $A_{ph} = \text{const}$  для данного острия при фиксированной мощности падающего излучения,  $I(\mathbf{r}_0; \theta)$  – интенсивность затухающего электромагнитного поля,  $\mathbf{r}_0 = (x_0, y_0, z_0)$  – координаты эффективного центра вершины зонда,  $\theta$  – угол между осью прибора и лучом, распространяющимся в образце в сторону детектора. Предполагается, что детектор является точечным или кольцевым. В противном случае  $\theta$  следует рассматривать как угол рассеяния, усредненный в пределах апертуры детектора.

Определим разрешение туннельного микроскопа (электронного или оптического) наименьшим периодом  $d_{\min}$  структуры, различимым на изображении образца. Условием наблюдения периодической структуры является присутствие в спектре рассеянных волн, помимо постоянной составляющей, хотя бы одной гармоники. При записи изображения структуры с периодом  $d$  вдоль оси  $x$  для выполнения указанного условия туннельный ток  $J(z, x)$  должен включать два компонента

$$J(z, x) = J_0(z, x) + J_1(z, x) = A_0 \exp(-2\kappa z) + A_1 [\exp(-2\kappa_1 z)] \cos^2(\pi x / d), \quad (5.22)$$

где  $A_0, A_1 = \text{const}$ ,  $\kappa$  и  $\kappa_1$  – коэффициенты затухания соответственно постоянной составляющей тока и первой гармоники. Будем предполагать, что запись изображения осуществляется в режиме  $J(z, x) = \text{const}$ . В этом случае осцилляции  $J_1$  при движении острия вдоль оси  $x$  гасятся периодическим смещением острия по вертикали, вызывающим противоположные по знаку изменения  $J_0$ .

Введя амплитуду  $\Delta z$  колебаний высоты острия, при  $\Delta z \ll z$  находим относительную величину осцилляций  $J_0$

$$\delta J_0/J_0 \approx 2\kappa \Delta z .$$

Из равенства  $|\delta J_0| = |\delta J_1|$  следует

$$2\kappa \Delta z = m \exp[-2\kappa(\frac{\kappa_1}{\kappa} - 1)z], \quad (5.23)$$

где  $m = A_1/A_0$  – постоянная, характеризующая глубину модуляции рельефа (так, для дифракционной решетки с треугольным профилем, изображенным на рис. 10,  $m \approx 0,8$ ). При исследовании в БСОМ представляют интерес структуры, период  $d$  которых удовлетворяет неравенству  $d \ll \lambda$ , где  $\lambda$  – длина волны источника света. Поскольку коэффициент затухания  $\kappa_1$  первой гармоники связан с периодом  $\lambda_x$  переменной составляющей сигнала:  $\kappa_1 \approx 2\pi/\lambda_x$ , а коэффициент затухания  $\kappa$  – с длиной волны  $\lambda$ :  $\kappa = (2\pi/\lambda)\sqrt{n^2 \sin^2 \theta - 1}$  (см. разд. 3.3), при  $d \ll \lambda$  справедливо соотношение  $\kappa_1 \gg \kappa$  и формула (5.23) преобразуется к виду

$$\Delta r = \frac{z_t}{2\pi} \ln\left(\frac{m}{2\kappa \Delta z}\right) \equiv B z_t, \quad (5.24)$$

где  $\Delta r \equiv d$  – разрешение,  $z_t$  – расстояние между виртуальным точечным острием и средним уровнем поверхностного рельефа.

Оценим параметры, входящие в формулу (5.24). Для видимого света в широком диапазоне углов рассеяния  $\kappa \sim 10^{-2} \text{ нм}^{-1}$ . Чтобы полезный сигнал выделялся на фоне шумов должно выполняться требование  $\Delta z \geq 0,1 \text{ нм}$  (на практике трудно добиться существенно лучшей стабилизации положения острия по вертикали). Подставляя указанные значения  $\kappa$ ,  $\Delta z$  в формулу (5.24), при типичных  $m = 0,1 \dots 1$  находим  $B \approx 1$ .

Мы пришли к простому определению пространственного разрешения БСОМ по горизонтали: оно равно минимальному расстоянию от эффективного центра острия до образца с поправочным коэффициентом, зависящим от уровня шумов и при типичных условиях близким к единице.

Чтобы количественно охарактеризовать разрешение, нам остается связать расстояние  $z_t$  с геометрическими параметрами зондов разного типа. Для безапертурного зонда со сферической вершиной радиусом  $a$ , край



которой отстоит на расстоянии  $z$  от образца, положение центра зонда определяется очевидным соотношением

$$z_t \approx a + z. \quad (5.25)$$

При сближении зонда с образцом разрешение улучшается, но даже при  $z \rightarrow 0$  не может оказаться существенно меньшим, чем радиус зонда:  $\Delta r_{\min} = Ba \approx a$ .

Оценку величины  $z_t$  при использовании апертурного зонда можно получить, исходя из спектра пространственных частот, пропущенных диафрагмой. Допуская, что предельная пространственная частота задана формулой (5.14), находим минимальный период структуры  $d_{\min} = 2a + z$ , который мы отождествляем с разрешением прибора. Коэффициент затухания  $\kappa_1$  для этой частоты определяется равенствами  $\kappa_1 = \pi / d_{\min} = \pi / (2a + z)$  и, следовательно, с ростом расстояния  $z_t$  от поверхности образца сигнал на частоте  $f_{\max}$  спадает как  $\exp[-2\pi z_t / (2a + z)]$ . Приравнивая эту величину введенному выше коэффициенту ослабления  $s$  сигнала, находим

$$z_t \approx \frac{2a + z}{2\pi} \ln(1/s). \quad (5.26)$$

Отсюда при разумном  $s = 0,01$   $z_t \approx 0,7(2a + z)$ . При  $z \rightarrow 0$  ослабление сигнала диафрагмой эквивалентно действию туннельного промежутка шириной  $1,4a$ . Таким образом, эффективный центр апертурного зонда с идеальной диафрагмой находится от торца зонда на расстоянии, несколько меньшем диаметра диафрагмы. По этой причине изменение расстояния  $z$  диафрагмы от образца мало сказывается на разрешении, так как в большинстве практических случаев действует неравенство  $z \ll 2a$ .

Заметим, что оба рассмотренных выше подхода к оценке разрешения апертурного БСОМ, основанных на использовании методов фурье-оптики и теории туннелирования, приводят к одинаковому выводу: *предельное разрешение близко к диаметру диафрагмы и может слегка отличаться от него в ту или другую сторону в зависимости от уровня шумов и чувствительности приемной схемы.*

## **6. Принципиальные и технические ограничения основных параметров БСОМ**

### **6.1. Пространственное разрешение и контраст. Влияние поляризации света и оптических характеристик системы зонд-образец**

Как показано в разделе 5, вопрос о пространственном разрешении БСОМ может быть сведен к определению размеров светового пятна, формируемого оптическим зондом. Зная распределение электромагнитного поля в зазоре зонд-образец и используя преобразование Фурье, можно найти ширину полосы пропускания пространственных частот. Она, в свою очередь, определяет ожидаемую четкость изображения. При рассмотрении модели апертурного БСОМ с зондом, оболочка которого вблизи вершины представляет идеальный проводник, такой подход приводит к оценке предельного пространственного разрешения, близкой к диаметру  $2a$  диафрагмы.

Аналогичные представления в принципе применимы и к реальным приборам, если известно распределение электромагнитного поля вблизи вершины зонда. Проблема состоит в том, что расчет структуры электромагнитного поля в нанометровом зазоре зонд-образец требует использования трудоемких вычислительных методов, корректность которых зависит от того, насколько удачно выбрана конфигурация зазора. Между тем, получить правильные представления о форме вершины зонда на практике удается, как правило, лишь после детального анализа изображения специально подобранных тест-объектов. Вид острия может меняться в процессе сканирования, например, из-за случайных соударений с образцом или прилипания посторонних частиц. Это означает, что привлечение самых точных методов измерения радиуса кривизны вершины зонда или радиуса диафрагмы и самых совершенных методов расчета полей может оказаться бесполезным при оценке пространственного разрешения в реальной ситуации.

При выборе оптимальных условий наблюдения образцов необходимо знать, как изменение параметров среды и условий освещения образца сказывается на качестве изображения. Чтобы избежать чрезмерных вычислительных трудностей, часто бывает удобнее искать приближенный ответ на эти вопросы, заменяя трехмерную задачу двумерной. Удачным приближением является использование двумерной модели апертурного БСОМ, в которой функция зонда выполняет плавно сужающаяся щель в бесконечно протяженном металлическом экране [22]. Щель заполнена диэлектриком, например, стеклом, представляющим световод для плоской

волны, распространяющейся по нормали к экрану. Под узким концом щели на расстоянии, примерно равным его ширине ( $2a = 30$  нм), помещен плоский образец из того же материала, что и световод. Нижний край экрана слегка скошен относительно плоскости образца, чтобы имитировать затупленную вершину зонда.

Ориентируем ось  $z$  по оси световода, а оси  $x$ ,  $y$  соответственно вдоль и поперек щели. Поскольку электромагнитное поле свободно распространяется в диэлектрике вдоль оси  $x$ , тогда как его проникновение в металлическую оболочку зонда ограничено толщиной скин-слоя, структура поля асимметрична относительно осей  $x$  и  $y$  и в сильной степени чувствительна к ориентации электрического вектора  $\mathbf{E}_{\text{вх}}$  падающей волны.

На рис. 18 изображены построенные путем численного решения уравнений Максвелла контуры постоянной интенсивности электромагнитного поля при двух взаимно перпендикулярных направлениях поляризации.

При  $s$ -поляризации (вектор  $\mathbf{E}_{\text{вх}}$  направлен по оси  $x$  вдоль щели, т.е. перпендикулярно плоскости рисунка) условие совпадения тангенциальных составляющих электрического вектора с двух сторон границы раздела металл-диэлектрик (форм. (3.22)) приводит к плавному уменьшению интенсивности поля при удалении от оси световода. Поле не исчезает на границе металла, как это происходило бы в случае идеального проводника, но проникает на глубину скин-слоя  $\delta_{\text{ск}}$ , зависящую от диэлектрической проницаемости металла. Для алюминия, выбранного в качестве материала экрана в модели [22] и на практике используемого чаще других металлов, характерны значения  $\epsilon = -34,5 + i \cdot 8,5$  и  $\delta_{\text{ск}} = 13$  нм при  $\lambda = 500$  нм.

Структура поля при  $p$ -поляризации (вектор  $\mathbf{E}_{\text{вх}}$  направлен по оси  $y$  поперек щели) характеризуется более равномерным распределением интенсивности внутри световода, а также заметным растеканием силовых линий вдоль торца зонда, приводящим к расширению светового пятна. В плоскости диафрагмы на узком конце щели поле концентрируется вблизи острых краев экрана. Основные различия между условиями распространения света по сужающемуся световоду при двух главных направлениях поляризации проявляются, таким образом, в большей степени сжатия  $s$ -поляризованного луча и в большей интенсивности  $p$ -поляризованного.

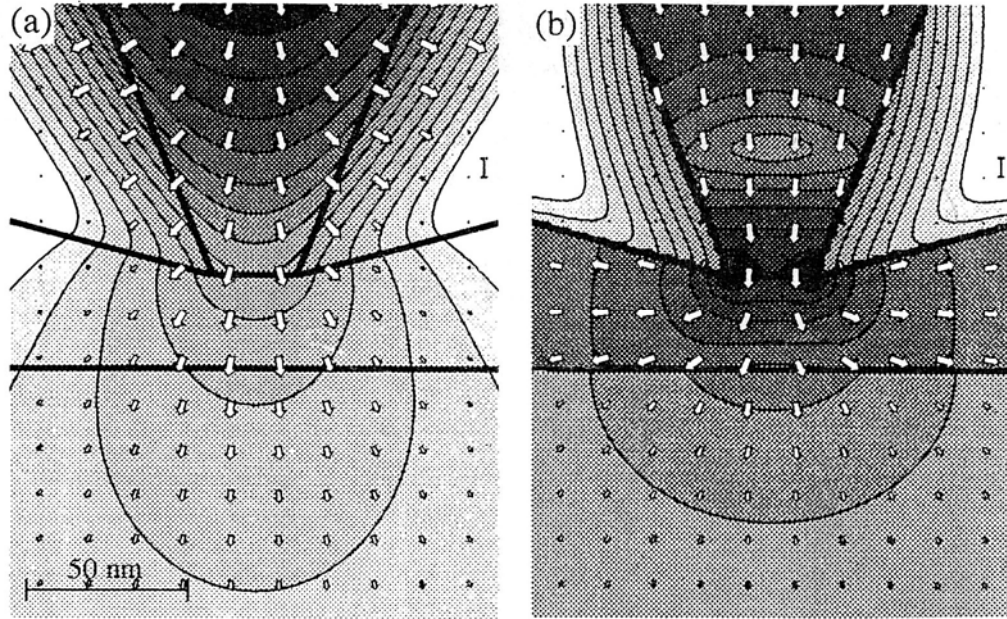


Рис. 18. Структура электромагнитного поля вблизи вершины зонда в двумерной модели апертурного БСОМ [22]. Из-за конечной проводимости экрана поле частично проникает из сердцевины зонда в оболочку.

На каждом последующем (при движении сверху вниз) контуре постоянной интенсивности поля  $\ln |E|^2$  уменьшается вдвое.

а) s – поляризация, электромагнитный вектор падающей волны перпендикулярен плоскости рисунка

б) p – поляризация, электромагнитный вектор падающей волны параллелен плоскости рисунка.

Как видно из графиков, изображенных на рис. 19, распределения интенсивности электромагнитного поля в плоскости диафрагмы в двумерной модели БСОМ при двух направлениях поляризации качественно совпадают с предсказываемыми теорией Бете-Боукампа (форм. (3.75)). Справедливость основных выводов теории, в свою очередь, подтверждается в опытах с реальными зондами. Это позволяет рассматривать картины, приведенные на рис. 18, как качественно правильные представления структуры полей в двух взаимно перпендикулярных сечениях апертурного зонда.

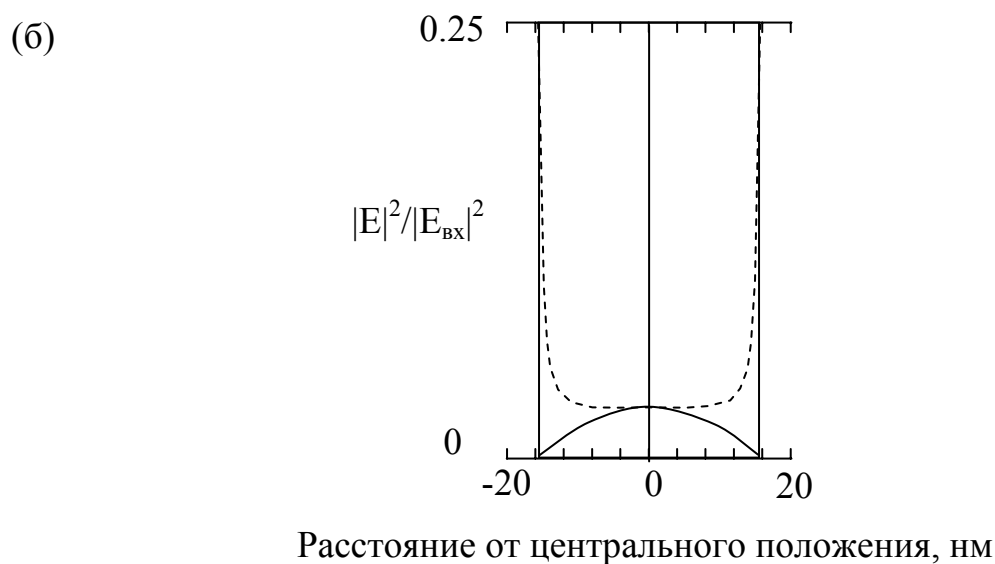
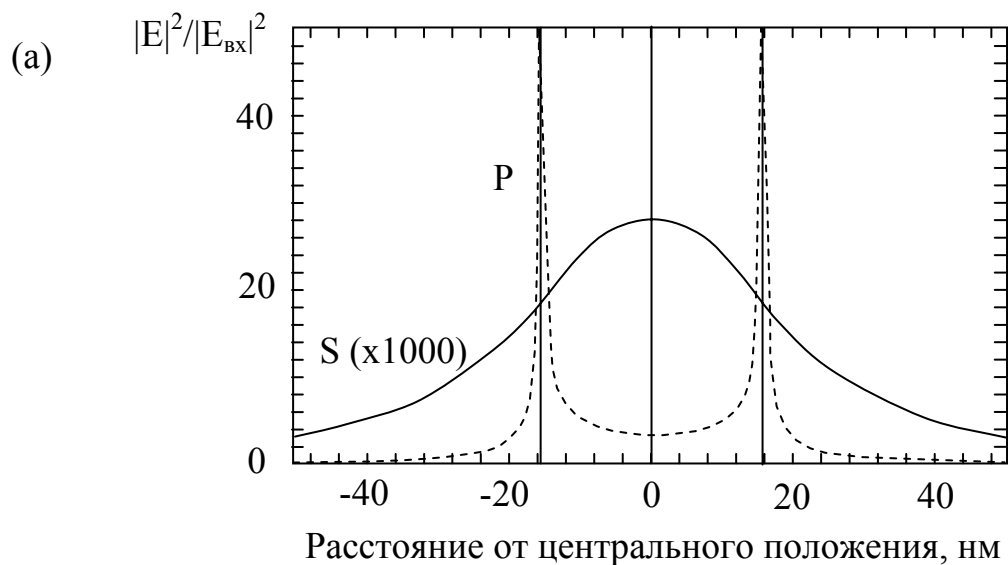


Рис. 19. Зависимость относительной интенсивности в плоскости диафрагмы от расстояния до оси зонда при двух направлениях поляризации: вдоль (пунктир) и поперек (сплошные кривые) плоскости рисунка.

а) двумерная модель БСОМ; излучение проходит через заполненную стеклом плавно сужающуюся щель в алюминиевом экране.

б) излучение проходит через круглую диафрагму в бесконечно тонком идеально проводящем экране.

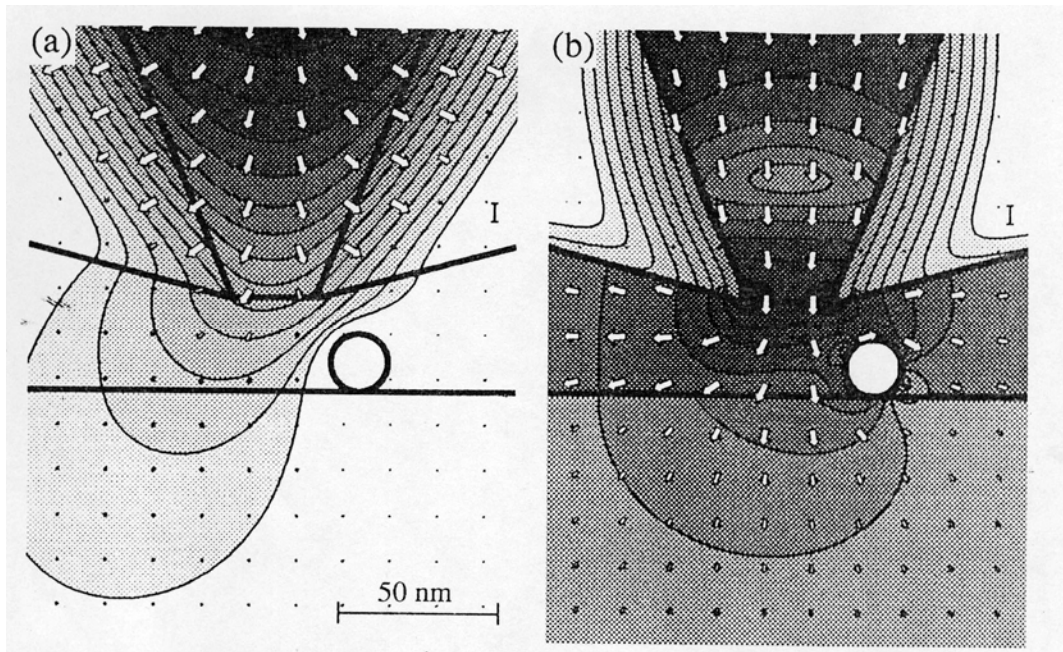


Рис. 20. Искажение аксиальной симметрии электромагнитного поля идеально проводящей частицей диаметром 20 нм, нанесенной на плоскую поверхность образца (в схеме, изображенной на рис. 18) на расстоянии 30 нм от оси зонда [22].  
а)  $s$  – поляризация; б)  $p$  – поляризация

Используя указанную аналогию, можно выделить следующие основные особенности распределения поля в зазоре зонд-образец апертурного БСОН.

1. Протяженность поля в плоскости образца, особенно при  $p$ -поляризации, может значительно превышать рассчитанную, исходя из геометрических размеров диафрагмы.

2. При малой ширине  $z$  зазора ( $z \ll 2a$ ) размер светового пятна на образце слабо зависит от  $z$ . Аналогичный вывод был получен в разделе 5 для идеальных зондов.

3. Изменение формы зазора зонд-образец существенно влияет на распределение поля. В частности, значительные искажения вносит наличие нанометровых металлических частиц, выступающих над плоской поверхностью образца. Рис. 20 иллюстрирует вид контуров равной интенсивности поля вблизи идеально проводящей частицы диаметром 20 нм.

При  $s$ -поляризации падающей волны частица отталкивает световое пятно, при  $p$ -поляризации – слегка притягивает его, одновременно нарушая аксиальную симметрию поля.

4. При настройке излучения в резонанс с частотой плазменных колебаний частицы поле концентрируется вблизи ее поверхности в области, меньшей размера диафрагмы. Этот вывод существенен для практических приложений БПО, связанных с модификацией материалов под действием интенсивного локализованного поля.

5. Важную роль играет соотношение показателей преломления материалов образца и зазора. Так, если на поверхности образца находится частица диэлектрика с более высоким, чем окружающая среда, показателем преломления, она, в противоположность металлической частице, притягивает световое пятно.

Таким образом, эффективный размер и форма светового пятна, а значит, и реализуемое разрешение, помимо диаметра диафрагмы и ширины зазора зависят от многих факторов. К ним относятся частота и поляризация падающего излучения, оптические характеристики материалов образца, зонда и зазора, форма поверхности образца и торца зонда, наличие выступов нанометрового размера, их взаимное расположение и т.д. Влияние указанных факторов можно учесть в общем виде, вводя вместо “геометрического” радиуса  $a$  диафрагмы эффективный или “оптический” радиус  $a_{eff}$ . Зависимость  $a_{eff}$  от  $a$  в первом приближении может быть представлена линейным соотношением

$$a_{eff} = \varphi_{pol}(s, p) \left( a + \frac{\delta_{sk}}{2} \right), \quad (6.1)$$

где  $\varphi_{pol}(s, p)$  – коэффициент, зависящий от направления поляризации, а также от геометрических и оптических характеристик системы. Для практических целей при определении разрешения оправдано использование усредненной по двум направлениям поляризации оценки  $\varphi_{pol}(s, p) \approx 1$ . Так, для распределений, приведенных на рис. 19а, при полуширине диафрагмы  $a = 15$  нм уменьшение интенсивности вдвое в плоскости диафрагмы происходит на расстоянии от оси  $a_{eff} \approx 22$  нм  $\approx a + (1/2)\delta_{sk}$  как для  $s$ -, так и для  $p$ -поляризованного света.

Считая, что реализуемое разрешение определяется формулой (5.15) с заменой  $a$  на  $a_{eff}$ , находим

$$\Delta r = f_{\max}^{-1} \approx 2a_{eff} + z. \quad (6.2)$$

При  $a, z \rightarrow 0$   $\Delta r \rightarrow \Delta r_{min} \approx \delta_{sk} \approx 13$  нм, что совпадает с экспериментальными оценками предельного разрешения апертурных БСОМ [3]. Этот вывод имеет простое физическое объяснение – при сколь угодно малой диафрагме размер светового пятна, формируемого зондом, не может быть меньше удвоенной (в обе стороны от оси) глубины проникновения поля в металлическую оболочку зонда. Связывая размер светового пятна с полушириной распределения интенсивности, приходим к оценке  $\Delta r_{min}$ , примерно совпадающей с толщиной скин-слоя.

Проблема определения пространственного разрешения и его зависимости от поляризации падающего света тесно связана с анализом контраста изображения. Одни и те же детали образца в  $s$ - и  $p$ -поляризованном свете могут выглядеть совершенно по-разному. Это явление, получившее название “поляризационный контраст”, объясняется особенностями согласования тангенциальных составляющих электрического вектора на поверхности частиц. Если электрический вектор направлен по касательной к поверхности металлической частицы, поле затухает в радиальном направлении монотонно по мере удаления частицы от оси прибора. В этих условиях при сканировании металлической частицы апертурным зондом мощность прошедшего света минимальна, когда ось зонда проходит через центр частицы (сплошная кривая на рис. 21а). При том же направлении электрического вектора сильно преломляющая свет диэлектрическая частица создает обратный контраст (пунктир, рис. 21а). Если электрический вектор направлен по нормали к поверхности частицы и вдоль строки развертки, прохождение зонда над частицей независимо от ее материала мало сказывается на интенсивности прошедшего света (рис. 21б), причем по характеру изображения трудно судить о размере и формы частицы.

Критерием пространственного разрешения прибора помимо ширины полосы пропускания может служить степень размытия резкого края протяженной детали образца или нанометровых частиц. Анализ кривых, приведенных на рис. 21а, позволяет сравнить оценки разрешения, полученные с использованием двух различных подходов. Полуширина сечения идеально проводящей частицы диаметром 20 нм при записи строки  $s$ -поляризованным светом (рис. 21а) составляет 120 нм; частицы диэлектрика того же размера – 90 нм. Вычитая диаметр частицы, приходим к оценкам разрешения соответственно 100 и 70 нм. Последняя величина близка к рассчитанной по формуле (6.2) при  $2a_{eff} \approx 50$  нм и  $z = 23$  нм (расстояние от плоскости диафрагмы до центра частицы).



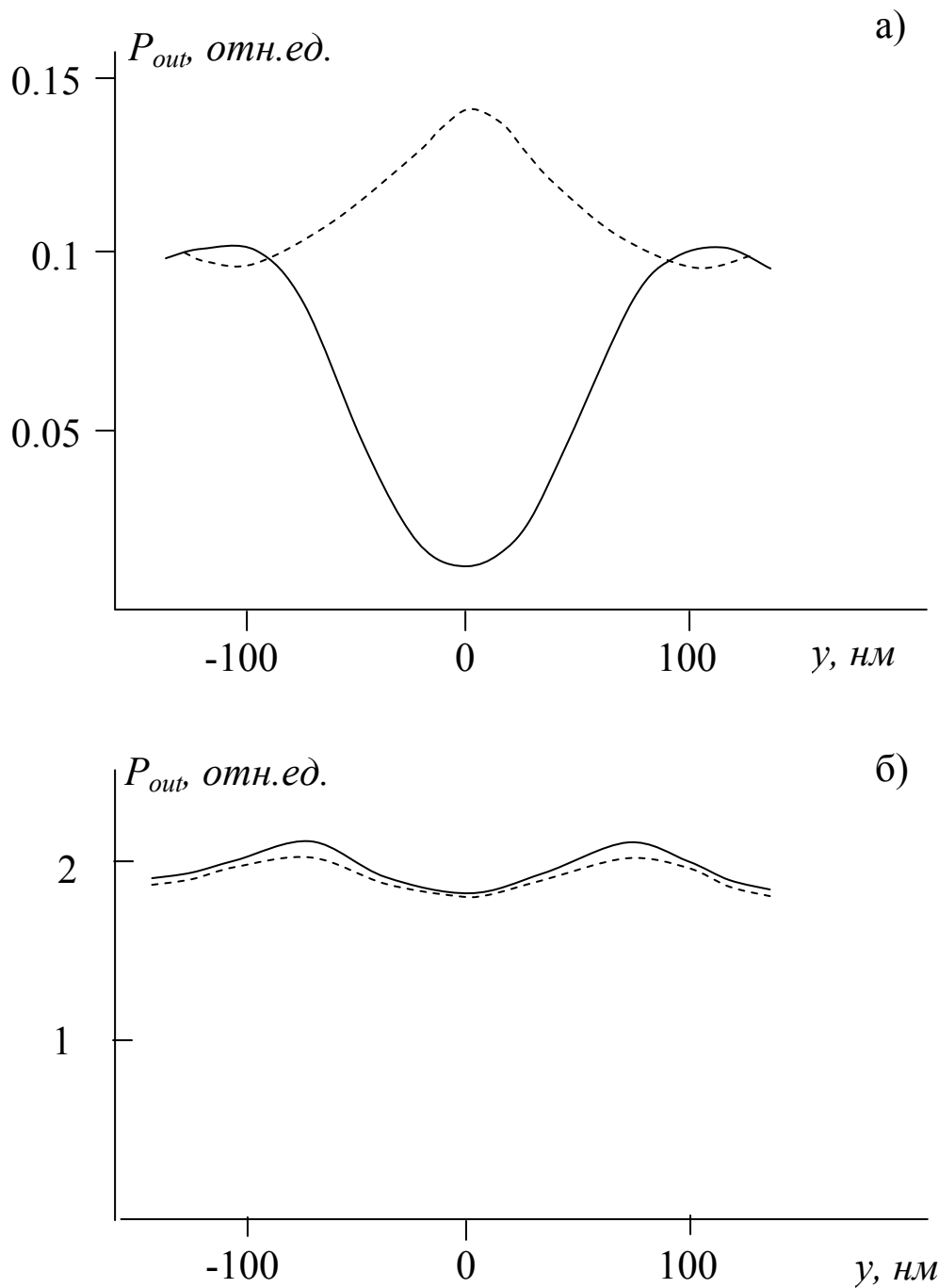


Рис. 21. Зависимость относительной мощности, прошедшей через зонд и частицу диаметром 20 нм, от расстояния между осью зонда и центром частицы в двумерной модели БСОМ. Схема аналогична изображенной на рис. 20, ширина зазора зонд-поверхность образца – 33 нм. Сплошные кривые – частица идеального проводника,  $\epsilon = -\infty$ ; пунктир – частица диэлектрика,  $\epsilon = 11.56$ .  
а) s-поляризация; б) p-поляризация.

Проявлением поляризационного контраста является асимметрия изображения проводящих образцов. Так, круглое металлическое кольцо на изображении превращается в эллиптические дуги (рис. 22). Меняя направление поляризации падающего света, можно повышать контраст тех или иных деталей изображения путем оконтуривания их границ.

## 6.2. Трудность разделения оптической и атомно-силовой составляющих сигнала

Одной из задач, решаемых с помощью БСOM, является выделение участков образца, различающихся по своим оптическим характеристикам (показателю преломления, спектральному составу излучения и т.д.). Такую задачу можно было бы однозначно решить, если бы система, управляющая перемещением острия, реагировала на изменение оптического сигнала. Между тем, в подавляющем большинстве БСOM при позиционировании зонда используются атомно-силовые сенсоры расстояния острие-образец.

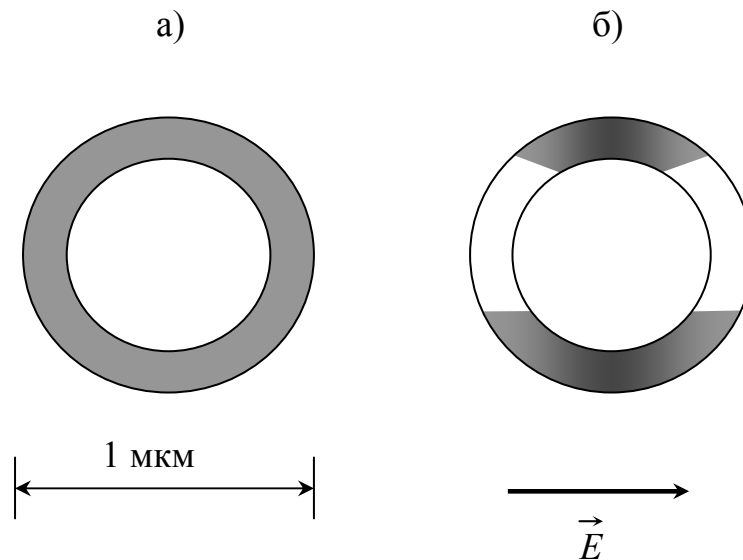


Рис. 22. Демонстрация поляризационного контраста. Алюминиевое кольцо (а) и его изображение в апертурном БСOM (б) [22]. Ориентация темных дуг меняется при изменении направления поляризации.

Это приводит к наложению на оптическую составляющую регистрируемого сигнала дополнительной информации, отражающей

топографический рельеф поверхности. Надежное выделение полезного сигнала возможно далеко не всегда, даже при наличии двух отдельных каналов обработки изображения.

Чтобы пояснить трудности, возникающие при интерпретации изображения, вернемся к простому примеру, рассмотренному в разд. 2, но в ситуации, более близкой к реальной. Пусть требуется найти распределение в образце некоторых оптических меток, например, молекул люминофора. Исследование проводится в апертурном БСОМ в режиме сбора фотонов, когда излучение молекул по волоконному зонду передается к детектору. Торцевой зонд, как правило, не бывает идеально гладким. На нем может находиться микровыступ (рис. 23а), например, выступающее над средним уровнем зерно алюминия, расстояние от которого до поверхности образца определяет величину сигнала в атомно-силовом канале.

Представим ситуацию, когда частицы люминофора сосредоточены на некотором участке (отрезок АС на рис. 23) вдоль строки развертки, и одна из частиц (в точке В) расположена выше других. Рост оптического сигнала вблизи точки А (рис. 23в) свидетельствует об увеличении локальной концентрации молекул люминофора.

Ширина фронта импульса, как следует из оценок разрешения, полученных в разделе 5.2, близка к  $2a + z$ , где  $a$  – радиус диафрагмы и  $z$  – расстояние от торца зонда до образца. Атомно-силовой сигнал начинает изменяться при прохождении выступа зонда над точкой В (рис. 23б), причем из-за малых размеров выступа скорости нарастания и спада импульса значительно выше, чем аналогичные характеристики оптического сигнала в точках А и С. При сканировании образца силы взаимодействия атомов острия и образца возрастают в момент прохождения зонда над точкой В. Из-за реакции системы обратной связи зонд сдвигается по вертикали в направлении от исследуемой поверхности. В результате значительная часть молекул люминофора, находящихся на участке, вырезаемом диафрагмой, отдаляется от нее, что приводит к спаду оптического сигнала.

Наложение на кривую записи оптического сигнала обращенного импульса атомно-силового сигнала является следствием “эффекта отрыва” (lift-off effect). Это название подразумевает образную аналогию резкого подъема острия, сталкивающегося с локальной неоднородностью рельефа, с взлетом летящего аппарата. Если точка В близка к одной из граничных точек области с повышенной концентрацией люминофора, два импульса с разной крутизной могут частично перекрываться, усугубляя трудность расшифровки изображения.

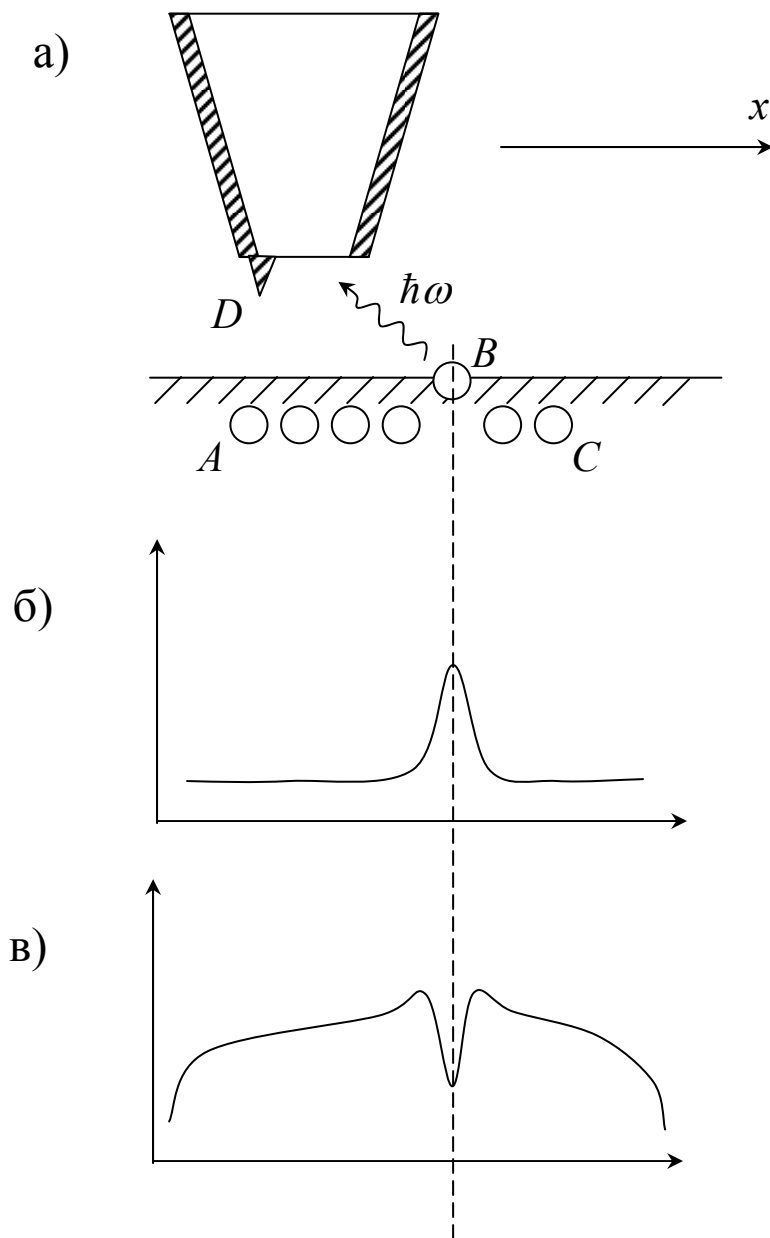


Рис. 23. Наложение топографической информации на оптическое изображение.

а) апертурный зонд над образцом с частицами люминофора, сосредоточенными на отрезке  $AC$ . Одна из частиц в точке  $B$  выступает над средним уровнем поверхности;  $D$  – микровыступ покрытия торца зонда

б) атомно-силовой сигнал;

в) оптический сигнал;

Уменьшение интенсивности при прохождении зонда над частицей  $B$  является следствием «эффекта отрыва».

Спад оптического сигнала в точке В можно ошибочно приписать прохождению острия над областью, отличающейся меньшей оптической активностью. При такой интерпретации воображаемое пространственное разрешение оказывается существенно лучшим, чем теоретический предел, определяемый размером диафрагмы. Известны многочисленные примеры наблюдения дискретных деталей, разделенных расстоянием, намного меньшим периода различимой на изображении решетки [3]. Было бы ошибкой связывать эти детали с истинной структурой образца.

В условиях, когда точная форма вершины зонда не известна, единственным надежным способом исключения “эффекта отрыва” является использование образцов, гладких настолько, чтобы изменение сил атомного взаимодействия острия с образцом при сканировании выбранного участка в режиме постоянной высоты было пренебрежимо мало. Лишь в этом случае любое изменение оптического сигнала указывает на прохождение зонда над оптически неоднородным участком.

### 6.3. Потери мощности в оптическом зонде

Важнейшей характеристикой апертурного зонда является коэффициент пропускания мощности  $\eta$ . Обычно его определяют отношением

$$\eta = P_{in}/P_{out}, \quad (6.3)$$

где  $P_{in}$ ,  $P_{out}$  – мощности излучения соответственно на входе и выходе зонда. Коэффициент пропускания в сильной степени зависит от формы острия, поэтому большое внимание уделяется поиску технологий, обеспечивающих изготовление зондов с оптимальной геометрией.

Чтобы представить, как должен выглядеть такой зонд, будем исходить из требования минимизации потерь на каждом из трех участков световода: цилиндрическом, коническом и непосредственно на вершине острия (рис. 24).

Обозначая коэффициенты пропускания на этих участках соответственно  $\eta_1, \eta_2, \eta_3$ , представим суммарный коэффициент пропускания в виде

$$\eta = \eta_1 \cdot \eta_2 \cdot \eta_3. \quad (6.4)$$

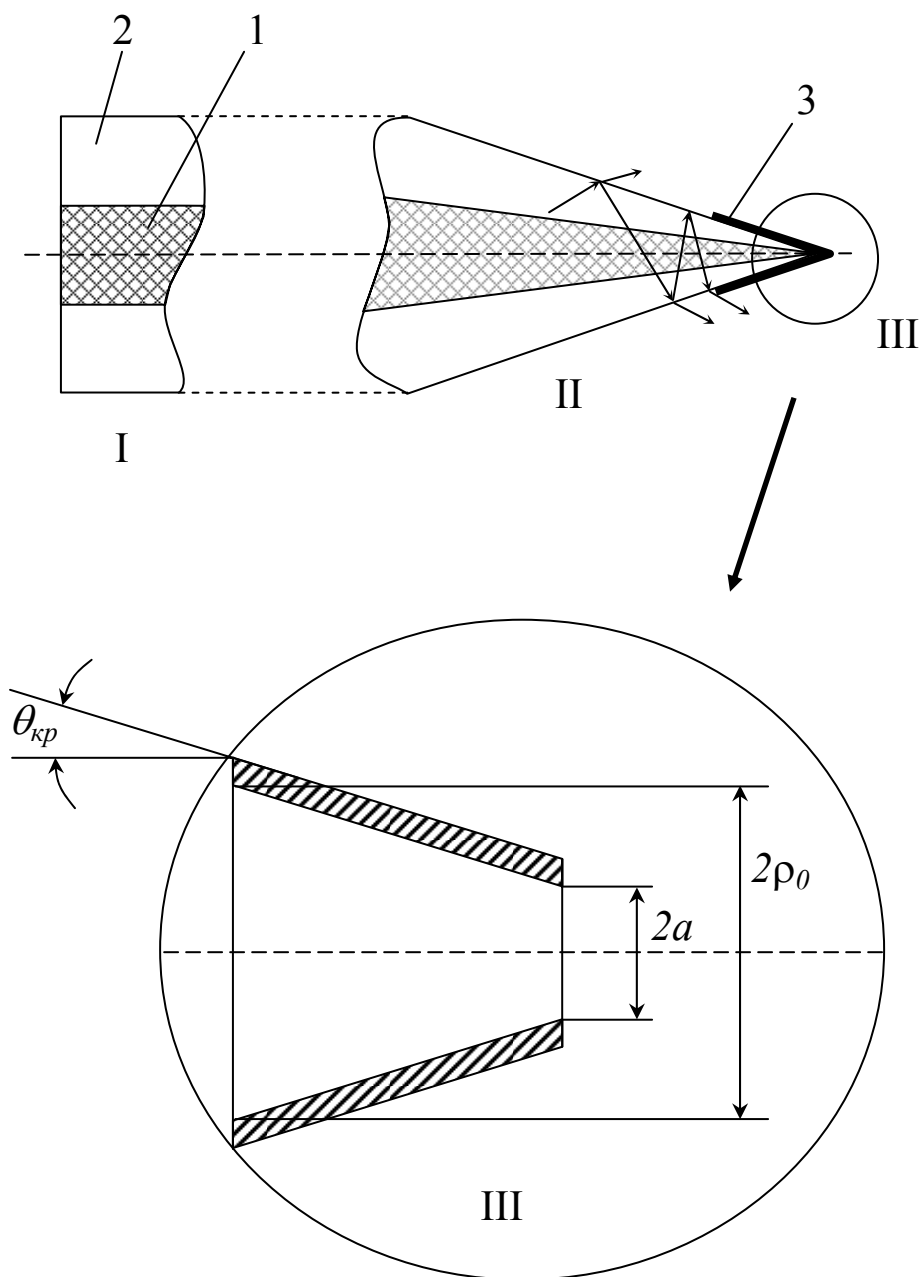


Рис. 24. Три участка волоконного зонда: цилиндрический (I), конический (II) и собственно вершина (III) (внизу с увеличением).

1, 2 – сердцевина и оболочка кварцевого волокна;

3 – металлическое покрытие.

Для наглядности сердцевина волокна увеличена в десятки раз.

$2\rho_0$  – диаметр отсечки основной моды.

После того, как излучение лазера запускается через согласующий элемент в волоконный световод, он распространяется почти без потерь на цилиндрическом отрезке зонда. За ним следует более или менее протяженный конический участок. Последовательное уменьшение угла падения светового луча на этом участке приводит к нарушению условия ПВО. В результате свет начинает проникать в оболочку и после одного или нескольких отражений от наружной поверхности выходит из волокна.

Строгий анализ потерь мощности в конической части зонда требует рассмотрения последовательной отсечки направляемых мод по мере сужения световода. Ограничимся приближенной оценкой  $\eta_2$  для одномодового волокна. Будем считать, что свет, выходящий из цилиндрического световода, попадает в однородную среду и распространяется в ней в пределах телесного угла, определяемого апертурой  $\theta_a$  волокна. Распространение света в постепенно сужающемся волокне возможно до сечения, радиус  $\rho_0$  которого соответствует условию отсечки основной моды в эквивалентном цилиндрическом световоде. Как будет показано ниже, для кварца в воздухе или в металлической оболочке  $\rho_0 \approx 0,2\lambda_0$ , где  $\lambda_0$  – длина волны света в вакууме. Если радиус  $a$  диафрагмы удовлетворяет неравенству  $a > \rho_0$ , конический отрезок зонда заканчивается диафрагмой.

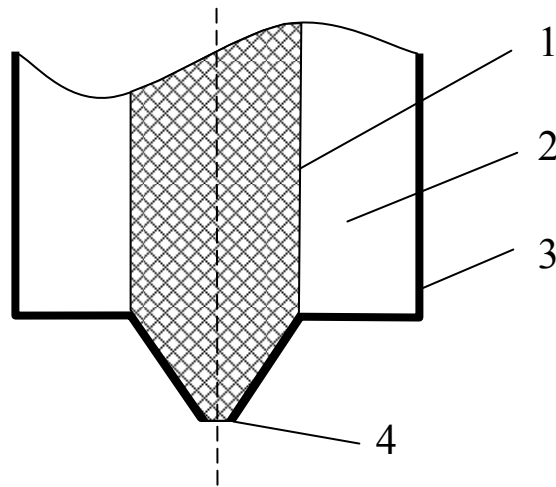


Рис. 25. Возможна форма зонда, изготовленного химическим травлением:

- 1, 2 – сердцевина и оболочка кварцевого волокна;
- 3 – металлическое покрытие;
- 4 – диафрагма.

Для наглядности сердцевина волокна увеличена в десятки раз.

Введя  $\rho_c \equiv \max \{a, \rho_0\}$  и пренебрегая вкладом многократно рассеянных лучей, находим

$$\eta_2 \approx [\rho_c / L \operatorname{tg}(\theta_a)]^2, \quad (6.5)$$

где  $L$  –длина конуса. Величина  $L$  может меняться от сотен микрометров в зондах, полученных при растяжении нагретого кварца, когда формирование конуса происходит по всему сечению волокна, до нескольких микрометров в зондах, изготовленных методами химического травления. Подбирая различные добавки к кварцу, можно обеспечить условия, при которых травление оболочки волокна происходит быстрее, чем сердцевины. В этом случае зонд по форме напоминает карандаш с выступающим наружу заточенным грифелем (рис. 25). Считается, что зонды типа “карандаш” с вершиной, изображенной на рис. 25, или с более плавным переходом от сердцевины к оболочке обеспечивают прохождение излучения с минимальными потерями мощности.

Оценка по формуле (6.5) и измерения показывают, что значения  $\eta$  в зондах с предельно короткими ( $L = 3 \dots 5 \text{ мкм}$ ) и удлинёнными ( $L \approx 100 \text{ мкм}$ ) конусами могут отличаться на 2...3 порядка. По этой причине все большее распространение получают технологии, использующие холодное травление кварца и позволяющие свести длину конической части зонда к минимальной величине, определяемой диаметром сердцевины и углом раствора конуса.

Последняя часть апертурного зонда снаружи покрыта металлом, поэтому при расчете потерь на этом отрезке необходимо учитывать модовый состав излучения в металлическом волноводе. Как было отмечено выше (раздел 3.9), при сужении волновода происходит последовательная отсечка мод высших порядков. Дальше других в идеальном металлическом волноводе распространяются  $TE_{11}$  и  $TM_{01}$  моды. В реальном волноводе, стенки которого имеют конечную проводимость, чисто поперечные моды преобразуются в смешанные. Роль основной моды вместо  $TE_{11}$  выполняет похожая на нее по структуре поля  $HE_{11}$  мода. Меняются и условия отсечки, так как эффективное сечение сердцевины растет из-за проникновения поля в металл. Дополнительные сложности вносит возможность существования поверхностных мод, распространяющихся вдоль границ раздела металл-кварц и металл-воздух.

Указанные обстоятельства затрудняют строгую постановку задачи расчета ослабления мощности на последнем отрезке конического зонда и не позволяют решить ее аналитически. Попробуем упростить задачу, аппроксимируя вершину зонда отрезком эквивалентного цилиндрического



волновода [23]. Начнем с определения вершины. Рассматривая кварцевый волновод с идеально проводящими стенками, будем отождествлять с вершиной отрезок между концом зонда и сечением отсечки  $TE_{11}$  моды. На этом отрезке поле существует в форме затухающих волн. Их амплитуда в цилиндрическом волноводе радиусом  $\rho$  меняется с расстоянием  $z$  по закону

$$E(z) = E(0) \exp(-\kappa z) = E(0) \exp \left[ -\frac{2\pi z}{\lambda_{11}} \left( 1 - \frac{\lambda_{11}^2}{\lambda^2} \right)^{1/2} \right], \quad (6.6)$$

где  $\kappa$  – коэффициент затухания,  $\lambda_{11} = 2\pi\rho/u_{11}$ ,  $u_{11} = 1,841$  – корень бесселевой функции (см. разд. 3.7). Свет с длиной волны  $\lambda = \lambda_0/n$  в среде с показателем преломления  $n$  может распространяться до сечения радиусом  $\rho_0 = 1,841\lambda_0/2\pi n$ , которое мы приняли за основание вершины. Из этого соотношения при  $\lambda_0 = 500$  нм,  $n = 1,46$  находим  $\rho_0 = 100$  нм. В большинстве случаев радиус  $a$  диафрагмы меньше, чем  $\rho_0$ , а значит, электромагнитное поле в непосредственной близости от диафрагмы содержит только затухающие компоненты.

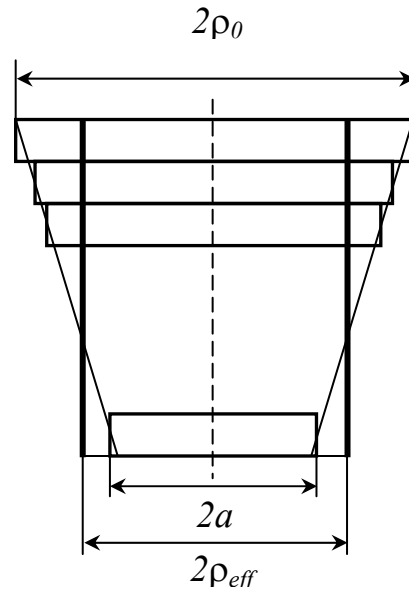


Рис. 26. Представление конической вершины зонда набором элементарных цилиндров.  $2\rho_{eff}$  – диаметр цилиндрического волновода, характеризуемого тем же коэффициентом пропускания, что и равная ему по длине коническая вершина зонда.

Заменяя конус набором  $s$  элементарных цилиндров (рис. 26), будем характеризовать коэффициент затухания поля на последнем участке зонда величиной

$$\kappa = \frac{1}{s} \sum_{q=1}^s \kappa_q, \quad (6.7)$$

где  $\kappa_q$  – коэффициент затухания на цилиндрическом отрезке с порядковым номером  $q$ . Суммарному коэффициенту затухания  $\kappa$  можно сопоставить радиус  $\rho_{eff}$  цилиндра, ослабляющего поле точно так же, как равная ему по высоте коническая вершина зонда. Как следует из геометрических соображений, величина  $\rho_{eff}$  должна принимать промежуточные значения между  $a$  и  $\rho_0$ , т.е.  $\rho_{eff} = \xi \rho_0$  при  $(a/\rho_0) < \xi < 1$ .

Принимая во внимание, что взаимодействие электромагнитного поля с веществом характеризуется квадратом амплитуды электрического вектора  $\mathbf{E}$ , определим прозрачность  $\eta_t$  вершины зонда отношением

$$\eta_t = \frac{|E(0)|^2}{|E(z)|^2}, \quad (6.8)$$

где  $E(z)$  – амплитуда электрического вектора на расстоянии  $z$  от конца зонда. В связи с тем, что электрический вектор в затухающем поле содержит как продольную, так и поперечную составляющие, отношение интенсивностей поля, определяемое формулой (6.8), строго говоря, не является тождественным отношению мощностей излучения в тех же сечениях зонда. В дальнейшем мы будем пренебрегать этим различием, допуская  $\eta_t \approx \eta_z$ .

Объединяя формулы (6.6) – (6.8), находим

$$\eta_t = \exp(-2\kappa z) = \exp\left[-\frac{4\pi z}{\lambda_0} n\varphi(\xi)\right], \quad (6.9)$$

$$\text{где} \quad \varphi(\xi) \equiv \frac{1}{\xi} \sqrt{1 - \xi^2} = \sqrt{\left(\frac{\rho_0}{\rho_{eff}}\right)^2 - 1}. \quad (6.10)$$

Заметим, что величины  $\rho_0$  и  $\rho_{eff}$  имеют один и тот же порядок, а значит,  $\varphi(\xi)$  не может существенно отличаться от единицы. Так, при  $\lambda_0 = 500$  нм, 20

нм  $\leq a \leq 80$  нм имеем  $\varphi(\xi) \approx 1 \pm 0,5$ . Если аппроксимировать  $\varphi(\xi)$  постоянной величиной, показатель степени в формуле (6.9) можно считать пропорциональным длине вершины  $z$ , которая связана с радиусом  $a$  диафрагмы и апертурой конуса при вершине  $\theta$  простым соотношением

$$z = (\rho_0 - a)/\operatorname{tg}\theta . \quad (6.11)$$

При выборе геометрических параметров зонда в первую очередь обращают внимание на радиус  $a$  диафрагмы, определяющий пространственное разрешение прибора. Если величина  $a$  фиксирована, единственным регулируемым параметром, существенно влияющим на пропускание света вершиной, является апертурный угол  $\theta$ . Моделирование электромагнитного поля вблизи вершины зонда показывает целесообразность увеличения  $\theta$  до  $40 \dots 50^\circ$ . При еще больших углах раствора конуса начинает заметно сказываться растекание поля за края диафрагмы, приводящее к увеличению размеров светового пятна.

В реальной ситуации из-за конечной проводимости металла поле проникает в оболочку на глубину скин-слоя и основной модой становится смешанная мода  $HE_{11}$ . Использование численных методов и моделирования позволяет представить картину распространения этой моды в коническом волноводе. Не вдаваясь в особенности решения задачи, можно предсказать качественно правильный результат, исходя из следующих соображений. Проникновение поля в глубь металла эквивалентно увеличению эффективного радиуса диафрагмы и уменьшению радиуса отсечки, если определять этот радиус границей металлического волновода. Поскольку оба параметра согласно формулам (6.9) – (6.11) входят в показатель экспоненты с разными знаками, их одновременное изменение должно существенно сказываться на величине коэффициента  $\eta_t$ . Сравнение результатов расчета  $\eta_t$  для  $TE_{11}$  и  $HE_{11}$  мод (рис. 27) показывает, что формулы (6.9) – (6.11) можно приближенно использовать и при расчете ослабления света в реальном волноводе, заменяя радиус  $a$  диафрагмы величиной  $a_0 \approx a + 30$  нм и оставляя  $\rho_0$  неизменным.

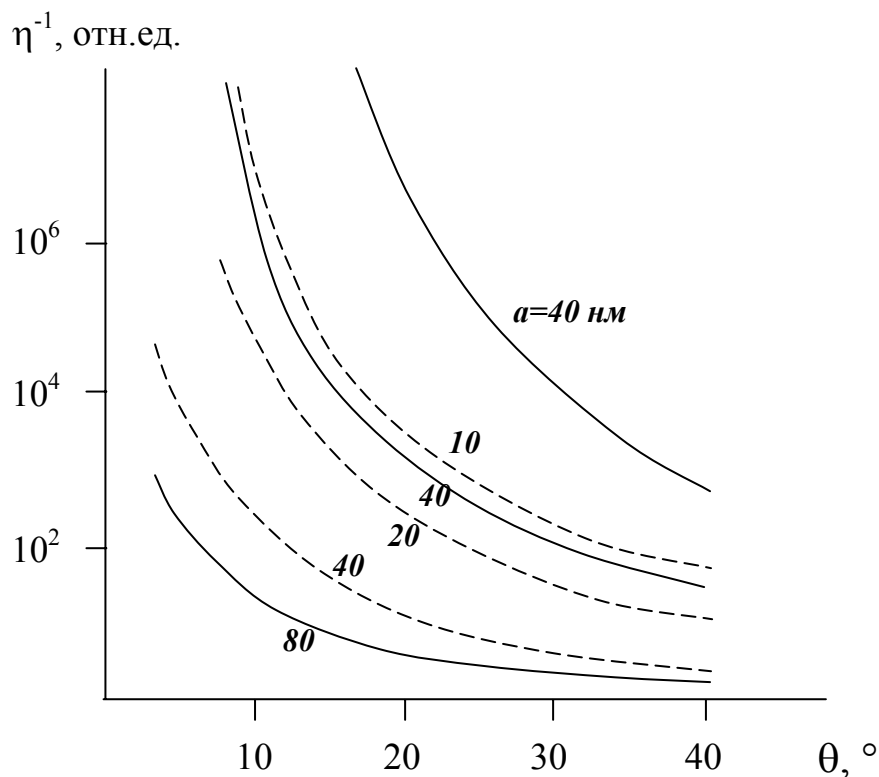


Рис. 27. Зависимость коэффициента ослабления мощности вершиной апертурного зонда от половинного угла раствора конуса  $\theta$  при разных значениях радиуса  $a$  диафрагмы; сплошные кривые –  $TE_{11}$  мода, зонд с идеально проводящим покрытием; пунктир –  $HE_{11}$  мода, зонд с алюминиевой оболочкой.

Оценим суммарные потери мощности в зонде при типичных значениях радиуса  $a$  диафрагмы в диапазоне  $30 \text{ нм} \leq a \leq 100 \text{ нм}$ . При  $a = 30 \text{ нм}$  распространение основной моды прекращается на конечном расстоянии  $z$  от торца зонда. Расчет по формулам (6.9) – (6.11) при  $\lambda_0 = 500 \text{ нм}$ ,  $n = 1,46$ ,  $\varphi(\xi) = 1,2$ ,  $\theta = 30^\circ$ ,  $z = (\rho_0 - a_0)/\text{tg}\theta = 70 \text{ нм}$  дает  $\eta_t = 5 \cdot 10^{-2}$ . Пропускание нескольких процентов мощности излучения характерно и для конической части зонда с формой, изображенной на рис. 25. В частности, при  $L = 5 \text{ мкм}$ ,  $\text{tg}\theta_a = 0,1$ ,  $\rho_c = \rho_0 = 100 \text{ нм}$  из формулы (6.5) находим  $\eta_2 = 5 \cdot 10^{-2}$ . Потери при вводе излучения в волокно обычно малы и часто их вообще исключают из

рассмотрения. Пусть, например,  $\eta_1 = 0,5$ . Тогда при  $\eta_3 \approx \eta_t$  коэффициент пропускания мощности зондом  $\eta = \eta_1 \cdot \eta_2 \cdot \eta_3 \approx 1 \cdot 10^{-3}$ , что соответствует лучшим экспериментальным результатам, полученным с зондами такого же радиуса. При  $a = 100$  нм  $HE_{11}$  мода распространяется до конца зонда и, следовательно,  $\eta_3 \approx 1$ ,  $\eta \approx 2 \cdot 10^{-2}$  при тех же значениях остальных параметров.

Увеличение длины конуса до  $L = 100$  нм в соответствии с формулой (6.5) должно привести к росту потерь на 2...3 порядка. Измеренные значения коэффициента пропускания мощности  $\eta$  зондами с удлинённым конусом, образующимся при нагреве и растяжении оптического волокна, обычно составляют  $10^{-5} \dots 10^{-7}$  при  $a = 30 \dots 50$  нм.

Таким образом, использование простых формул (6.5), (6.9) – (6.11) позволяет правильно оценить порядок величины  $\eta$  и обойти трудности, возникающие при численном расчете и моделировании распределения поля в апертурном зонде.

#### **6.4. Предельная чувствительность при различении малых частиц и спектральном анализе**

Одним из основных достоинств БСОМ, выделяющим его среди других зондовых микроскопов, является возможность проведения спектрального анализа образца. Оптические спектры содержат разнообразную информацию, позволяющую, в частности, судить о химическом составе поверхности. Точность анализа зависит от чувствительности прибора, определяемой его способностью обнаружить присутствие в образце малого числа молекул примеси. Чувствительность является функцией ряда параметров, включая мощность излучения на выходе зонда и скорость развертки.

Поясним вид этой функциональной зависимости на конкретном примере, представляющем интерес для различных областей применения БСОМ. Пусть образец содержит частицы флуорофора, например, красителя, избирательно реагирующего с функционально различными молекулярными группами поверхности. Свечение флуорофора возбуждается излучением лазера, сжимаемым апертурным зондом до размера выходной диафрагмы. Интенсивность свечения определяется мощностью излучения вблизи образца, которую можно представить в виде

$$P_t = \alpha \eta P, \quad (6.12)$$

где  $\alpha$  – коэффициент усиления мощности вблизи заостренной вершины зонда,  $\eta$  – коэффициент пропускания или энергетическая эффективность зонда,  $P$  – мощность лазера. Скорость эмиссии фотонов частицей, содержащей  $N$  оптически активных молекул, характеризуется соотношением

$$j = N\Phi_f \frac{\sigma}{\pi a^2} \frac{P_t}{\hbar\omega} , \quad (6.13)$$

где  $\Phi_f$  – квантовый выход флуоресценции,  $\sigma$  – сечение захвата молекулой фотона с энергией  $\hbar\omega$ ,  $a$  – радиус диафрагмы (для краткости мы не учитываем различия между радиусами диафрагмы и светового пятна).

Если детектором является счетчик, создающий 1 импульс на  $n$  фотонов, скорость счета составит

$$J_1 = j \beta / n , \quad (6.14)$$

где  $\beta$  – доля эмитированных фотонов, проходящих через входную диафрагму детектора. При записи изображения зонд последовательно перемещается от одной точки образца к другой, одновременно меняя свое положение по высоте. Время  $\tau_p$  его задержки в каждой точке цифрового раstra равно отношению времени записи кадра (обычно  $\sim 1$  мин.) к числу элементов изображения ( $\sim 10^4$ ). Если размер частицы меньше размера элемента изображения (пиксела), число  $J$  регистрируемых фотонов при прохождении зонда над частицей равно

$$J = J_1 \tau_p . \quad (6.15)$$

На полезный сигнал накладываются шумы, вызванные дробовым эффектом и пропорциональные  $J^{1/2}$ , а также шумы, связанные с флуктуациями темнового тока. Последними, как правило, можно пренебречь. Считая условием надежного обнаружения частицы не менее чем трехкратное превышение уровня сигнала над амплитудой шумов, приходим к требованию

$$J \geq 10 . \quad (6.16)$$

При спектральном анализе ширина полосы пропускания детектора  $\Delta\Omega$  всегда существенно превосходит определяемую разрешением спектрометра ширину линии  $\Delta\omega$ . Объединяя формулы (6.12) – (6.15) и вводя в них

спектральный коэффициент ослабления сигнала  $\gamma_s = \Delta\omega/\Delta\Omega$ , представим окончательный результат в виде

$$J = \gamma_s N \Phi_f \alpha \eta \frac{\sigma}{\pi a^2} \frac{P}{\hbar \omega} \frac{\beta}{n} \tau_p. \quad (6.17)$$

Отсюда при  $\omega = 4 \cdot 10^{15} \text{ с}^{-1}$ ,  $a = 50 \text{ нм}$ ,  $P = 10 \text{ мВт}$  и разумных значениях  $\alpha = 10$ ,  $\Phi_f = 0,5$ ,  $\tau_p = 1 \cdot 10^{-2} \text{ с}$ ,  $\sigma = 2 \cdot 10^{-16} \text{ см}^2$ ,  $\beta/n = 1 \cdot 10^{-2}$ ,  $\eta = 1 \cdot 10^{-6}$  находим

$$J = 30 \gamma_s N. \quad (6.18)$$

Сравнение формул (6.16), (6.18) показывает, что выбранные условия наблюдения позволяют обнаружить единичные молекулы ( $N=1$ ), если  $\gamma_s = 1$ , т.е. если детектор настроен точно на частоту флуоресценции. При переходе от (6.17) к (6.18) мы приняли  $\eta = 1 \cdot 10^{-6}$ , предполагая тем самым, что речь идет о зонде с удлинённым конусом, образующимся при растяжении волокна. Переход к технологиям химического травления волокна, как было показано выше, позволяет увеличить  $\eta$  на 2...3 порядка при том же радиусе  $a$  диафрагмы. Этого выигрыша достаточно, чтобы обеспечить надежное превышение сигнала над шумом и при спектральном анализе при типичном  $\gamma_s \sim 10^{-2}$ . Подавлению шумов может способствовать также увеличение радиуса диафрагмы и замедление скорости развертки.

Приведенные оценки уровня сигнала указывают на отсутствие принципиальных препятствий к проведению в БСОМ спектрального анализа на молекулярном уровне.

Опыты по наблюдению отдельных молекул красителей и анализу изменения их спектров при непрерывном и импульсном лазерном возбуждении относят к числу наиболее значительных достижений БПО [4].

Заметим, что хотя размер различимых в БСОМ молекул составляет около 1 нм, их изображение представляется увеличенным примерно в  $10^2$  раз, причем искажение истинного размера растет пропорционально радиусу  $a$  диафрагмы. Молекулы, отстоящие одна от другой на расстоянии, меньшем  $2a$ , выглядят как одна частица. Проблему разрешения плотноупакованных молекулярных структур можно пытаться решить, используя реакцию детектора на взаимодействие молекул образца с одной или несколькими оптически активными молекулами, нанесенными на кончик острия.

## 6.5. Нагрев зонда.

Из-за малого коэффициента пропускания (или энергетической эффективности) апертурных зондов практически вся мощность излучения, вводимая в волоконный световод, поглощается в металлической оболочке. В одномодовом волокне, начиная с сечения отсечки радиусом  $\rho_0$ , интенсивность прошедшего излучения резко падает и основная часть электромагнитной энергии отбрасывается в обратном направлении. Если бы каждое столкновение рассеянных фотонов со стенками приводило к поглощению, можно было бы ожидать преимущественного нагрева зонда на расстоянии от вершины  $z_0 \approx \rho_0 / \operatorname{tg} \theta$ , где  $\theta$  – апертурный угол (половина угла раствора конуса). В действительности поглощение света происходит лишь после многократных отражений от зеркальной границы раздела кварц-алюминий и область максимального нагрева смещается на расстояние  $z_1 \approx p z_0$ , где  $p$  – среднее число отражений. Учитывая, что для мод низших порядков  $\rho_0 = 0,1 \dots 0,5$  мкм и приняв  $p \approx 10$ , находим  $z_1 = 10 \dots 50$  мкм при  $\operatorname{tg} \theta = 0,1$ .

Оценим повышение температуры покрытого алюминием заостренного кварцевого волокна в области, отстоящей на расстоянии  $z_1$  от вершины [24]. На рис. 28 указаны направления основных тепловых потоков от нагретой отраженным светом области зонда к более холодным участкам и в окружающую среду. Обусловленный теплопроводностью поток  $Q_1$  вблизи вершины распространяется в основном по алюминиевому покрытию толщиной  $d_{Al} \approx 100$  нм, коэффициент теплопроводности  $\chi_{Al}$  которого на два порядка превосходит коэффициент теплопроводности  $\chi_q$  кварца.

Выравнивание тепловых потоков по диэлектрической сердцевине и металлической оболочке зонда происходит на расстоянии от вершины  $z_2 \approx 100$  мкм, когда возросшее сечение сердцевины компенсирует ее меньшую теплопроводность:  $\chi_q \cdot z_2 \cdot \operatorname{tg} \theta \approx \chi_{Al} \cdot d_{Al}$ . Таким образом, на отрезке  $z_1 \leq z \leq z_2$  применимо приближенное соотношение, описывающее распространение тепла по алюминиевой оболочке зонда

$$Q_1 \approx -2\pi r d_{Al} \chi_{Al} \frac{dT}{dz}, \quad (6.19)$$

где  $Q_1$  – переносимая тепловая мощность,  $r = z \cdot \operatorname{tg} \theta$  – радиус сечения конуса на расстоянии  $z$  от вершины.



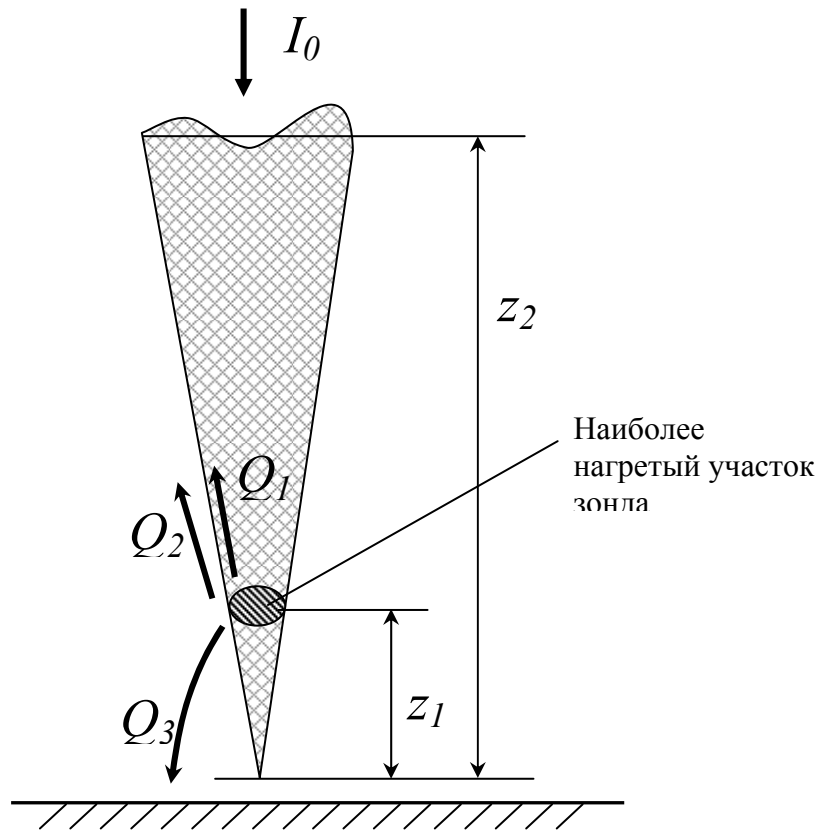


Рис. 28. Схема переноса тепла от участка зонда, нагреваемого отраженным и рассеянным излучением.

Из формулы (6.19) находим разность температур в двух сечениях зонда

$$\Delta T = T(z_1) - T(z_2) = \frac{Q_1}{2\pi\chi_{Al}d_{Al}\operatorname{tg}\vartheta} \ln \frac{z_2}{z_1} \quad (6.20)$$

Нам остается выяснить, какую часть мощности  $P$  излучения, теряемой в волокне, переносит поток  $Q_1$ . Измерения и численные оценки показывают, что вызванный конвекцией поток  $Q_2$  вдоль поверхности зонда и направленный к образцу поток  $Q_3$  соизмеримы с  $Q_1$ . На это указывает, в частности, обнаруженная экспериментально возможность уменьшения нагрева зонда в несколько раз при его сближении с проводящим образцом.

Мы вправе принять, например,  $Q_1 \approx Q_2 + Q_3$  или  $Q_1 \approx 0,5P$ . Подстановка в формулу (6.13)  $d_{Al}=100$  нм,  $\chi_{Al}=220$  Вт/м·К,  $z_2=100$  нм,  $z_1=20$  нм дает

$$\Delta T/P \approx 30 \text{ К/мВт}.$$

Согласно последнему соотношению плавление алюминиевой оболочки зонда ( $\Delta T \approx 600\text{К}$ ) должно происходить при критической мощности  $P_c \approx 20$  мВт, если считать температуру зонда в сечении  $z_2$  близкой к комнатной. В действительности предельная мощность может быть еще ниже ( $P_c \approx 10$  мВт), т.к. разрушение зондов из-за растрескивания оболочки происходит при температурах, меньших точки плавления алюминия. Последняя величина соответствует усредненным результатам многочисленных измерений критической мощности, что оправдывает приближения, использованные при оценке потока  $Q_1$ . Опасность разрушения можно уменьшить, используя зонды с большими углами раствора конуса или со специально подобранными покрытиями.

Если образец, например, микросхема нагревается в процессе работы сильнее, чем зонд, направление потока  $Q_3$  меняется на противоположное. Из-за термического расширения вершины зонда расстояние между ней и образцом уменьшается тем сильнее, чем больше локальный нагрев образца. Эффект существенно возрастает, если вершину полностью покрыть металлом. На этом принципе, поясняемом рис. 29, основано применение зондов в качестве микроболометров, обладающих чувствительностью  $\sim 10^{-1}$  нм/мкВт при разрешении по горизонтали до 1 мкм.

Ограничение мощности, подводимой к образцу по апертурному зонду, величиной  $\sim \eta P_c$  существенно затрудняет развитие зондовых методов лазерной обработки материалов.

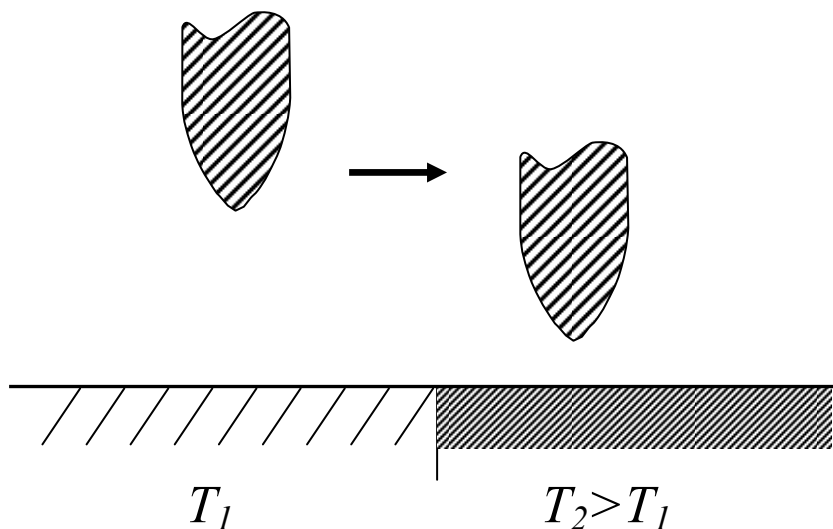


Рис. 29. Схема, поясняющая возможность обнаружения перепада температуры вдоль поверхности образца. Из-за термического расширения зонда расстояние между ним и горячим участком поверхности уменьшается.

#### 6.6. Расширение зоны светового воздействия из-за диффузионных процессов в образце

Реализуемое на практике разрешение БСОМ может существенно отличаться от рассчитанного, исходя из размеров и формы вершины острия. Дело в том, что реакция образца на воздействие локализованного зондом электромагнитного поля обычно распространяется на площадь, значительно превышающую размер светового пятна. Расширение зоны светового воздействия является следствием различных процессов, таких как диффузия фотовозбужденных носителей заряда, перенос тепла в приповерхностном слое и др.

Поясним этот эффект на примере, относящемся к одной из перспективных областей применения БСОМ – исследованию фотолюминесценции полупроводниковых наноструктур. Задачей оператора

БСОМ является различие областей с характерными размерами  $l \sim 10$  нм, отличающихся от окружающей матрицы по электронным или оптическим свойствам. Такими областями могут быть, например, квантовые точки – возникающие из-за флуктуаций толщины активного слоя или выращиваемые специальными методами наночастицы с типичной плотностью на поверхности  $\sim 10^{10} \text{ см}^{-2}$ . Создаваемые фотонами неравновесные носители образуют электрон-дырочные пары и рекомбинируют через время  $\tau \sim 10^{-9}$  с, генерируя вторичное излучение. За это время они смещаются на расстояние  $L_{dif} \approx (D\tau)^{1/2}$ , где  $D$  – коэффициент диффузии (такая зависимость  $L_{dif}$  от  $\tau$  является общей для всех диффузионных процессов). Настроивая приемник излучения на частоту фотолюминесценции квантовых точек, можно получить их изображение в виде светлых пятен радиусом  $R \approx L_{dif}$ , выделяющихся на более темном фоне матрицы.

Чтобы оценить  $D$  и  $L_{dif}$ , используем соотношение Эйнштейна, связывающее коэффициент диффузии носителей с их подвижностью  $\mu$ :  $D = \mu kT/e$ , где  $k$  – постоянная Больцмана,  $T$  – температура,  $e$  – заряд электрона. Обычно исследования фотолюминесценции полупроводниковых наноструктур проводят при гелиевых температурах. Для наименее подвижных носителей – дырок при  $T = 5\text{ К}$  и характерных значениях  $\tau = 1 \cdot 10^{-9}$  с,  $\mu = 1 \cdot 10^3 \text{ см}^2/\text{В} \cdot \text{с}$ , находим  $D = 0,4 \text{ см}^2/\text{с}$ ,  $L_{dif} = 0,2 \text{ мкм}$ . Последняя величина на порядок превышает минимальный радиус светового пятна в апертурном БСОМ и характерный размер квантовых точек. В связи с этим полученное изображение воспроизводит искомое распределение наночастиц в существенно искаженном виде.

Принцип образования изображения поясняет рис. 30. При сборе фотонов линзой детектор, настроенный на частоту  $\omega_{PL}$  фотолюминесценции, регистрирует свечение образца при всех положениях зонда, когда расстояние  $r$  между осью зонда и квантовой точкой удовлетворяет неравенству  $r \leq a + L_{dif}$  (рис. 30а). В результате частица размером  $l$  превращается на изображении в круг диаметром  $2(a + L_{dif})$ . При  $L_{dif} \gg l$  использование предельно узких зондов нецелесообразно, так как уменьшение радиуса  $a$  диафрагмы мало сказывается на видимом размере частицы. Лучшее пространственное разрешение дает схема, в которой зонд является одновременно источником и приемником фотонов (рис. 30б). Недостатком такой схемы является ослабление интенсивности сигнала из-за сложения потерь мощности при распространении излучения в прямом и обратном направлениях.

Рассеяние фотовозбужденных носителей на ионах кристаллической решетки сопровождается ее нагревом. Скорость распространения тепла из зоны возбуждения определяется коэффициентом температуропроводности  $\alpha_T$

$= \chi / \rho_m c_p$ , где  $\chi$  – теплопроводность образца,  $\rho_m$  – его плотность,  $c_p$  – теплоемкость. За время действия импульса длительностью  $\tau_T$  нагреваемая область расширяется до радиуса, определяемого соотношением  $L_T \approx (a_T \tau_T)^{1/2}$ , по аналогии с формулой, связывающей длину и время диффузии.

Для широкого круга материалов (за исключением лучших теплоизоляторов)  $a_T \geq 10^{-2} \text{ см}^2/\text{с}$ , а значит, в тех случаях, когда необходимо сжать зону светового воздействия до размеров  $\sim 0,1 \text{ мкм}$ , возбуждение должно носить импульсный характер с длительностью импульса  $\tau < (0,1 \text{ мкм})^2 / (0,01 \text{ см}^2/\text{с}) \sim 10^{-8} \text{ с}$ . Распространение тепла можно рассматривать как следствие диффузии фононов – тепловых колебаний решетки. Высокая скорость диффузии препятствует использованию БСОН для записи информации с высокой плотностью путем прямого термического воздействия излучения на образец при длительности импульса, превышающей несколько наносекунд. Опыт показывает, что лучше подходят другие методы, например, вдавливание нагретого острия в термопластичный образец или создание холмиков на поверхности образца за счет переноса атомов с острия.

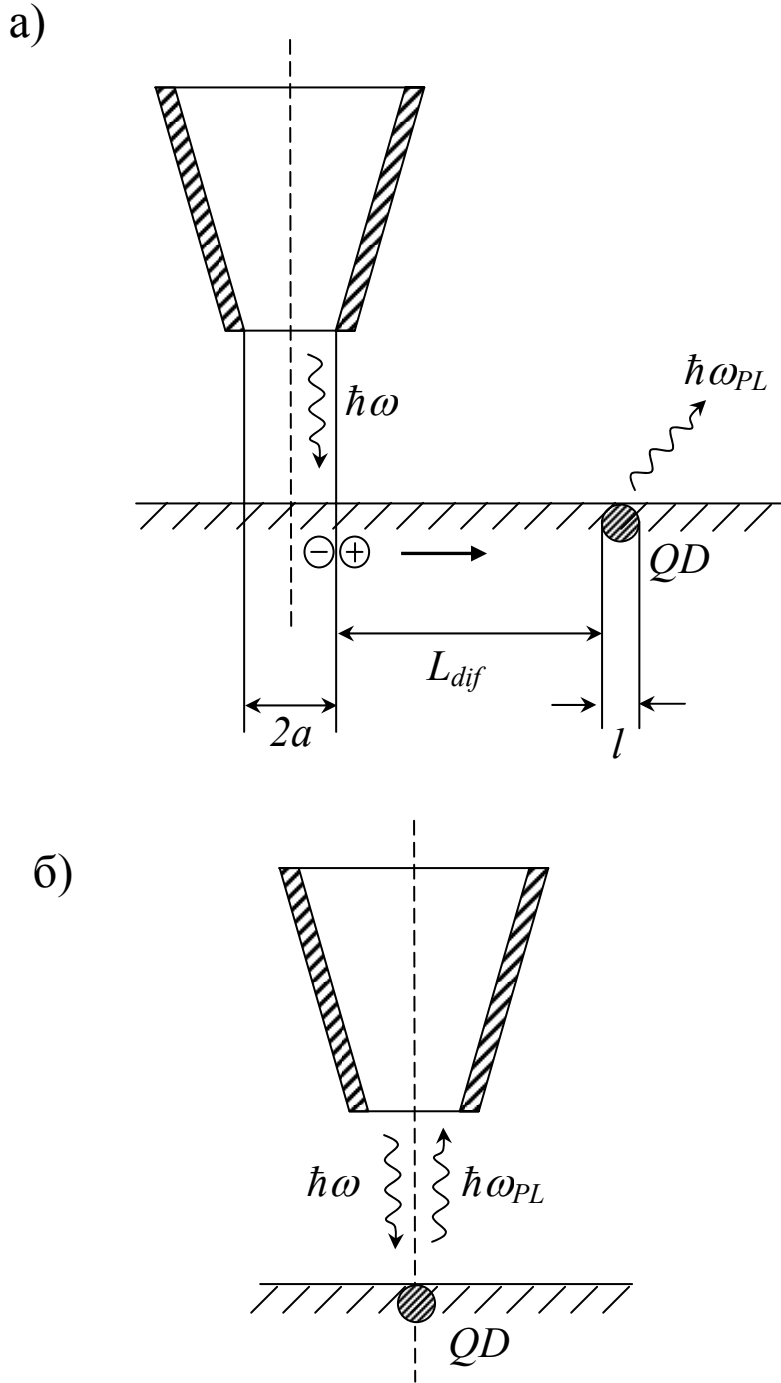


Рис. 30. Два способа наблюдения в БСОМ квантовой точки (QD). Излучение на частоте фотолуминесценции  $\omega_{PL}$  собирается линзой с широкого участка образца (схема а) или зондом непосредственно над квантовой точкой (схема б). В случае (а) фотовозбужденные носители смещаются до рекомбинации на расстояние  $L_{dif}$ , что приводит к свечению QD при любом положении зонда в пределах круга радиусом  $a+L_{dif}$ .

## 7. Заключение

Хотя БПО относится к числу самых молодых разделов физики, она уже завоевала широкое признание. Помимо очевидного общенаучного интереса, развитие исследований стимулируется перспективой решения важных прикладных задач [25].

Одно из наиболее интересных направлений – целенаправленное воздействие сверхтонкого оптического зонда на избранные участки образца. Использование элементов БПО в уже известных методах модификации и анализа поверхности может существенно расширить возможности этих методов и дать новый толчок к их развитию. Так, в демонстрационных опытах, проведенных с помощью БСОМ, удалось в 20 раз повысить плотность записи информации на магнитооптических средах, тогда как многолетние предшествующие усилия сопровождались темпами роста в среднем на 60% в год. Введение образца в ближнее поле источника излучения при регистрации наведенного фототока позволило почти на порядок повысить точность локализации дефектов в полупроводниках по сравнению с используемыми ранее методами.

Большие надежды связывают с развитием высоколокального спектрального анализа. Оптическое зондирование квантовых гетероструктур позволило анализировать их энергетику на уровне, доступном ранее лишь для атомной физики. Перспективно привлечение спектральных методов БПО к выявлению флуоресцентных меток в биологических тканях и молекулах ДНК.

Широкое распространение БСОМ в течение ряда лет сдерживалось трудностью конкуренции с другими СЗМ, обладающими существенно лучшим разрешением. Сейчас стало очевидным, что БСОМ наиболее эффективен как часть комбинированного прибора, способного в зависимости от материала образца и типа зонда выполнять также функции АСМ, СТМ и др. Ценность такого прибора определяется в большей степени не резкостью оптического изображения, а важностью содержащейся в этом изображении дополнительной информации.

Мы рассмотрели физические явления, лежащие в основе механизма образования изображения в БСОМ и позволяющие реализовать “сверхразрешение”. Не меньший интерес представляет и чисто техническая сторона проблемы. Очевидно, что задача перемещения зонда на высоте порядка 1 нм над образцом с одновременной регистрацией оптического сигнала не является тривиальной. Для ее решения предложены многочисленные варианты схем, отличающихся как конструктивными особенностями, так и принципом детектирования сигнала. Большое внимание

уделяется оптимизации оптических зондов с целью достижения высокой выходной мощности при малом размере зоны засветки. Значительные успехи достигнуты в разработке многоцелевых приборов, способных не только создавать изображение поверхности, но и решать задачи, относящиеся к записи и считыванию информации, нанолитографии, спектроскопии с высоким пространственным и временным разрешением и т.д. Эти вопросы заслуживают отдельного обсуждения.

### Список литературы.

1. E.Synge. *Phil.Mag.*, **6**, 356-362, (1928)
2. D.W.Pohl, W.Denk, M.Lanz. *Appl.Phys.Lett.*, **44**, 651-654, (1984)
3. E.Betzig, J.K.Trautman. *Science*, **257**, 189-195, (1992)
4. E.Betzig, J.K.Trautman et al. *J.Appl.Phys.*, **73**, 5791-5797, (1993)
5. K.Dickmann, J.Jersch. *Laser und Optoelektronik*, **27**, 76-83, (1995)
6. R.Dunn. *Chem.Rev.*, **99**, 2891-2927, (1999)
7. R.Bachelot, P.Gleizes, A.C.Voccara. *Appl.Opt.*, **36**, 2160-2170, (1997)
8. Дж.А.Стрэттон. Теория электромагнетизма. ГИТТЛ, М.-Л., (1948)
9. Г.С.Ландсберг. Оптика. “Наука”, М., (1976)
10. Р.Спроул. Современная физика. ГИФМЛ, М., (1961)
11. М.Н.Либенсон, В.С.Макин, С.Д.Пудков. Поверхностные электромагнитные волны в оптике. Знание, Л., (1990)
12. R.C.Reddick, R.J.Warmack et al. *Rev.Sci.Instr.*, **61**, 3669-3773, (1990)
13. M.Specht, J.D.Pedarnig et al. *Phys.Rev.Lett.*, **68**, 476-479, (1992)
14. C.J.Bouwkamp. *Rep.Prog.Phys.*, **17**, 35-100, (1954)
15. D.W.Pohl, U.C.Fisher, U.T.Durig. *J.Microscopy*, **152**, 853-860, (1988)
16. E.Wolf, M.Nieto-Vesperinas. *J.Opt.Soc.Amer. A*, **2**, 886-889, (1985);
17. Дж.Гудмен. Введение в фурье-оптику. “Мир”, М., (1970)
18. A.Roberts. *J.Appl.Phys.*, **65**, 2886-2889, (1989)
19. Г.С.Жданов. М.Н.Либенсон, Г.А.Марциновский. *Усп.физ.наук*, **168**, 801-804, (1998)
20. J.Tersoff, D.R.Hamann. *Phys.Rev.B*, **31**, 805-813, (1985)
21. G.S.Zhdanov. *Proc.SPIE*, **5506**, 206-214, (2004)
22. L.Novotny, D.W.Pohl, P.Regli. *J.Opt.Soc.Amer. A*, **11**, 1768-1779, (1994)
23. Г.С.Жданов. *Поверхность*, №7, 91-94, (1998)
24. Либенсон М.Н., Марциновский Г.А., *Поверхность*, 2000, № 11, С.80-83.
25. Zhu X., M.Ohtsu (eds.). *Near-Field Optics: Principles and Applications*. “World Sci.Publ.”, Singapore, (1999)



## ОГЛАВЛЕНИЕ

Предисловие.....	3
1. Введение. История возникновения и развития ближнепольной оптики .....	4
2. Принцип действия и основные узлы БСОМ .....	5
3. Распространяющиеся и затухающие волны – две стороны физических явлений.....	12
3.1. Волновые процессы .....	12
3.2. Уравнения Максвелла. Электромагнитные волны .....	14
3.3 Физический смысл комплексного волнового вектора .....	18
3.4. Корпускулярно-волновой дуализм. Уравнение Шредингера. Волновая функция .....	21
3.5. Туннельный эффект .....	23
3.6. Туннелирование фотонов. Поверхностные электромагнитные волны ..	27
3.7. Структура поля в волноводах. Отсечка .....	31
3.8. Электромагнитное поле осциллятора .....	34
3.9. Прохождение света через малую диафрагму .....	39
3.10. Усиление электромагнитного поля у вершины зонда .....	43
3.11. Определение термина “ближнее поле”. Механизм переноса энергии в системе источник - зонд – детектор .....	46
4. Элементы фурье-оптики ближнего поля .....	49
4.1. Метод разложения по плоским волнам .....	49
4.2. Спектр пространственных частот шероховатой поверхности .....	52
4.3 Пары преобразований Фурье, часто встречающиеся в прикладной оптике .....	54
5. Пространственное разрешение оптических приборов. Преодоление дифракционного предела разрешения .....	58
5.1. Теория Аббе и критерий Рэля .....	58
5.2. Передаточная функция БСОМ .....	61
5.3. Предел разрешения и принцип неопределенности .....	68
5.4. Пространственное разрешение при туннелировании .....	69
6. Принципиальные и технические ограничения основных параметров БСОМ.....	76
6.1. Пространственное разрешение и контраст. Влияние поляризации света и оптических характеристик системы зонд-образец .....	76
6.2. Трудность разделения оптической и атомно-силовой составляющих сигнала.....	84
6.3. Потери мощности в оптическом зонде .....	87
6.4. Предельная чувствительность при различении малых частиц и спектральном анализе .....	95

6.5. Нагрев зонда .....	98
6.6. Расширение зоны светового воздействия из-за диффузионных процессов в образце .....	101
7. Заключение .....	105
Список литературы .....	106