

Структура световых волн в волноводе, сужающемся до субволновых поперечных размеров

Т.И.Кузнецова, В.С.Лебедев

Рассчитано просачивание Н-волны через волновод, радиус которого плавно изменяется от величины порядка длины волны до субволновых размеров (порядка $1/10$ длины волны). Разработан метод анализа полей и на численных примерах изучена область его применимости. Исследована зависимость коэффициента пропускания сужающегося волновода от длины волны излучения и крутизны профиля волновода. Показано, что более резкое уменьшение радиуса обеспечивает большие коэффициенты пропускания поля.

Ключевые слова: оптика ближнего поля, сужающийся волновод, коэффициент пропускания волновода.

1. Введение

В последние годы был выполнен ряд исследований, относящихся к прохождению излучения через системы, имеющие субволновую апертуру. Практический интерес к этой проблематике связан с развитием оптической микроскопии сверхвысокого разрешения, в основе которой лежит локализация поля в области субволновых размеров. В классических теоретических работах по этой тематике подробно изучалась дифракция на круглом отверстии [1–3]. В первых работах по сканирующему оптическому микроскопу ближнего поля «квазиточечный» источник света создавался с помощью отверстия в непрозрачном экране [4]. Исследовались также и более протяженные структуры, например микропипетки [5, 6]. Особенно эффективным оказалось использование сужающихся волокон [7, 8], к тому же выяснилось, что целесообразно напылять слой металла на боковую поверхность микропипетки или волокна. К указанной тематике относится также ряд теоретических работ (см., напр., [9–12] и приведенные там ссылки).

Ориентируясь на экспериментальные исследования, в которых использовались волокна с металлическими покрытиями, мы будем рассматривать диэлектрический волновод с идеально отражающими металлическими стенками. При этом мы сосредоточим внимание на изменении поперечного сечения в направлении оптической оси и его влиянии на структуру поля. По аналогии со свойствами цилиндрических волноводов (см., напр., [13]) естественно представить, что в волноводе переменного сечения может проявиться так называемая отсечка: при уменьшении диаметра волновода прекращается каналирование поля и наблюдается сильное его затухание вдоль продольной координаты.

Представляет интерес количественная оценка этого затухания. Подобная оценка была сделана ранее в [14]. В

этой работе для расчета амплитуды прошедшей волны использовался локальный коэффициент затухания. Он вводился по таким же правилам, как и в круглом цилиндрическом волноводе, однако в волноводе переменного радиуса этот коэффициент оказывался зависящим от продольной координаты. Для расчета полного затухания проводилось интегрирование локального коэффициента. При таком подходе продольные изменения параметров волновода учитываются слишком упрощенно. Дело в том, что в сужающемся волноводе возникают дополнительные волны, существенно изменяющие структуру полного поля, которые заранее нельзя исключить из рассмотрения.

Здесь уместно обратиться к аналогии со случаем прохождения световой волны через отверстие в отражающем экране [1]. В работе [1] было установлено, что возникновение отраженной волны в области перед экраном является существенным фактором, влияющим на просачивание поля через малое отверстие. Этот фактор сильно изменяет коэффициент пропускания по сравнению с тем, который мог бы быть получен при расчетах, основанных на принципе Гюйгенса.

В настоящей работе развита теория сужающегося волновода, включающая в рассмотрение наряду с прямой отраженную волну, оценивается также трансформация исходного поля в волны высших типов. Метод исследования изложен применительно к волноводам, имеющим круглое сечение с переменным радиусом. Разработанный метод примыкает к так называемому методу поперечных сечений, развитому в [15,16], но при этом учитывает специфику волокон, применяемых в оптическом микроскопе ближнего поля: малые радиусы, сильные отличия профиля от цилиндрического, резкое уменьшение радиуса на выходном участке до размеров, существенно меньших длины волны, сильное отражение от стенок.

2. Формулировка задачи и основные уравнения

Приведем исходные уравнения и граничные условия для поля в круглом волноводе с переменным радиусом

Физический институт им. П.Н.Лебедева РАН, Россия, 119991 Москва, Ленинский просп., 53;
e-mail: tkuzn@mail1.lebedev.ru, vlebedev@mail1.lebedev.ru

Поступила в редакцию 9 января 2002 г., после доработки – 31 мая 2002 г.

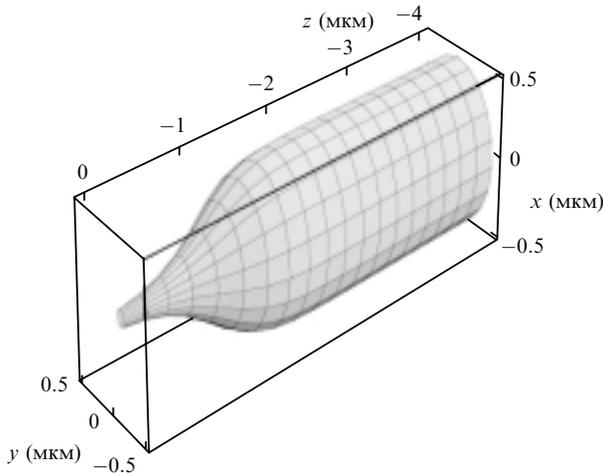


Рис.1. Вид сужающегося волновода гипергеометрического типа при входном и выходном радиусах $a_{in} = a_{-\infty} = 0.5$ мкм, $a_{out} = a(z = 0) = 0.05$ мкм и параметре волновода $l = 0.2$ мкм.

(рис.1). Пусть z – координата вдоль оси волновода, ρ – расстояние до оси волновода, φ – полярный угол, a – зависящий от продольной координаты внутренний радиус волновода. Среда внутри волновода однородна и имеет диэлектрическую проницаемость $\epsilon = \text{const}$. Будем считать стенки волновода идеально отражающими. Ограничимся рассмотрением полей, компоненты которых не зависят от угла φ . Зависимость компонент от времени выберем в виде $\exp(-i\omega t)$ и далее в выражениях для полей этот временной множитель опустим. При этом система уравнений Максвелла для волн магнитного типа сводится к уравнению

$$\frac{\partial^2 E}{\partial z^2} + \frac{\partial}{\partial \rho} \left[\frac{1}{\rho} \frac{\partial}{\partial \rho} (\rho E) \right] + \epsilon \frac{\omega^2}{c^2} E = 0, \tag{1}$$

причем отличные от нуля компоненты поля связаны соотношениями

$$E_\varphi = E, \quad H_\rho = i \frac{c}{\omega} \frac{\partial E}{\partial z}, \quad H_z = -i \frac{c}{\omega} \frac{1}{\rho} \frac{\partial (\rho E)}{\partial \rho}. \tag{2}$$

Граничное условие имеет вид

$$E_\varphi|_{\rho=a(z)} = 0. \tag{3}$$

В сформулированной задаче основная особенность состоит в том, что граница волновода не является цилиндрической поверхностью: $a(z) \neq \text{const}$.

Мы будем строить решение уравнения Гельмгольца (1) для волн магнитного типа, используя в каждом поперечном сечении разложение по полной системе функций. Различным продольным координатам будут соответствовать свои собственные функции. Будем работать с функциями $J_1(q_n \rho)$, где J_1 – функция Бесселя первого рода первого порядка, q_n – поперечное волновое число, определяемое соотношениями

$$q_n(z) = \frac{\xi_n}{a(z)}, \quad J_1(\xi_n) = 0 \quad (n = 1, 2, \dots), \tag{4}$$

причем корни ξ_n пронумерованы в порядке их возрастания. Из-за зависимости $a(z)$ собственные функции оказы-

ваются зависящими от z как от параметра. В каждом сечении z они образуют полную ортонормированную систему и удовлетворяют граничному условию (3). На целесообразность использования такой системы функций было указано в [15, 16] (см. также [17], с. 435).

Представим решение уравнения (1) в виде

$$E = \sum_{n=1}^{\infty} A_n(z) \Phi_n(\rho, z), \quad \Phi_n(\rho, z) = J_1[q_n(z)\rho]. \tag{5}$$

Отметим, что в разложении (5) нет разделения поля на прямые и обратные волны (в отличие от подхода, использованного в [11, 15, 16]). Это означает, что каждый из коэффициентов A_n в (5) включает в себя, по существу, и прямую и отраженную волны. В дальнейшем коэффициенты A_n будут называться амплитудами мод ($n = 1, 2, \dots$). Введенные таким образом амплитуды A_n могут принимать как положительные, так и отрицательные значения; эти амплитуды зависят от продольной координаты z . Нахождению зависимости $A_n(z)$ посвящены этот и два последующих раздела.

Разложим производную $\partial \Phi_n / \partial z$ по функциям Φ_m :

$$\frac{\partial \Phi_n}{\partial z} = \sum_{m=1}^{\infty} c_{nm} \Phi_m, \quad c_{nm} = \frac{1}{N} \int_0^a \frac{\partial \Phi_n}{\partial z} \Phi_m \rho d\rho, \tag{6}$$

$$N = \int_0^a \Phi_m^2 \rho d\rho.$$

Учитывая (6), напомним разложение $\partial^2 \Phi_n / \partial z^2$ по функциям Φ_m :

$$\frac{\partial^2 \Phi_n}{\partial z^2} = \sum_{m=1}^{\infty} \left(\Phi_m \frac{dc_{nm}}{dz} + c_{nm} \sum_{p=1}^{\infty} c_{mp} \Phi_p \right). \tag{7}$$

Подставим (4), (5) и (6) в (1) и в полученном уравнении приравняем нулю коэффициенты при каждой из собственных функций. Тогда

$$\frac{d^2 A_n}{dz^2} + A_n \left[\epsilon \frac{\omega^2}{c^2} - q_n^2(z) \right] + 2 \sum_{n=1}^{\infty} \frac{dA_n}{dz} c_{np} + \sum_{n=1}^{\infty} A_n \left(\frac{dc_{np}}{dz} + \sum_{m=1}^{\infty} c_{nm} c_{mp} \right) = 0. \tag{8}$$

Выражение для коэффициентов c_{nm} (см. (6)) имеет вид

$$c_{nm} = \left(\frac{1}{a} \frac{da}{dz} \right) \tilde{c}_{nm}, \quad \tilde{c}_{nm} = \begin{cases} -\frac{2\xi_n \xi_m}{\xi_n^2 - \xi_m^2} \frac{J_0(\xi_n)}{J_0(\xi_m)}, & n \neq m, \\ 1, & n = m. \end{cases} \tag{9}$$

Здесь $J_0(z)$ – функция Бесселя первого рода нулевого порядка. Поскольку все коэффициенты c_{nm} одинаково зависят от координаты z (см. (9)) наиболее естественным для дальнейшего анализа оказывается введение новых амплитуд волн $Z_n(z) = \text{const} \cdot a(z) A_n(z)$. Полагая константу равной $1/a_{-\infty}$, получаем

$$A_n(z) = \frac{a_{-\infty}}{a(z)} Z_n(z), \tag{10}$$

и на входе ($z \rightarrow -\infty$) функции Z_n и A_n совпадают.

С помощью (10) приводим систему (8) к виду

$$\frac{d^2 Z_p}{dz^2} + Z_p \left[\varepsilon \frac{\omega^2}{c^2} - q_p^2(z) \right] + Z_p \left(\frac{1}{a} \frac{da}{dz} \right)^2 \sum_{m \neq p}^{\infty} \tilde{c}_{pm} \tilde{c}_{mp} = \psi_p(z), \quad (11)$$

где

$$\psi_p = -2 \frac{1}{a} \frac{da}{dz} \sum_{n \neq p}^{\infty} \frac{dZ_n}{dz} \tilde{c}_{np} - \left[\frac{d}{dz} \left(\frac{1}{a} \frac{da}{dz} \right) \right] \sum_{n \neq p}^{\infty} Z_n \tilde{c}_{np} - \left(\frac{1}{a} \frac{da}{dz} \right)^2 \sum_{n \neq p}^{\infty} Z_n \sum_{m \neq n, p}^{\infty} \tilde{c}_{nm} \tilde{c}_{mp}. \quad (12)$$

Сделанные в этом разделе преобразования не содержат приближений, и система уравнений (11), (12) полностью эквивалентна исходному уравнению (1). Отметим, что коэффициенты связи волн с разными индексами имеют порядок $a^{-1} da/dz$, а самовоздействие волны, возникающее из-за нерегулярности профиля волновода, пропорционально $(a^{-1} da/dz)^2$. Этого удалось добиться за счет преобразования (10). Переход от амплитуд A_n к новым независимым переменным, определяемым соотношением (10), полезен еще и в том отношении, что благодаря ему явно видно повышение плотности электромагнитной энергии из-за уменьшения сечения волновода.

3. Адиабатическое приближение

Решение системы уравнений (11), (12) может быть найдено в рамках теории возмущений по параметру $a^{-1} \times (da/dz) c(\sqrt{\varepsilon} \omega)^{-1}$. Малость этого параметра означает, что характерный продольный масштаб изменения радиуса волновода существенно превышает длину световой волны. Примем, что производную da/dz можно считать сколь угодно малой, т. е. что профиль волновода изменяется адиабатически медленно. При этом построим сначала решение в нулевом приближении, пренебрегая в (11), (12) всеми слагаемыми, содержащими $a^{-1} \times da/dz$.

Полагая правую часть в (11) равной нулю и отбрасывая в левой части (11) последний член, получаем

$$\frac{d^2 Z_p}{dz^2} + \left[\varepsilon \frac{\omega^2}{c^2} - q_p^2(z) \right] Z_p = 0. \quad (13)$$

Система уравнений (13) показывает, что в нулевом приближении все волны, отвечающие разным индексам p , являются независимыми, как это было бы и в регулярном волноводе. Однако уже здесь имеются два отличия от регулярного волновода. Первое – зависимость q_p от z , которая приводит к взаимодействию и связи двух волн, идущих в направлениях $\pm z$, а также к тому, что амплитуды этих волн изменяются при изменении z . Второе отличие от регулярного волновода состоит в том, что амплитудой волны с индексом p является не функция Z_p , а функция A_p (см. (10)), обратно пропорциональная радиусу волновода. Таким образом, даже нулевое приближение учитывает важные особенности сужающегося волновода.

Выберем следующую зависимость радиуса волновода от продольной координаты z :

$$a(z) = \left[\frac{1 + \exp(-z/l)}{(1/a_{\infty})^2 + (1/a_{-\infty})^2 \exp(-z/l)} \right]^{1/2}, \quad (14)$$

где l – параметр, характеризующий область сужения волновода. Формула (14) описывает постепенно сужающееся волокно, радиус которого изменяется от $a_{\text{in}} \equiv a_{-\infty}$ на входном конце при $z \rightarrow -\infty$ до a_{∞} при $z \rightarrow +\infty$. Рассмотрим также точку $z = 0$, поскольку область $-\infty < z < 0$ хорошо аппроксимирует протяженное волокно с заостренным выходным концом. Радиус волновода при $z = 0$ будем считать выходным радиусом a_{out} , для которого имеем выражение

$$a_{\text{out}} \equiv a(0) = a_{\infty} \left[\frac{2}{1 + (a_{\infty}/a_{-\infty})^2} \right]^{1/2}. \quad (15)$$

При значительном различии величин $a_{-\infty}$ и a_{∞} выходной радиус волокна определяется соотношением $a_{\text{out}} \approx \sqrt{2} a_{\infty}$. Отметим также, что оценки по формуле (14) показывают, что существенное изменение радиуса волновода по продольной координате фактически происходит на отрезке длиной $\sim 5l$.

Волновод данного типа будем называть далее гипергеометрическим, поскольку поле в нем описывается гипергеометрическими функциями. На рис.1 изображена поверхность такого волновода, описываемая формулой (14), со следующими параметрами: $a_{\text{in}} = 0.5$ мкм, $a_{\text{out}} = 0.05$ мкм, $l = 0.2$ мкм. Выбранная зависимость (14) удобна тем, что, хорошо описывая форму сужающегося волокна, она в то же время позволяет получить в явном виде решение основного уравнения (13). Далее введем обозначения

$$m = \left(\frac{c \xi_1}{\omega \sqrt{\varepsilon} a_{-\infty}} \right)^2, \quad M = \left(\frac{c \xi_1}{\omega \sqrt{\varepsilon} a_{\infty}} \right)^2. \quad (16)$$

Здесь $\xi_1 \approx 3.832$ – наименьший ненулевой корень уравнения $J_1(\xi) = 0$, определяющий поперечное волновое число q_1 основной волны. Большинство расчетов в настоящей работе будет проведено для излучения с длиной волны $\lambda = 2\pi c/\omega = 500$ нм, кроме того, положим $\sqrt{\varepsilon} = 1.5$. Тогда, учитывая указанные выше значения a_{in} и a_{∞} , получаем соотношения $m < 1$ и $M > 1$.

После подстановки выражений (4), (14) и (16) в уравнение (13) оно принимает вид

$$\frac{d^2 Z_p}{dz^2} + \varepsilon \frac{\omega^2}{c^2} \left[1 - \frac{\xi_p^2}{\xi_1^2} \frac{M + m \exp(-z/l)}{1 + \exp(-z/l)} \right] Z_p = 0. \quad (17)$$

Решение уравнения (17) выражается через гипергеометрические функции. Оно приведено, например, в [18] при обсуждении задачи об отражении частицы от одномерного потенциального барьера.

Остановимся прежде всего на рассмотрении волны низшего типа. В этом случае ($p = 1$) уравнение (17) имеет два линейно независимых решения:

$$U(z) = \exp(zX/l) \times {}_2F_1[-X + iY, -X - iY; 1 - 2X; -\exp(-z/l)], \quad (18)$$

$$V(z) = \exp(-zX/l) \times {}_2F_1[X + iY, X - iY; 1 + 2X; -\exp(-z/l)]. \quad (19)$$

Здесь ${}_2F_1$ – гипергеометрическая функция;

$$X = \frac{\omega\sqrt{\epsilon}l}{c} (M - 1)^{1/2}; \quad Y = \frac{\omega\sqrt{\epsilon}l}{c} (1 - m)^{1/2}. \quad (20)$$

Отметим, что при $z \rightarrow +\infty$ функция $U(z)$ возрастает, а функция $V(z)$ убывает. В качестве решения уравнения (17), описывающего амплитуду основной волны ($p = 1$) при отсутствии каких-либо излучающих или усиливающих источников на $+\infty$, мы выбираем затухающую на $+\infty$ функцию

$$Z_1(z) = V(z). \quad (21)$$

Приведем асимптотические выражения для функции (19) при $z \rightarrow \pm\infty$. При $z \rightarrow \infty$ функция $V(z)$ принимает вид

$$V(z) = \exp\left(-\frac{zX}{l}\right), \quad \exp\left(-\frac{z}{l}\right) \ll 1. \quad (22)$$

При $z \rightarrow -\infty$ функция $V(z)$ представляет собой две бегущие волны, распространяющиеся навстречу друг другу:

$$V(z) = \frac{\Gamma(1 + 2X)\Gamma(-i2Y)}{\Gamma(X - iY)\Gamma(1 + X - iY)} \exp\left(\frac{iZY}{l}\right) + \frac{\Gamma(1 + 2X)\Gamma(i2Y)}{\Gamma(X + iY)\Gamma(1 + X + iY)} \exp\left(-\frac{iZY}{l}\right), \quad (23)$$

$$\exp\left(\frac{z}{l}\right) \ll 1.$$

Здесь $\Gamma(x)$ – полная гамма-функция. Вводя обозначение

$$A_{in} = \frac{\Gamma(1 + 2X)\Gamma(-i2Y)}{\Gamma(X - iY)\Gamma(1 + X - iY)}, \quad (24)$$

вместо (23) можно написать

$$V(z) = A_{in} \exp\left(\frac{iZY}{l}\right) + A_{in}^* \exp\left(-\frac{iZY}{l}\right). \quad (25)$$

Равенство модулей амплитуд прямой и отраженной волн связано с тем, что в узкой части волновода ($z \rightarrow \infty$) поток электромагнитной энергии обращается в нуль (см. асимптотическое выражение (22)). При $z \rightarrow -\infty$ поле имеет вид стоячей волны. Характеристики поля при промежуточных значениях z , полученные по формулам (19) и (21), будут обсуждаться в следующем разделе.

4. Численный анализ характеристик поля

Для более точного описания зависимости поля в волноводе от продольной координаты z следует вернуться к амплитуде A_n , которая выражается через построенное выше решение и радиус $a(z)$ (14). Для первой волны ($n = 1$) соотношение (10) имеет вид

$$A_1(z) = \frac{a_{in}}{a(z)} V(z). \quad (26)$$

На рис.2–4 приведены результаты расчетов для этой амплитуды, нормированной на модуль амплитуды падающей волны (24), т. е. для $A_1(z)/|A_{in}|$.

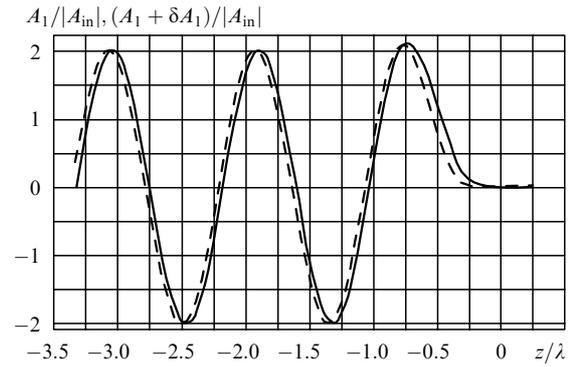


Рис.2. Зависимости нормированных амплитуд $A_1/|A_{in}|$ и $(A_1 + \delta A_1)/|A_{in}| \times |A_{in}|^{-1}$ электрического поля основной волны магнитного типа ($\lambda = 500$ нм, $\xi_1^m = 3.832$) от продольной координаты z/λ в волноводе с параметрами $a_{in} = \lambda/2$, $a_\infty = \lambda/20$ и $l = \lambda/10$. Сплошная кривая – расчет амплитуды поля в адиабатическом приближении по скорости изменения профиля волновода, штриховая кривая – расчет с учетом возмущения δA_1 (30), определяемого самовоздействием волны A_1 .

На рис.2 показан вид нормированной амплитуды в диапазоне изменения продольной координаты $-3.25 < z/\lambda < 0.25$. Сплошная кривая соответствует расчету амплитуды поля в нулевом приближении (см. (19), (24), (25)) по скорости изменения профиля волновода. Уже этот расчет наглядно демонстрирует основные закономерности поведения поля в сужающемся волноводе. Они проявляются в наличии осцилляционной структуры при отрицательных z и существенном затухании поля при увеличении z в области резкого уменьшения радиуса волновода. Практически тот же вид осцилляций сохраняется и при продвижении дальше в отрицательную область $z < -3.25\lambda$, затухание продолжается и при больших положительных $z > 0.25$ нм. Расчеты, результаты которых представлены на рис.2, проведены для $\lambda = 500$ нм и $l = 50$ нм. Результаты аналогичного расчета зависимости амплитуды поля от продольной координаты z приведены сплошной кривой на рис.3 для случая более резкого изменения ($l = 25$ нм) профиля волновода. Кроме указанных на рис.2 и 3 вариантов, расчеты были выполнены и для $l = 75, 100$ и 250 нм. Во всех случаях зависимости амплитуды поля от продольной координаты выглядели аналогично.

На рис.4 показано поведение поля основной волны магнитного типа ($\lambda = 500$ нм, $\xi_1^m = 3.832$) в той области изменения продольной координаты, где оно испытывает резкое затухание. Расчеты выполнены для четырех случаев, отвечающих различным скоростям изменения ра-

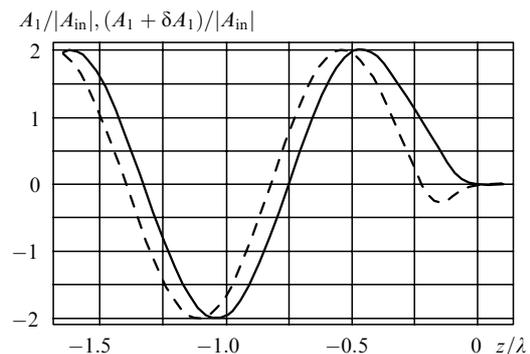


Рис.3. То же, что и на рис.2, но для параметра волновода $l = \lambda/20$.

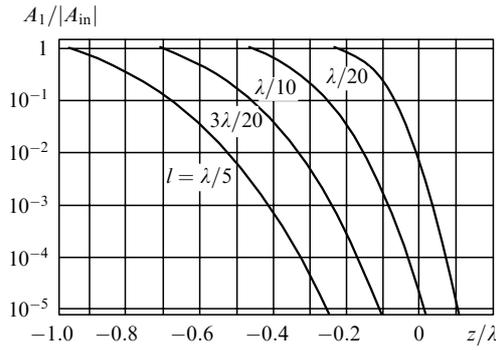


Рис.4. Затухание волн магнитного типа ($\lambda = 500$ нм, $\xi_1^m = 3.832$) с увеличением продольной координаты z в области существенного уменьшения радиуса волновода при $a_{in} = \lambda/2$ и $a_{\infty} = \lambda/20$. Кривые соответствуют результатам расчета нормированной амплитуды поля $A_1/|A_{in}|$ при различных l .

диуса волновода в продольном направлении. Полученные результаты демонстрируют чрезвычайно сильную зависимость амплитуды поля от параметра l . Так, например, в точке $z = 0$ нормированная амплитуда поля $A_1(z = 0)/|A_{in}|$ составляет соответственно 1.8×10^{-5} и 6.4×10^{-3} при $l = \lambda/10$ и $\lambda/20$. Таким образом, для более плавного изменения профиля волновода (большие l) затухание поля оказывается более сильным. Это связано с тем, что при меньшей длине переходной области затухание начинается позднее (т. е. критический радиус достигается при больших z). Следовательно, малая длина переходной области (т. е. тупые углы на выходе) всегда оказывается более предпочтительной для получения больших амплитуд выходных полей.

5. Анализ применимости нулевого приближения

Выше была построена зависимость амплитуды первой волны от продольной координаты $Z_1(z)$ на основе уравнения (17) при $p = 1$. Пусть в нулевом приближении все волны с $p \neq 1$ отсутствуют ($Z_p(z) = 0$). Исходя из этого предположения и полученных зависимостей $Z_1(z)$, определим поправку к решению, появляющуюся из-за отброшенных в (11), (12) величин, содержащих $a^{-1}da/dz$. Считая, что амплитуда Z_1 больше всех возникающих поправок, получаем из (11), (12) неоднородные уравнения для волн высших типов (для Z_p при $p \neq 1$) и для поправки δZ_1 к амплитуде основной волны. Эти уравнения здесь не приводятся в связи с ограниченностью объема статьи и будут даны в отдельной работе. Укажем, что решение их находится с помощью функции Грина, причем в качестве фундаментальной пары решений при нахождении δZ_1 выбираются функции (18), (19), а при нахождении амплитуд Z_p – функции

$$W_p(z) = \exp(z Y_p / l) \times {}_2F_1[X_p + Y_p, -X_p + Y_p; 2Y_p + 1; -\exp(z/l)], \quad (27)$$

$$V_p(z) = \exp(-z Y_p / l) \times {}_2F_1[X_p + Y_p, X_p - Y_p; 2X_p + 1; -\exp(-z/l)], \quad (28)$$

где

$$X_p = \frac{\omega \sqrt{\epsilon} l}{c} \left(\frac{\xi_p^2}{\xi_1^2} M - 1 \right)^{1/2};$$

$$Y_p = \frac{\omega \sqrt{\epsilon} l}{c} \left(\frac{\xi_p^2}{\xi_1^2} m - 1 \right)^{1/2}. \quad (29)$$

На решения Z_p налагались очевидные условия ограниченности на $\pm\infty$. На поправку δZ_1 , кроме того, налагалось требование, чтобы приходящая из $-\infty$ волна, содержащаяся в сумме $Z_1 + \delta Z_1$, имела модуль амплитуды, совпадающий с модулем амплитуды волны того же направления, содержащейся в Z_1 (т. е. неизменность входного потока). Окончательные формулы мы приведем для величин $A_p = Z_p(a_{-\infty}/a(z))$ и $\delta A_1 = \delta Z_1(a_{-\infty}/a(z))$.

5.1. Поправка к амплитуде основной волны

Для поправки δA_1 к амплитуде основной волны ($p = 1$) результат расчета выглядит следующим образом:

$$\delta A_1 = \left\{ CV(z) - \frac{l}{2X} \left[U(z) \int_z^\infty V(\zeta) f_1(\zeta) d\zeta + V(z) \int_{-\infty}^z U(\zeta) f_1(\zeta) d\zeta \right] \right\} \left(\frac{a_{-\infty}}{a(z)} \right), \quad (30)$$

где

$$f_1(z) = - \left(\frac{1}{a} \frac{da}{dz} \right)^2 \sigma_1 Z_1; \quad \sigma_1 = \sum_{k \neq 1}^{\infty} \tilde{c}_{1k} \tilde{c}_{k1}; \quad (31)$$

$$C = \left[1 + B^2 \left(\frac{\beta \gamma^* - \gamma \beta^*}{2\beta \beta^*} \right)^2 \right]^{1/2} - B \frac{\beta \gamma^* + \gamma \beta^*}{2\beta \beta^*} - 1; \quad (32)$$

$$B = - \frac{l}{2X} \int_{-\infty}^{\infty} V(\zeta) f_1(\zeta) d\zeta; \quad (33)$$

$$\beta = \frac{\Gamma(2X + 1) \Gamma(-i2Y)}{\Gamma(X - iY) \Gamma(X + 1 - iY)}; \quad (34)$$

$$\gamma = \frac{\Gamma(-2X + 1) \Gamma(-i2Y)}{\Gamma(-X - iY) \Gamma(-X + 1 - iY)}.$$

Вычисление поправок к решению, полученному в адиабатическом приближении, было проведено по формулам (30)–(34) для волновода с радиусами $a_{in} = 250$ нм, $a_{\infty} = 25$ нм и широкого набора параметров l (250, 100, 75, 50 и 25 нм). Как и следовало ожидать, поправки оказывались особенно малыми при больших параметрах l . На рис.2 для $l = 50$ нм штриховой кривой показаны результаты расчета амплитуды $A_1 + \delta A_1$ основной волны магнитного типа ($\lambda = 500$ нм, $\xi_1^m = 3.832$), полученной на основе формулы (30), и их сравнение с амплитудой A_1 (сплошная кривая), рассчитанной в адиабатическом приближении (см. (14), (19), (26)). Видно, что при учете возмущения δA_1 картина стоячих волн немного смещена влево от начала координат, а затухание амплитуды поля начинается при несколько меньших значениях z по сравнению с теми, которые имеют место при расчете в нулевом приближении. Из результатов расчетов, представленных на рис.2, следует, что при $l \gtrsim 50$ нм учет поправки δA_1 (30) приводит к небольшим изменениям амплитуды волны A_1 , полученной в нулевом приближении. Таким образом, в рассматриваемых случаях нулевое

(адиабатическое) приближение вполне разумно количественно описывает структуру основной волны.

Наша задача состояла в установлении тех значений l , при которых поправка к адиабатическому приближению становилась заметной. В связи с этим на рис.3 приведены результаты аналогичного расчета амплитуды $A_1 + \delta A_1$ (штриховая кривая) для $l = 25$ нм. В этом случае сравнение с результатом расчета амплитуды поля в нулевом приближении (сплошная кривая) демонстрирует большее смещение картины стоячих волн в сторону отрицательных z . Кроме того, при $l = 25$ нм возникает дополнительный экстремум в той области продольной координаты волновода, где поле уже начинает заметно ослабевать. Аналогичный экстремум имеется и в первом варианте расчета ($l = 50$ нм), однако он настолько мал, что на рис.2 практически незаметен. Значение $l = 25$ нм определяет границу, начиная с которой адиабатическое приближение дает лишь грубую оценку для поля в гипергеометрическом волноводе. Однако следует отметить, что проведенные нами недавно расчеты для конуса (точно решаемая задача) и их сопоставление с представленными выше результатами показывает, что адиабатическое приближение дает разумную количественную оценку даже при $l = 25$ нм.

5.2. Трансформация поля в волны высших типов

Перейдем теперь к обсуждению трансформации поля в волны высших типов. Выражение для амплитуд A_p ($p \geq 2$) выглядит следующим образом:

$$A_p = \left(\frac{a_{-\infty}}{a(z)} \right) \left\{ -\frac{l}{2X_p} \frac{\Gamma(X_p + Y_p)\Gamma(X_p + Y_p + 1)}{\Gamma(2Y_p + 1)\Gamma(2X_p)} \times \left[W_p \int_z^\infty V_p(\zeta)f_p(\zeta)d\zeta + V_p \int_{-\infty}^z W_p(\zeta)f_p(\zeta)d\zeta \right] \right\}, \quad (35)$$

где

$$f_p(z) = -2 \left(\frac{1}{a} \frac{da}{dz} \right) \tilde{c}_{1p} \frac{dZ_1}{dz} - \frac{d}{dz} \left(\frac{1}{a} \frac{da}{dz} \right) \tilde{c}_{1p} Z_1 - \left(\frac{1}{a} \frac{da}{dz} \right)^2 \sigma_p Z_1; \quad (36)$$

$$\sigma_p = \sum_{k \neq 1, p}^{\infty} \tilde{c}_{1k} \tilde{c}_{kp}. \quad (37)$$

По формуле (35) были рассчитаны зависимости от про-

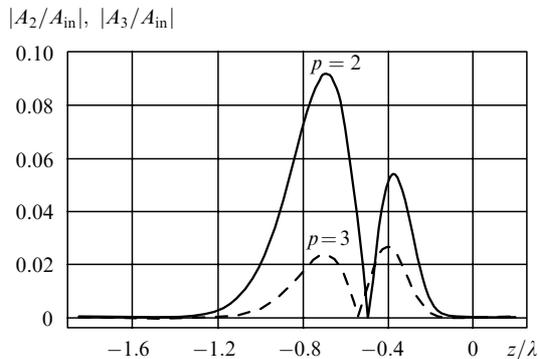


Рис.5. Зависимости нормированных амплитуд A_2 ($p=2, \xi_2^m = 7.016$) и A_3 ($p=3, \xi_3^m = 10.170$) от продольной координаты z/λ . Сплошная и штриховая кривые рассчитаны по формулам (10) и (35) соответственно для $a_{in} = \lambda/2, a_{\infty} = \lambda/20$ и $l = \lambda/10$.

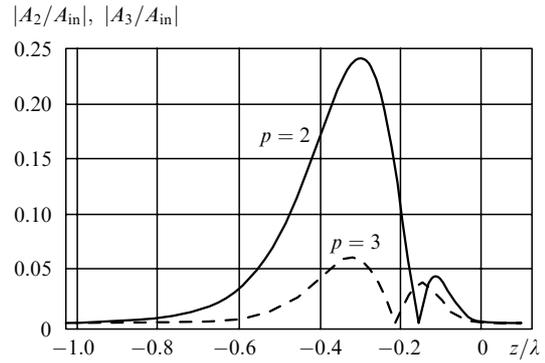


Рис.6. То же, что и на рис.5, но для параметра волновода $l = \lambda/20$.

дольной координаты z амплитуд второй волны $A_2(z)$ ($p = 2, \xi_2^m = 7.016$) и третьей волны $A_3(z)$ ($p = 3, \xi_3^m = 10.170$) магнитного типа. Соответствующие зависимости при $l = 50$ и 25 нм ($\lambda = 500$ нм) приведены на рис.5 и 6. В обоих случаях амплитуды A_2 и A_3 нормированы на модуль входной амплитуды основной волны ($p = 1, \xi_1^m = 3.832$), т. е. на величину $|A_{in}|$.

Как видно из рис.5 и 6, амплитуды второй и третьей волн оказываются существенно меньше амплитуды первой. Они заметно отличны от нуля только на малом участке оси z (там, где происходит существенное сужение волновода) и практически равны нулю вне этого участка. Построенные зависимости позволяют сделать вывод о том, что при $l \geq 50$ нм невозмущенное решение (волна A_1) является хорошим приближением к точному решению задачи.

6. Коэффициент пропускания волновода

Выше были рассмотрены свойства электромагнитных волн в бесконечном волноводе. Для волновода конечной длины оптическая схема выглядит следующим образом. Сначала излучение распространяется в вакууме, затем попадает на слой металла, содержащий тонкий канал (волновод), проходит по каналу конечной длины до выходного сечения и далее выходит в открытое пространство, где преобразуется в соответствии с волновыми законами. Результат этого преобразования зависит от того, рассматривается ли поле в ближней или в дальней зоне.

В данной работе основной интерес для нас представляет изучение поля в ближней зоне. Поэтому при обсуждении коэффициента пропускания волновода мы будем иметь дело с плотностями энергии поля, а не с потоком энергии. Это связано с тем, что в приборах ближнего поля определяющую роль играют нераспространяющиеся волны, с которыми не связаны потоки энергии. Введем ряд величин, которые представляют собой плотности энергии электрического поля, проинтегрированные по поперечному сечению в соответствующих участках оптической схемы (для определенности ниже рассматриваются волны магнитного типа). Для исходного излучения (при $z = -\infty$) введем величину

$$P_{-\infty} = 2\pi \int_0^{a_{beam}} |E_\phi^{(-\infty)}|^2 \rho d\rho, \quad (38)$$

где a_{beam} – радиус входного пучка; $E_\phi^{(-\infty)}$ – напряженность

поля только одной, идущей в сторону волновода волны (отраженные волны в (38) не включаются в отличие от последующих формул). Далее введем интегральную величину

$$P(z_0) = 2\pi \int_0^{a(z_0)} |E_\varphi(\rho, z_0)|^2 \rho d\rho, \quad (39)$$

которая относится ко входному сечению $z = z_0$ ($z_0 < 0$) волновода радиусом $a(z_0)$. Соответственно

$$P(z = 0) = 2\pi \int_0^{a_{\text{out}}} |E_\varphi(\rho, z = 0)|^2 \rho d\rho \quad (40)$$

– это интеграл от плотности энергии в выходном сечении $z = 0$ волновода радиусом $a_{\text{out}} \equiv a(0)$. В формуле (40) используется поле E_φ , рассчитанное в волноводе без учета влияния границы со свободным пространством ($z = 0$). В формуле

$$P_{\text{out}} = 2\pi \int_0^{a_{\text{out}}} |E_\varphi^{\text{out}}(\rho, z = 0)|^2 \rho d\rho \quad (41)$$

величина $E_\varphi^{\text{out}}(\rho, z = 0)$ – это поле $E_\varphi(\rho, z = 0)$, измененное в результате отражения от выходного торца волновода ($z = 0$).

Если дополнительно интересоваться полями в дальней зоне, то следует ввести еще одну интегральную величину

$$P_\infty = \int_S |E_\varphi(\rho, z)|^2 dS, \quad (42)$$

где интеграл берется по поверхности S полусферы, радиус которой стремится к бесконечности.

Трансформация поля на каждом участке оптической схемы характеризуется соответствующим коэффициентом передачи:

$$T_{-\infty w} = \frac{P(z_0)}{P_{-\infty}}, \quad T_w = \frac{P(0)}{P(z_0)}, \quad (43)$$

$$T_{w0} = \frac{P_{\text{out}}}{P(0)}, \quad T_{0\infty} = \frac{P_\infty}{P_{\text{out}}}.$$

Здесь $T_{-\infty w}$ – коэффициент передачи из свободного пространства в волновод; T_w – коэффициент пропускания самого волновода; T_{w0} – коэффициент трансформации, обусловленный влиянием выходного торца волновода; $T_{0\infty}$ – коэффициент передачи свободного пространства из ближней зоны в дальнюю.

Интегральная плотность энергии P_{out} (41) в ближней зоне на выходе волновода выражается через интегральную плотность входного излучения $P_{-\infty}$ и полный коэффициент передачи T , учитывающий три этапа преобразования:

$$P_{\text{out}} = P_{-\infty} T, \quad T = T_{-\infty w} T_w T_{w0}. \quad (44)$$

Для нахождения же интегральной плотности энергии в дальней зоне P_∞ (42) в выражение для результирующего коэффициента пропускания $T_{-\infty\infty}$ нужно включить еще один сомножитель $T_{0\infty}$, отвечающий трансформации поля в свободном пространстве:

$$T_{-\infty\infty} = T_{-\infty w} T_w T_{w0} T_{0\infty}, \quad P_\infty = P_{\text{out}} T_{0\infty}. \quad (45)$$

Поясним смысл трансформации поля в свободном пространстве. Пусть в некоторой плоскости распределение поля имеет вид пятна радиусом a (где $a \ll \lambda$), а при $\rho > a$ поле обращается в нуль. Фурье-спектр такого пятна занимает область поперечных пространственных частот $k_\rho \sim 1/a$. Только малая доля этих гармоник отвечает распространяющимся волнам. Волны с $k_\rho < \omega/c = 2\pi/\lambda$ распространяются, остальные же затухают при удалении от рассматриваемой плоскости. В поток энергии и в интенсивность поля в дальней зоне вносят вклад только незатухающие волны, доля которых составляет порядка $(a/\lambda)^2$ от всего пространственного спектра. В интенсивность же поля в ближней зоне вносят вклад все пространственные гармоники, и для нее роль затухающих волн оказывается определяющей. Именно в силу этого обстоятельства интегральная плотность, а не поток энергии характеризует интенсивность поля в ближней зоне.

Обратим теперь внимание на ситуацию, когда поле находится внутри волновода с субволновым поперечным сечением. Коэффициент пропускания T_w связывает интегральные плотности энергии во входном и выходном сечениях волновода. Этот коэффициент и представляет наибольший интерес для оптики ближнего поля, являясь основным предметом исследований в данной работе. Надо иметь в виду, что в рассматриваемой ситуации коэффициент $T_{w0} \approx 1$. При этом (т. е. в отсутствие отражения от выходного торца) поток энергии в волноводе равен нулю.

Если же учитывать отражение от границы волновода со свободным пространством, то интенсивность поля на выходном торце будет пропорциональна $|1 + R|^2$, где R – амплитудный коэффициент отражения. Поток электромагнитной энергии будет при этом определяться величиной $\text{Im}R$. Если коэффициент R мал, то он слабо влияет на интенсивность выходного поля. Вместе с тем этот коэффициент, а именно его мнимая часть, полностью определяет поток энергии. Эти соображения также указывают на необходимость анализа ближнего поля в терминах плотности энергии, а не потока энергии.

Комплекс вопросов, связанных с преобразованием поля на его входном и выходном сечениях, будет подробно обсуждаться в Приложениях 1, 2.

В дальнейшем наши конкретные расчеты будут направлены на изучение поведения коэффициента пропускания T_w самого волновода. Рассмотрим полубесконечный ($-\infty < z \leq 0$) сужающийся волновод гипергеометрического типа с параметрами $z_0 \rightarrow -\infty$, $a_{\text{in}} \equiv a(z_0) \sim \lambda$ и $a_{\text{out}} \equiv a(0) \ll \lambda$ и найдем спектральную зависимость коэффициента пропускания T_w . При этом возьмем фиксированные значения a_{in} , a_{out} и l , а изменять будем длину волны входного излучения. Амплитуда $A_1(0)$ будет давать нам амплитуду поля на выходе $z = 0$, т. е. $A_{\text{out}} = A_1(0)$.

В этом разделе наряду с волнами магнитного типа будут обсуждаться также волны электрического типа. Разумеется, строгое исследование волн электрического типа должно учитывать как зависимость самой амплитуды от изменяющегося радиуса, так и зависимость от него граничного условия (которое в общем случае связывает между собой радиальную и продольную компоненты электрического поля). Однако в случае достаточно медленных изменений профиля можно воспользоваться собственными функциями регулярного волновода и внести в них параметрическую зависимость от z . В рассмат-

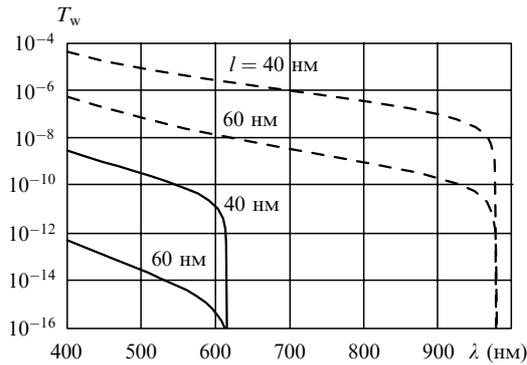


Рис.7. Коэффициент пропускания волновода T_w в зависимости от длины волны λ для $a_{in} = 250$ нм, $a_{out} = 35$ нм и различных l . Сплошные и штриховые кривые – результаты расчетов соответственно для основных волн ($p = 1$) магнитного ($\xi_1^m = 3.832$) и электрического ($\xi_1^e = 2.405$) типов при $\lambda < \lambda_c$ (длины волн отсечки $\lambda_c^m = 614.9$ нм и $\lambda_c^e = 979.8$ нм).

риваемых примерах был использован именно такой прием, что свелось к замене собственного значения волнового числа $\xi_1^m = 3.832$ (используемого в волнах магнитного типа) на число $\xi_1^e = 2.405$, являющееся корнем уравнения $J_0(\xi) = 0$.

На рис.7 представлены полученные в настоящей работе результаты расчетов для коэффициентов пропускания T_w волновода в зависимости от длины световой волны λ . Эти коэффициенты связаны с отношением амплитуды поля на выходе волновода к модулю входной амплитуды $A_{out}/|A_{in}|$ с помощью следующего соотношения: $T_w = [a_{out}A_{out}/(a_{in}|A_{in}|)]^2$. Все расчеты выполнены здесь для основных волн ($p = 1$) магнитного и электрического типов в нулевом приближении по скорости изменения профиля волновода при входном и выходном радиусах $a_{in} = 250$ нм, $a_{out} = 35$ нм и параметре $l = 40$ и 60 нм.

Во всех указанных случаях имеет место качественно аналогичное поведение коэффициента пропускания волновода T_w в зависимости от длины волны. С увеличением λ происходит сильное уменьшение коэффициента T_w , которое становится особенно резким при приближении λ к длине волны отсечки $\lambda_c = 2\pi\sqrt{\epsilon}a_{in}/\xi_1$ (выше которой волны перестают быть распространяющимися даже на входном конце волновода). При выбранных здесь радиусах a_{in} и диэлектрической проницаемости волокна ϵ ($\sqrt{\epsilon} = 1.5$) длины волн отсечки λ_c^m и λ_c^e составляют соответственно 614.92 и 979.78 нм для основных волн магнитного и электрического типов. Сравнение значений T_w для основной волны магнитного типа (при $l = 40$ нм) при различных длинах волн показывает, что T_w уменьшается приблизительно в 220 раз при увеличении λ от 400 до 600 нм. Аналогичное сравнение для основной волны электрического типа показывает, что коэффициент пропускания T_w уменьшается в 400 раз при увеличении длины волны от 400 до 900 нм.

В то же время обращают на себя внимание большие количественные различия в коэффициентах T_w для волн разных типов. Так, например, при $\lambda = 500$ нм и $l = 40$ нм коэффициент пропускания T_w для волны электрического типа превышает соответствующую величину для волны магнитного типа приблизительно в 2.6×10^4 раз. Результаты, представленные на рис.7, наглядно демонстрируют также сильную зависимость выходного поля от параметра l , характеризующего крутизну профиля волновода.

7. Заключение

Развитая техника исследования волн позволила установить пространственную структуру поля в сужающихся волноводах. Определена область геометрических параметров, близких к экспериментальным, которые реализуются при использовании оптического волокна в качестве источника освещения в микроскопе ближнего поля.

При выбранных здесь условиях волна, на которой осуществляется возбуждение волновода, остается преобладающей по всей его длине. Происходит лишь слабая перекачка энергии основной волны в волны высших типов. Расчеты указывают на сильную зависимость характеристик поля от длины переходного участка, на котором осуществляется сужение волновода. Меньшие длины переходной области (т. е. большая крутизна профиля волновода) обеспечивают большее пропускание за счет уменьшения вклада переходного участка в затухание. Увеличение крутизны профиля волновода приводит и к другому эффекту: увеличивается перекачка энергии в волны высших типов, что в конце концов ведет к нарушению применимости используемой техники расчета.

Для конкретного рассмотрения были выбраны такие варианты ($l \geq \lambda/10$), для которых трансформация в высшие волны оказывается незначительной. Качественно ясно, что для обеспечения наилучшего пропускания целесообразно применять волноводы с более быстрым изменением радиуса. В данной работе мы получили, что при $l \approx \lambda/10$ пропускание волновода T_w достигает значений $\sim 10^{-5} - 10^{-4}$ для волн электрического типа. Если перейти к $l \approx \lambda/20$, то оценка на основе адиабатического приближения показывает, что пропускание волновода возрастает на порядок. Справедливость этой оценки подтверждают проведенные нами недавно расчеты для конуса, допускающего точное решение. Результаты расчета, полученные для конуса, неплохо согласуются с результатами расчета в адиабатическом приближении для участка гипергеометрического волновода с соответствующим наклоном стенок.

Проведенные в работе расчеты показывают, что весьма резкой оказывается зависимость коэффициента пропускания сужающегося волновода от длины волны. Безусловно, предпочтительны короткие волны, что на качественном уровне вполне согласуется со свойствами регулярных волноводов. То же можно сказать и о зависимости коэффициента пропускания от поперечной структуры возбуждающей волны: очевидны преимущества волн с малыми поперечными волновыми числами. Поэтому большой интерес представляют волны электрического и магнитного типов с ненулевыми азимутальными числами. Представляется принципиально важным распространить развитый здесь подход на больший круг задач с учетом специфики указанных волн.

Работа выполнена при финансовой поддержке РФФИ (грант № 00-02-17245) и программы МПНТ «Оптика. Лазерная физика».

Приложение 1. Отражение волн на входном сечении волновода

Рассмотрим падение волны из свободного пространства на круглый цилиндрический волновод радиусом a . Будем считать, что во входном сечении $z = z_0$ ($z_0 < 0$) волновод переходит в идеально проводящую бесконеч-

ную плоскость, простирающуюся от $\rho = a$ до $\rho = \infty$. Эта картина отражает тот факт, что стенки волновода не являются тонкими, они заведомо толще, чем его внутренний диаметр.

От поперечной структуры падающей волны радиальным образом зависит соласование полей в свободном пространстве и в волноводе. Для упрощения расчетов выберем вид поля, близкий к волноводной моде, а именно бesselев пучок

$$E_\varphi = AJ_1(\alpha\rho) \left\{ \exp \left[i(z - z_0) \left(\frac{\omega^2}{c^2} - \alpha^2 \right)^{1/2} \right] - \exp \left[-i(z - z_0) \left(\frac{\omega^2}{c^2} - \alpha^2 \right)^{1/2} \right] \right\}, \quad (\text{П1.1})$$

где A – амплитуда падающей волны. В это выражение включены и падающая волна и отраженная, вследствие чего будет выполнено условие обращения поля в нуль на фланце, находящемся в плоскости $z = z_0$. Поперечная компонента волнового вектора α – заданная характеристика исходной волны. Будем считать, что входной бesselев пучок обрезан при $\rho = a_{\text{beam}}$ (т. е. a_{beam} – его радиус), и положим для удобства $J_1(\alpha a_{\text{beam}}) = 0$. Таким образом, $\alpha a_{\text{beam}} = \xi_n$, где ξ_n – один из нулей функции Бесселя с $n \neq 1$. Поскольку входной радиус волновода $a \equiv a(z_0) \ll a_{\text{beam}}$, то выполняется соотношение $\alpha a \ll 1$.

К полному виду (П1.1) будет добавлен непрерывный набор отраженных волн с другими волновыми векторами. Однако удобно сначала написать выражение для поля в волноводе:

$$E_\varphi^w = \sum_{s=1}^{\infty} B_s \Phi_s(\rho) \exp \left[i(z - z_0) \left(\frac{\omega^2}{c^2} \varepsilon - q_s^2 \right)^{1/2} \right], \quad (\text{П1.2})$$

где B_s – постоянные коэффициенты;

$$\Phi_s(\rho) \equiv J_1(q_s \rho); \quad J_1(q_s a) = 0. \quad (\text{П1.3})$$

Фазу для полей в (П1.2) и в аналогичных выражениях будем выбирать следующим образом:

$$\text{arg} \left(\frac{\omega^2}{c^2} \varepsilon - q_s^2 \right)^{1/2} = \begin{cases} 0, & q_s^2 < \frac{\omega^2}{c^2} \varepsilon, \\ \frac{\pi}{2}, & q_s^2 > \frac{\omega^2}{c^2} \varepsilon. \end{cases} \quad (\text{П1.4})$$

Учитывая соотношение (П1.2) и основываясь на условии непрерывности поля при $z = z_0$, $\rho < a$, выражение для полного поля в свободном пространстве можно записать так:

$$E_\varphi^{(0)} = AJ_1(\alpha\rho) \left\{ \exp \left[i(z - z_0) \left(\frac{\omega^2}{c^2} - \alpha^2 \right)^{1/2} \right] - \exp \left[-i(z - z_0) \left(\frac{\omega^2}{c^2} - \alpha^2 \right)^{1/2} \right] \right\} + \sum_{s=1}^{\infty} B_s g_s(\rho, z), \quad (\text{П1.5})$$

$$g_s(\rho, z) = \int_0^\infty \tilde{g}_s(\chi) J_1(\chi\rho) \times \exp \left[-i(z - z_0) \left(\frac{\omega^2}{c^2} - \chi^2 \right)^{1/2} \right] \chi d\chi, \quad (\text{П1.6})$$

где знак корня выбирается в соответствии с (П1.4); χ – поперечное волновое число в свободном пространстве;

$$\tilde{g}_s(\chi) = \int_0^a \Phi_s(\rho') J_1(\chi\rho') \rho' d\rho'. \quad (\text{П1.7})$$

С учетом (П1.3) функция $\tilde{g}_s(\chi)$ вычисляется в явном виде:

$$\tilde{g}_s(\chi) = \frac{q_s a}{\chi^2 - q_s^2} J_1(\chi a) J_0(q_s a). \quad (\text{П1.8})$$

При $\chi = q_s$ это выражение выглядит следующим образом:

$$\tilde{g}_s(q_s) = \frac{a^2}{2} J_0^2(q_s a). \quad (\text{П1.9})$$

Радиальная компонента магнитного поля в свободном пространстве равна $i(c/\omega) \partial E_\varphi^{(0)}/\partial z$, а в волноводе – $i(c/\omega) \partial E_\varphi^w/\partial z$. Для обеспечения непрерывности магнитного поля на отверстии нужно потребовать выполнения при $z = z_0$, $\rho < a$ равенства

$$\left. \frac{\partial E_\varphi^{(0)}}{\partial z} \right|_{z=z_0} = \left. \frac{\partial E_\varphi^w}{\partial z} \right|_{z=z_0}. \quad (\text{П1.10})$$

С учетом (П1.2), (П1.5) равенство (П1.10) сводится к следующему выражению:

$$A2i \left(\frac{\omega^2}{c^2} - \alpha^2 \right)^{1/2} J_1(\alpha\rho) + \sum_{s=1}^{\infty} B_s \int_0^\infty \left[-i \left(\frac{\omega^2}{c^2} - \chi^2 \right)^{1/2} \right] \tilde{g}_s(\chi) J_1(\chi\rho) \chi d\chi = \sum_{s=1}^{\infty} B_s i \left(\frac{\omega^2}{c^2} \varepsilon - q_s^2 \right)^{1/2} \Phi_s(\rho). \quad (\text{П1.11})$$

Умножая правую и левую части (П1.11) на $\Phi_n(\rho)$ и интегрируя по $\rho d\rho$, получаем

$$2i \left(\frac{\omega^2}{c^2} - \alpha^2 \right)^{1/2} A \int_0^a J_1(\alpha\rho) \Phi_n(\rho) \rho d\rho + \sum_{s=1}^{\infty} B_s \int_0^a \Phi_n(\rho) \rho d\rho \int_0^\infty \left[-i \left(\frac{\omega^2}{c^2} - \chi^2 \right)^{1/2} \right] \times \tilde{g}_s(\chi) J_1(\chi\rho) \chi d\chi = i \left(\frac{\omega^2}{c^2} \varepsilon - q_s^2 \right)^{1/2} B_n \int_0^a \Phi_n^2(\rho) \rho d\rho. \quad (\text{П1.12})$$

Учитывая выражение (П1.8) и переходя к безразмерным величинам $x = \chi a$ и $\xi_n = q_n a$, приводим (П1.12) к виду

$$2ia \left(\frac{\omega^2}{c^2} - \alpha^2 \right)^{1/2} AG(\xi_n, \alpha a) + \sum_{s=1}^{\infty} B_s b_{sn} = i \left(\frac{\omega^2 a^2}{c^2} \varepsilon - \xi_n^2 \right)^{1/2} B_n, \quad (\text{П1.13})$$

где

$$G(\xi_n, \alpha a) = \frac{1}{N_n} \int_0^a J_1(\alpha\rho) J_1(q_n \rho) \rho d\rho; \quad (\text{П1.14})$$

$$N_n = \int_0^a J_1^2(q_n \rho) \rho d\rho;$$

$$b_{sn} = \frac{J_0^2(\xi_n)}{2} \int_0^\infty \left[-i \left(\frac{\omega^2 a^2}{c^2} - x^2 \right)^{1/2} \right] \times G(\xi_n, x) G(\xi_s, x) x dx; \quad (\text{П1.15})$$

$$G(\xi_n, x) = \frac{2\xi_n}{x^2 - \xi_n^2} \frac{J_1(x)}{J_0(\xi_n)}. \quad (\text{П1.16})$$

Напомним, что $J_1(\xi_n) = 0$, так что при $x = \xi_n$ функция $G(\xi_n, x)$ обращается в единицу.

Система уравнений (П1.13) позволяет получить амплитуды B_n всех волн, входящих в волновод. Мы здесь ограничимся приближенным решением для амплитуды B_1 низшей моды. Основанием для использования этого приближения является тот факт, что модули недиагональных элементов b_{sn} оказываются много меньше диагональных. Выбирая в (П1.13) $n = 1$ и полагая $b_{s1} = 0$ при $s \neq 1$, находим амплитудный коэффициент пропускания при переходе из свободного пространства в волновод

$$D_{-\infty w} \equiv \frac{B_1}{A} = 2i \left(\frac{\omega^2 a^2}{c^2} - \alpha^2 a^2 \right)^{1/2} \times G(\xi_1, \alpha a) \left[b_{11} + i \left(\frac{\omega^2 a^2}{c^2} \varepsilon - \xi_1^2 \right)^{1/2} \right]^{-1}. \quad (\text{П1.17})$$

Вычисление коэффициента b_{11} , входящего в (П1.17), позволяет получить численные оценки амплитудного коэффициента пропускания.

С учетом соотношений $\alpha a \ll 1$ и $\alpha a_{\text{beam}} = \xi_n$ выражение (П1.17) приводится к виду

$$D_{-\infty w} = \frac{-2i\xi_n}{\xi_1 J_0(\xi_1)} \left(\frac{\omega a}{c} \right) \left(\frac{a}{a_{\text{beam}}} \right) f \left(\frac{\omega a}{c} \right), \quad (\text{П1.18})$$

$$f \left(\frac{\omega a}{c} \right) = \left[b_{11} + i \left(\frac{\omega^2 a^2}{c^2} \varepsilon - \xi_1^2 \right)^{1/2} \right]^{-1}.$$

Энергетический коэффициент передачи $T_{-\infty w}$ из свободного пространства в волновод (см. (43)) связан с коэффициентом $D_{-\infty w}$ (П1.18) формулой

$$T_{-\infty w} = \left[\int_0^a J_1^2(q_1 \rho) \rho d\rho / \int_0^{a_{\text{beam}}} J_1^2(x \rho) \rho d\rho \right] \left| \frac{B_1}{A} \right|^2 = \frac{J_0^2(\xi_1)}{J_0^2(\xi_n)} \left| \frac{B_1}{A} \right|^2 \frac{a^2}{a_{\text{beam}}^2}. \quad (\text{П1.19})$$

Подставляя (П1.18) в (П1.19), получаем

$$T_{-\infty w} = \eta_n \left(\frac{\omega}{c} \right)^2 \left(\frac{a^6}{a_{\text{beam}}^4} \right) \mathcal{F} \left(\frac{\omega a}{c} \right), \quad (\text{П1.20})$$

где $\eta_n = \{2\xi_n / [\xi_1 J_0(\xi_n)]\}^2$ – постоянные коэффициенты; $\mathcal{F}(\omega a/c) = |f(\omega a/c)|^2$ – безразмерная функция. Обратим внимание на то, что в формулу (П1.20) входит радиус $a(z_0)$, который практически совпадает с радиусом $a_{\text{in}} \equiv a(-\infty)$, если входное сечение $z = z_0 < 0$ гипергеометрического волновода выбирается на его достаточно пологом («цилиндрическом») участке. Таким образом, формулы (П1.15)–(П1.20) определяют поведение и величину коэффициента передачи $T_{-\infty w}$ из свободного пространства в волновод при различных параметрах a/a_{beam} , ε и $\omega a/c$.

Приложение 2. Отражение волн на выходном сечении волновода

Рассмотрим теперь, как изменяется затухающая волна в цилиндрическом волноводе из-за наличия границы со свободным пространством. Возьмем волноводную моду вида

$$E_\varphi = A_1 \Phi_1(\rho) \exp \left[-z \left(q_1^2 - \frac{\omega^2}{c^2} \varepsilon \right)^{1/2} \right], \quad (\text{П2.1})$$

которая соответствует бесконечному волноводу (A_1 – амплитуда низшей моды). Будем считать, что в сечении $z = 0$ волновод обрывается, переходя в идеально проводящую плоскость, простирающуюся от $\rho = a$ (имеется в виду $a = a_{\text{out}}$) до бесконечности (аналогично тому, как это предполагалось для входного торца). Поле, измененное из-за влияния границы, запишем в виде

$$E_\varphi^w = A_1 \Phi_1(\rho) \exp \left[-z \left(q_1^2 - \frac{\omega^2}{c^2} \varepsilon \right)^{1/2} \right] + \sum_{s=1}^{\infty} B_s \Phi_s(\rho) \exp \left[z \left(q_s^2 - \frac{\omega^2}{c^2} \varepsilon \right)^{1/2} \right]. \quad (\text{П2.2})$$

Поле в свободном пространстве представим в виде интеграла Фурье – Бесселя. Учтем, что при $z = 0$, $\rho < a$ оно должно совпадать с полем в волноводе, а при $z = 0$, $\rho > a$ – обращаться в нуль. Это поле (с учетом уравнения (1)) должно иметь следующий вид:

$$E_\varphi^{(0)} = A_1 g_1(\rho, z) + \sum_{s=1}^{\infty} B_s g_s(\rho, z), \quad (\text{П2.3})$$

$$g_s(\rho, z) = \int_0^\infty \tilde{g}_s(\chi) J_1(\chi \rho) \times \exp \left[iz \left(\frac{\omega^2}{c^2} - \chi^2 \right)^{1/2} \right] \chi d\chi. \quad (\text{П2.4})$$

Отметим, что для задач об отражении от входного и выходного сечений волновода функции g_s различаются только своей зависимостью от z , а зависимость от ρ у них одинаковая (см. (П1.6), (П2.4)). Поэтому функции $g_s(\chi)$ (см. (П1.7)) оказываются одинаковыми в обоих случаях. Для обеспечения непрерывности магнитного поля нужно потребовать выполнения при $\rho < a$ равенства

$$\frac{\partial E_\varphi^{(0)}}{\partial z} \Big|_{z=0} = \frac{\partial E_\varphi^w}{\partial z} \Big|_{z=0}. \quad (\text{П2.5})$$

С учетом (П2.1)–(П2.3) это равенство принимает вид

$$\sum_{s=1}^{\infty} \left(q_s^2 - \frac{\omega^2}{c^2} \varepsilon \right)^{1/2} (B_s - \delta_{s1} A_1) \Phi_s(\rho) = \sum_{s=1}^{\infty} (B_s + \delta_{s1} A_1) \times \int_0^\infty \left[i \left(\frac{\omega^2}{c^2} - \chi^2 \right)^{1/2} \right] \tilde{g}_s(\chi) J_1(\chi \rho) \chi d\chi, \quad (\text{П2.6})$$

где δ_{s1} – символ Кронекера. Умножая обе части (П2.6) на $\Phi_n(\rho)$ и интегрируя по $\rho d\rho$ в пределах $0 \leq \rho \leq a$, находим

$$(B_n - \delta_{n1} A_1) \left(q_n^2 - \frac{\omega^2}{c^2} \varepsilon \right)^{1/2} \int_0^a \Phi_n^2(\rho) \rho d\rho = \sum_{s=1}^{\infty} (B_s + \delta_{s1} A_1) \\ = \int_0^{\infty} i \left(\frac{\omega^2}{c^2} - x^2 \right)^{1/2} \tilde{g}_s(x) \tilde{g}_n(x) x dx. \quad (\text{П2.7})$$

Учитывая явное выражение (П1.8) для $\tilde{g}_s(x)$ и переходя к безразмерным переменным $x = \kappa a$ и $\xi_n = q_n a$, приходим (П2.7) к виду

$$(B_n - \delta_{n1} A_1) \left(\xi_n^2 - \frac{\omega^2 a^2}{c^2} \varepsilon \right)^{1/2} \\ = - \sum_{s=1}^{\infty} (B_s + \delta_{s1} A_1) b_{ns}, \quad (\text{П2.8})$$

где b_{sn} дается формулами (П1.15) и (П1.16). Система уравнений (П2.8) позволяет найти амплитуды B_n всех волн, отраженных назад, в волновод. Для простоты ограничимся приближенным решением для B_1 . Как и в Приложении 1, при построении решения используется тот факт, что недиагональные элементы b_{sn} существенно меньше диагональных. Выбирая в (П2.8) $n = 1$ и полагая $b_{s1} = 0$ для $s \neq 1$, находим

$$(B_1 - A_1) \left(\xi_1^2 - \frac{\omega^2 a^2}{c^2} \varepsilon \right)^{1/2} = -(B_1 + A_1) b_{11}, \quad (\text{П2.9})$$

откуда получаем

$$R \equiv \frac{B_1}{A_1} = \frac{1 - \mu}{1 + \mu}, \quad \mu = b_{11} \left(\xi_1^2 - \frac{\omega^2 a^2}{c^2} \varepsilon \right)^{-1/2}. \quad (\text{П2.10})$$

Величина R представляет собой амплитудный коэффициент отражения от открытого конца волновода радиусом $a \equiv a_{\text{out}}$. Используем далее тот факт, что при $\omega a/c \ll 1$ имеют место следующие соотношения:

$$\left| \text{Re } b_{11} \left(\xi_1^2 - \frac{\omega^2 a^2}{c^2} \varepsilon \right)^{-1/2} - 1 \right| \ll 1, \quad (\text{П2.11})$$

$$\text{Im } b_{11} \approx - \frac{1}{15 \xi_1^2} \frac{\omega^5 a^5}{c^5}.$$

С учетом (П2.11) получаем

$$\text{Re } R = \frac{1}{2} \left[1 - \text{Re } b_{11} \left(\xi_1^2 - \frac{\omega^2 a^2}{c^2} \varepsilon \right)^{-1/2} \right] \\ - \frac{1}{2} \text{Re } b_{11} \frac{\omega^2 a^2}{c^2} (\varepsilon - 1) \left(\xi_1^2 - \frac{\omega^2 a^2}{c^2} \varepsilon \right)^{-3/2}, \quad (\text{П2.12})$$

$$\text{Im } R = \frac{\omega^5 a^5}{c^5} \left[30 \xi_1^2 \left(\xi_1^2 - \frac{\omega^2 a^2}{c^2} \varepsilon \right)^{1/2} \right]^{-1}. \quad (\text{П2.13})$$

Из конкретных расчетов по формулам (П2.11)–(П2.13) следует, что при $a \equiv a_{\text{out}} \ll \lambda$ действительная часть коэффициента отражения $\text{Re } R$ мала и составляет $\sim 0.04 - O(a_{\text{out}}^2/\lambda^2)$, а его мнимая часть $\text{Im } R$ оказывается еще более малой величиной. Таким образом, коэффициент передачи T_{w0} (см. разд.6) при переходе из волновода с субволновым поперечным сечением в ближнюю зону свободного пространства близок к единице с очень хорошей точностью. Это означает, что вычисление поля на выходе волновода конечной длины можно проводить без учета эффектов отражения на выходе.

Отметим, что с использованием соотношения (П2.13) может быть также получено следующее выражение для энергетического коэффициента передачи $T_{0\infty}$ свободного пространства из ближней зоны в дальнюю:

$$T_{0\infty} \approx \frac{2}{15 \xi_1^2} \frac{\omega^4 a^4}{c^4}. \quad (\text{П2.14})$$

Таким образом, при $a_{\text{out}}/\lambda \ll 1$ мы получили коэффициент $T_{0\infty}$, который оказывается пропорциональным $(a_{\text{out}}/\lambda)^4$.

1. Bethe H.A. *Phys. Rev.*, **66**, 163 (1944).
2. Bouwkamp C.J. *Rep. Prog. Phys.*, **17**, 35 (1954).
3. Marchand E.W., Wolf E. *J. Opt. Soc. Am.*, **59**, 79 (1969).
4. Pohl D.W., Denk W., Lanz M. *Appl. Phys. Lett.*, **44**, 651 (1994).
5. Kopelman R., Tan W. In: *Microscopy and spectroscopy images of the chemical state* (New York, Basel, Hon Kong: Marcel Dekker, 1993).
6. Harootunian A., Betzig E., Isaacson M.S., Lewis A. *Appl. Phys. Lett.*, **49**, 674 (1986).
7. Meixner A.J., Zeisel D., Vopp M.A., Tarrach G. *Opt. Eng.*, **34**, 2324 (1995).
8. Tang M., Cai S.M., Liu Z.F. *Opt. Commun.*, **146**, 21 (1998).
9. Klimov V.V., Letokhov V.S. *Opt. Commun.*, **106**, 151 (1994).
10. Зуев В.С., Кузнецова Т.И. *Квантовая электроника*, **24**, 462 (1997).
11. Knoll B., Keilmann F. *Opt. Commun.*, **162**, 177 (1999).
12. Климов В.В., Первенцев Я.А. *Квантовая электроника*, **29**, 9 (1999).
13. Вайнштейн Л.А. *Электромагнитные волны* (М.: Радио и связь, 1988).
14. Palanker D.V., Simanovskii D.M., Huie P., Smith T.I. *J. Appl. Phys.*, **88**, 6808 (2000).
15. Каценеленбаум Б.З. *Теория нерегулярных волноводов с медленно меняющимися параметрами* (М.: Изд-во АН СССР, 1961).
16. Шевченко В.В. *Плавные переходы в открытых волноводах* (М.: Наука, 1969).
17. Маркузе Д. *Оптические волноводы* (М.: Мир, 1974).
18. Ландау Л.Д., Лифшиц Е.М. *Квантовая механика* (М.: Наука, 1974).