

Введение

Одной из важнейших характеристик лазера является направленность генерируемого излучения.

В твердотельных и газовых лазерах поперечные размеры d световых пучков много больше длины волны излучения λ , поэтому направленность излучения, определяемая соотношением λ/d , оказывается очень малой и составляет десятые и сотые доли градуса.

В полупроводниковых лазерах толщина активного слоя, как правило, сравнима с длиной волны генерируемого излучения. Другой, перпендикулярный размер светового пучка полупроводникового инжекционного лазера (ПИЛ) может быть значительно больше и совпадает с размером самого ПИЛ. Следствием этих обстоятельств является существенное уширение диаграммы направленности по сравнению с другими типами лазеров.

На рис. 1 схематически представлена картина излучения ПИЛ в дальней зоне. Полные углы θ_1 и θ_2 , измеренные на уровне половины максимальной интенсивности, характеризуют расходимость луча соответственно в плоскости активного слоя и перпендикулярно ей. Обычно θ_1 составляет $\approx 10^\circ$, тогда как θ_2 принимает значительно большие значения ($35^\circ - 60^\circ$) и зависит от толщины активного слоя и распределения электромагнитного поля внутри активного слоя.

Для изучения пространственных характеристик излучения ПИЛ воспользуемся волноводной моделью ПИЛ, позволяющей рассматривать задачу в два этапа. На первом этапе пренебрегают влиянием конечной протяженности лазера и рассматривают электромагнитные волны в бесконечном диэлектрическом волноводе. На втором этапе вычисляют дифракционную расходимость получившегося волноводного решения на открытом конце волновода.

1. Волноводная модель ПИЛ

Схема полупроводникового инжекционного лазера приведена на рис. 2, где 1 – p -эмиттер, 2 – активный слой, 3 – n -эмиттер, R_1 и R_2 – выходные зеркала с коэффициентами отражения R_1 и R_2 . При протекании тока накачки от слоя 1 к слою 3 в активном слое возникает спонтанное излучение.

С увеличением плотности тока накачки j интенсивность спонтанного излучения возрастает, и при некоторой величине j_0 в актив-

ном слое создается инверсная населенность, приводящая к появлению усиления света и возникновению стимулированного излучения. Световая волна, распространяясь в плоскости активного слоя между зеркалами R_1 и R_2 , усиливается и испытывает потери, которые можно разделить на 3 типа:

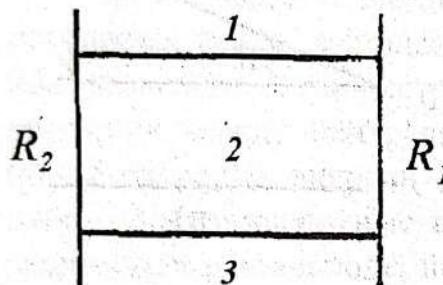
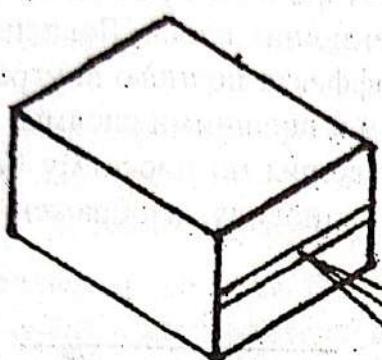


Рис. 2

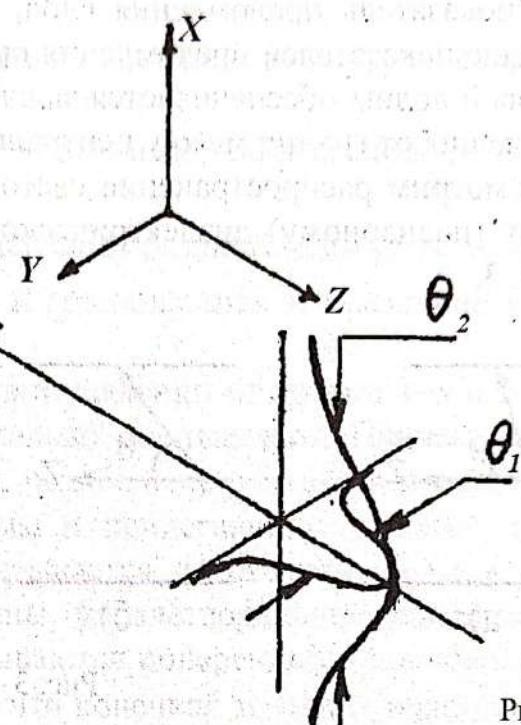


Рис. 1

- 1) поглощение на свободных носителях в активном слое;
- 2) потери вследствие ухода части света из активного слоя;
- 3) потери при отражении на выходных зеркалах.

Пороговое усиление, необходимое для достижения генерации, должно компенсировать потери при однократном проходе волны между зеркалами в прямом и обратном направлениях, т.е.

$$I_0 = I_0 R_1 R_2 \exp(-2\alpha L + 2gL). \quad (1)$$

Здесь I_0 – интенсивность возникшей световой волны, α – коэффициент потерь при распространении вдоль активного слоя, g – коэффициент усиления в активном слое, L – расстояние между зеркалами R_1 и R_2 . Или, переписывая (1) в другом виде, получим

$$g_{\text{пор}} \geq \alpha - \frac{\ln(R_1 R_2)}{2L}. \quad (2)$$

Ясно, что при распространении излучения между зеркалами усиливаться в большей степени будут те фотоны, которые при своем дви-

жении каким-то образом остаются, удерживаются, локализуются в активном слое. Фотоны, которые вылетают за пределы области усиления, попадают в пассивные эмиттеры и там поглощаются.

Явление локализации света можно обеспечить с помощью системы, называемой диэлектрическим волноводом и характеризуемой тем, что показатель преломления слоя, в котором распространяется свет, больше показателей преломления прилегающих слоев. Локализация световой волны обеспечивается за счет эффекта полного внутреннего отражения от границ между центральным и внешними слоями.

Рассмотрим распространение световой волны по плоскому бесконечному (планарному) диэлектрическому волноводу, изображенному на рис. 3.

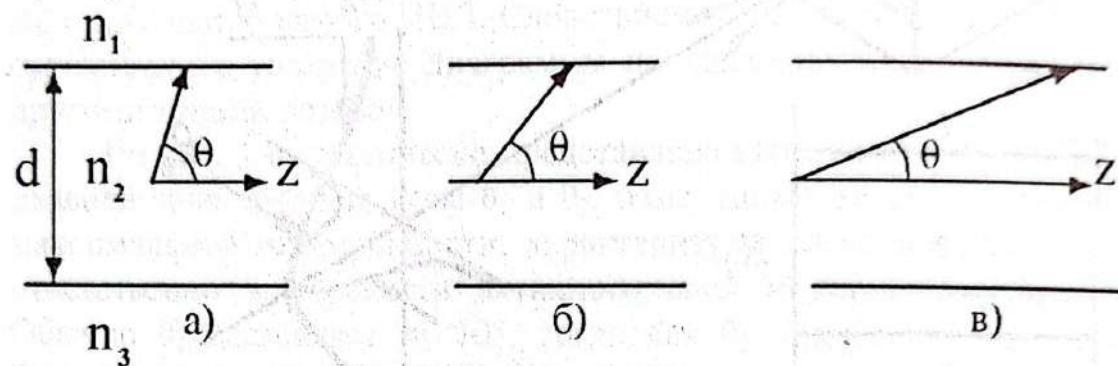


Рис. 3

Показатель преломления активного слоя n_2 больше, чем показатели преломления пассивных слоев n_1 и n_3 , причем считаем для определенности $n_2 > n_1 \geq n_3$. Угол θ между направлением распространения волны (направлением волнового вектора \vec{k}) с осью z , лежащей в плоскости волновода, назовем углом распространения волны.

Если электромагнитная волна распространяется вдоль слоя под углом θ таким, что на границах 1–2 и 2–3 она не испытывает полного внутреннего отражения, т.е. $\theta > \arccos\left(\frac{n_3}{n_2}\right)$, то большая часть фотонов будет уходить из области усиления в пассивные слои 1 и 3, вследствие чего волна быстро затухнет (рис. 3а).

Если угол распространения θ таков, что на границе 2–3 происходит полное внутреннее отражение, но на границе 1–2 волна преломля-

ется, т.е. $\arccos\left(\frac{n_1}{n_2}\right) < \theta < \arccos\left(\frac{n_3}{n_2}\right)$, то локализации света в активном слое также не происходит (рис. 3б).

Если угол распространения еще меньше, а именно $\theta < \arccos\left(\frac{n_1}{n_2}\right)$, то волна испытывает полное внутреннее отражение на обеих границах волновода и возникает локализация света в активном слое (рис. 3в).

В гомолазерах на p - n -переходе различие между n_2 и n_1 , n_3 составляет всего лишь 0,1–1%. В гетеролазерах это различие значительно выше и составляет $\approx 10\%$.

Интенсивность волны при удалении от границ 1–3 и 2–3 вглубь пассивных слоев экспоненциально уменьшается. Причем это затухание происходит тем быстрее, чем больше разница в показателях преломления между центральным и прилегающими слоями. Основной поток световой энергии оказывается сосредоточенным в активном слое. Электромагнитные волны, удовлетворяющие условиям локализации поля в волноводе, называются поперечными волноводными модами. Далее будет показано, что конечная толщина волновода d приводит к тому, что значения угла распространения являются дискретными, а число поперечных мод, существующих в волноводе, определяется толщиной волновода d и соотношениями между n_2 и n_1 , n_3 .

Приведенное выше качественное рассмотрение вопроса о локализации поля в активном слое было проведено на примере модели планарного диэлектрического волновода. Как показывают теоретические и экспериментальные исследования, именно эта модель наилучшим образом описывает волноводную структуру существующих в настоящее время конструкций ПИЛ. Перед тем, как более подробно рассмотреть модель трехслойного планарного диэлектрического волновода и связь его модовой структуры с характеристиками излучения ПИЛ, рассмотрим монохроматические электромагнитные волны в пространстве одного, двух и трех измерений и более простой случай планарного металлического волновода.

2. Монокроматические волны в пространстве одного, двух и трех измерений

Напомним некоторые сведения о колебаниях и волнах. Для простоты рассмотрим не электромагнитные колебания и волны, а колебания и волны, не требующие учета поляризации. Те, что описываются скалярными величинами, например, колебание струны, поверхности воды или натянутой мембранны, акустические волны. Гармоническое колебание какой-то скалярной величины U угловой частоты ω периода T описывается формулами:

$$U(t) = A \cos(-\omega t + \phi) = A \cos\left(-2\pi \frac{t}{T} + \phi\right), \quad (3)$$

где ϕ – начальная фаза колебания, A – его амплитуда. Период T – время, за которое аргумент функций меняется на 2π . Это колебание является решением уравнения колебаний

$$\frac{\partial^2 U}{\partial t^2} = -\omega^2 U. \quad (4)$$

Плоская волна, распространяющаяся вдоль оси z со скоростью c , описывается формулой

$$U(z, t) = U\left(\frac{z}{c} - t\right). \quad (5)$$

Она является решением одномерного волнового уравнения

$$\frac{\partial^2 U}{\partial t^2} = c^2 \frac{\partial^2 U}{\partial z^2}. \quad (6)$$

Если волна монокроматическая с угловой частотой ω и длиной волны $\lambda = \frac{2\pi}{\omega} c = c T$, то она описывается формулой

$$U\left(\frac{z}{c} - t\right) = A \cos\left(2\pi\left(\frac{z}{\lambda} - \frac{t}{T}\right) + \phi\right) = A \cos(kz - \omega t + \phi), \quad (7)$$

где $k = \frac{2\pi}{\lambda}$ – волновое число. Видно, что в любой фиксированной точке z происходит колебание с периодом T . Фиксируя момент времени t можно убедиться в том, что волна периодична по координате z с периодом λ . Увеличение координаты z на Δz приводит к увеличе-

нию аргумента функции на $2\pi \frac{\Delta z}{\lambda}$. Иначе говоря, на промежутке Δz происходит "набег фазы" $2\pi \frac{\Delta z}{\lambda} = k\Delta z$. Рассмотрим такую волну в другой системе координат, такой, что новые оси x , y , z составляют соответственно углы α , β и γ с нормалью к волновому фронту. Легко видеть, что в новой системе координат волна обладает пространственной периодичностью вдоль каждой из осей с периодами $\lambda_x = \frac{\lambda}{\cos \alpha}$; $\lambda_y = \frac{\lambda}{\cos \beta}$; $\lambda_z = \frac{\lambda}{\cos \gamma}$.

Соответственно (7) может быть записано в виде

$$\begin{aligned} U(\vec{r}, t) &= A \cos \left[2\pi \left(\frac{x}{\lambda_x} + \frac{y}{\lambda_y} + \frac{z}{\lambda_z} - \frac{t}{T} \right) + \phi \right] = \\ &= A \cos \left[\frac{2\pi}{\lambda} (x \cos \alpha + y \cos \beta + z \cos \gamma) - \frac{2\pi t}{T} + \phi \right] = \\ &= A \cos (\vec{k} \cdot \vec{r} - \omega t + \phi), \end{aligned} \quad (8)$$

где $|\vec{k}| = \frac{2\pi}{\lambda}$, вектор \vec{k} направлен по нормали к волновому фронту,

$k_x = \frac{2\pi}{\lambda_x}$; $k_y = \frac{2\pi}{\lambda_y}$; $k_z = \frac{2\pi}{\lambda_z}$. Длина волны – минимальное расстоя-

ние между соседними волновыми фронтами – $\lambda = \frac{2\pi}{k}$ совпадает с какой-либо из величин λ_x ; λ_y ; λ_z , если волновой вектор k направлен вдоль соответствующей оси. В этом случае остальные две величины обращаются в бесконечность. Разложение волнового вектора \vec{k} на компоненты k_x ; k_y ; k_z вовсе не означает, что волна "раскладывается" на суперпозицию (сумму) отдельных волн с волновыми векторами k_x ; k_y ; k_z . Волна одна, она распространяется во всем пространстве (в том числе и по координатным осям x ; y ; z). Величины k_x ; k_y ; k_z указывают на величину пространственной периодичности вдоль каждой оси (в общем случае различную) и равны "набегу фазы" при пере-

мещении на единицу длины вдоль соответствующей координаты. То есть изменению аргумента волны при фиксированных значениях времени и остальных координат. Как видно, вдоль каждой из координат волна имеет бегущий вид. Монохроматическая волна вида (8) – есть частное решение волнового уравнения для пространства трех измерений:

$$\frac{\partial^2 U}{\partial t^2} = c^2 \left(\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} \right). \quad (9)$$

Решения этого уравнения могут быть гораздо сложнее, чем плоская бегущая волна. Например, в трехмерном вакууме в отсутствии какого-либо поглощения энергии электромагнитная волна может быть бегущей волной для одного направления, чисто стоячей для другого, экспоненциальной затухающей волной для третьего направления. Вид решения определяется граничными условиями. Ответ на естественный вопрос, что надо понимать под величиной U в случае электромагнитного поля, отложим до четвертой части описания.

Найти решение уравнения (9) можно двумя способами. Первый – сразу искать решение, удовлетворяющее граничным условиям задачи. Второй – из частных, более простых решений уравнения (9), не удовлетворяющих граничным условиям, составить линейную комбинацию, такую, для которой граничные условия будут выполнены. Ниже будут применены оба способа. Начнем со второго.

Но чтобы понять разницу между этими двумя способами, перейдем с языка математиков, формально решающих абстрактно сформулированную задачу, на язык физиков, более наглядный. В нем удобно использовать такие понятия, как физическая реальность (далее ФР) и математическая модель (далее ММ). Например, натянутая струна конечной длины L с закрепленными концами – это физическая реальность, как и колебания различных точек струны. Они существуют на этой длине L и нигде больше. Но вместо колебания струны конечной длины L можно рассмотреть две абстрактные одномерные волны равной амплитуды, бегущие навстречу друг другу, образующие при интерференции такую стоячую волну, что на промежутке длины L она совпадает с физической реальностью – реальной струной. Для двумерного случая колебания прямоугольной мембранны конечных размеров – эти колебания и есть физическая реальность. А те несколько бегущих волн в бесконечном пространстве, таких, что их интерференционная картина, сумма их колебаний в каждой точке пространства совпадает с колебаниями соответствующей точки физической реальности, – это

математическая модель колебания двумерной мембраны. Целесообразность введения математической модели в том, что она дает простое и наглядное решение. Оно, в отличие от ФР, обладает пространственной периодичностью и заполняет собой все пространство. Но одна ячейка этой периодической двумерной структуры точно совпадает с физической реальностью, реальной мембраной конечных размеров.

Одним из простейших решений уравнений Максвелла является плоская электромагнитная волна с линейной поляризацией. Выбором системы координат и начала отсчета времени можно привести выражение для волны к виду

$$\vec{E} = (E_x, 0, 0); \quad E_x = E_0 \cos[2\pi(\frac{z}{\lambda} - \frac{t}{T})]; \quad (10)$$

$$\vec{H} = (0, H_y, 0); \quad H_y = H_0 \cos[2\pi(\frac{z}{\lambda} - \frac{t}{T})]. \quad (11)$$

Такая волна называется ТЕМ-волной. Название указывает на то, что и электрическое, и магнитное поля волны тангенциальны (касательны) к плоскости волнового фронта, т.е. и \vec{E} , и \vec{H} перпендикулярны направлению распространения волны (оси z).

Известно, что в одномерной стоячей волне расстояние между узлами составляет половину длины волны. Как объяснить гораздо большее, видимое глазом расстояние между темными полосами на экране, при демонстрации интерференционной картины от двух отверстий, освещаемых когерентным светом?

Рассмотрим интерференцию двух плоских монохроматических электромагнитных волн одинаковой амплитуды и частоты, симметрично направленных относительно оси z (левая часть рис. 4). Векторы \vec{E} волн направлены вдоль оси y , перпендикулярно плоскости рисунка. Напряженность электрического поля интерференционной картины во всем пространстве опишется в виде

$$\begin{aligned} & E_0 \sin(k_z z + k_x x - \omega t) - E_0 \sin(k_z z - k_x x - \omega t) = \\ & = 2E_0 \sin(k_x x) \cos(k_z z - \omega t) = \\ & = 2E_0 \sin(k_x \sin \theta) \cos(k_z \cos \theta - \omega t) = \\ & = E_a(x) \cos(k_z \cos \theta - \omega t). \end{aligned} \quad (12)$$

Знак "−" в левой части выражения (12) объясняется подходящим выбором начала координат и начала отсчета времени. При другом выборе в выражениях для обеих волн появятся начальные фазы. Задача

однородна относительно y , зависимости от y нет. В левой части рисунка изображены две интерферирующие волны, пунктиром обозначено мгновенное положение линий нулевой амплитуды каждой из двух плоских волн. В правой части рисунка сплошными линиями изображены не зависящие от времени линии нулевой амплитуды интерференционной картины. Пунктиром обозначено мгновенное положение линий нулевой амплитуды интерференционной картины.

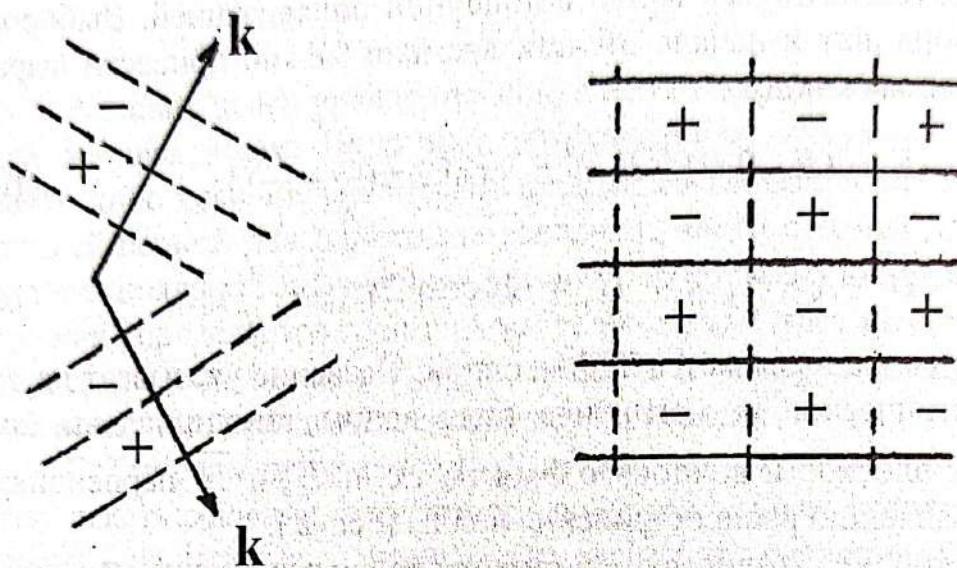


Рис. 4

ние линий нулевой амплитуды интерференционной картины, чье положение меняется со временем – перемещается вдоль оси z со скоростью $\frac{c}{\cos \theta}$. Знаки "+" и "-" соответствуют горбам и впадинам мгновенного распределения электромагнитного (электрического) поля между линиями нулей. Через полпериода $T/2 = \pi/\omega$ эти знаки сменятся на противоположные. Как видно из (12) и правой части рис. 4, вдоль оси z распространяется со скоростью $\frac{c}{\cos \theta}$ (большей скорости света) волна с частотой ω и постоянной распространения $h = k_z = k \cos \theta$, волна, имеющая более сложный характер.

Во-первых, в отличие от ТЕМ-волны, эта волна называется ТЕ-вольной, так как только электрическое поле перпендикулярно оси z – направлению распространения. Магнитное поле волны имеет продольную составляющую $H_z \neq 0$. (Если бы на рис. 4 волны были бы

однородна относительно y , зависимости от y нет. В левой части рисунка изображены две интерферирующие волны, пунктиром обозначено мгновенное положение линий нулевой амплитуды каждой из двух плоских волн. В правой части рисунка сплошными линиями изображены не зависящие от времени линии нулевой амплитуды интерференционной картины. Пунктиром обозначено мгновенное положе-

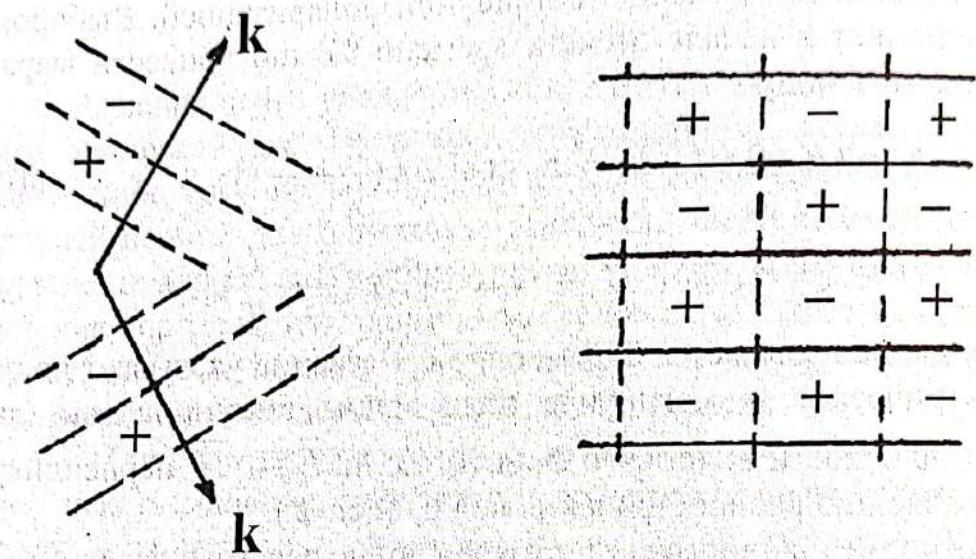


Рис. 4

ние линий нулевой амплитуды интерференционной картины, чье положение меняется со временем – перемещается вдоль оси z со скоростью $\frac{c}{\cos \theta}$. Знаки "+" и "-" соответствуют горбам и впадинам мгновенного распределения электромагнитного (электрического) поля между линиями нулей. Через полпериода $T/2 = \pi/\omega$ эти знаки сменяются на противоположные. Как видно из (12) и правой части рис. 4, вдоль оси z распространяется со скоростью $\frac{c}{\cos \theta}$ (большей скорости света)

волна с частотой ω и постоянной распространения $h = k_z = k \cos \theta$, волна, имеющая более сложный характер.

Во-первых, в отличие от ТЕ-волны, эта волна называется ТЕ-волной, так как только электрическое поле перпендикулярно оси z – направлению распространения. Магнитное поле волны имеет продольную составляющую $H_z \neq 0$. (Если бы на рис. 4 волны были бы

линейно поляризованы так, что напряженность магнитного поля была перпендикулярна плоскости рисунка, то в результате получилась бы ТМ-волна, волна, в которой перпендикулярно оси z было бы только магнитное поле, и имелась бы продольная составляющая $E_z \neq 0$ электрического поля.) Во-вторых, в отличие от ТЕ-волны амплитуды ТЕ-волны и ТМ-волны не постоянны вдоль фронта, а промодулированы по закону $\sin(k_x x)$. Величину $q = k_x$ называют поперечным волновым числом. Плоскостям $x k \sin \theta = m\pi$, где $m = 0, \pm 1, \pm 2\dots$, соответствует нулевая амплитуда. Минимальное расстояние между этими плоскостями

$$\Delta x = \frac{\lambda}{2 \sin \theta}. \quad (13)$$

Видно, что при стремлении угла θ к 90° , расстояние Δx стремится к $\frac{\lambda}{2}$; при стремлении θ к нулю расстояние Δx стремится к бесконечности, т.е. ТЕ-волна переходит в ТЕМ-волну, волну с постоянной по всему фронту амплитудой.

Предлагается подумать над тем, как изменится интерференционная картина на рисунке 4, если к каждой из двух изображенных слева бегущих волн будет добавлено по волне той же частоты, той же поляризации и той же амплитуды, но двигающейся в противоположном направлении.

3. Планарные (плоские) металлический и диэлектрический волноводы

Рассмотрим на рисунке 4 пару плоскостей, в которых результирующая амплитуда электрического поля постоянно равна нулю. Так как в этих плоскостях поля нет, то можно в эти плоскости ввести проводящие бесконечные металлические пластины, картина электромагнитного поля от этого не изменится. Такие две параллельные плоскости, электромагнитное поле между которыми является бегущей вдоль оси z , вдоль плоскостей ТЕ-волной, называются металлическим волноводом. Волна, распространяющаяся в волноводе при отсутствии полей и волн вне объема волновода, – это физическая реальность. Но часто полезно бывает рассмотреть и математическую модель – сумму (суперпозицию) двух парциальных волн – плоских монохроматиче-

ских волн, распространяющихся под углом θ к оси z во всем свободном пространстве (где нет этих параллельных металлических плоскостей). Эти описания тождественно совпадают внутри объема волновода. То есть можно представить (см. рис. 5), что в этом промежутке при

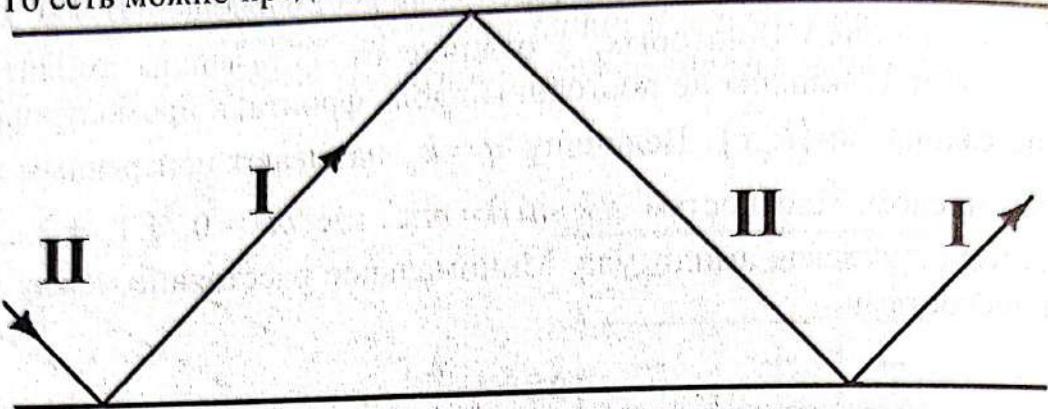


Рис. 5

отражении от верхней и нижней плоскостей происходит взаимопревращение волн — волна I переходит в волну II, которая после отражения от нижней плоскости переходит в волну I. Рис. 5 создает неверное впечатление, что такое взаимопревращение волн может происходить при любом расстояния d между металлическими плоскостями и при любом угле распространения θ . Но, как следует из (13), для волн с длиной волны λ существует ограничение на минимальный, "критический размер" волновода d . Эта величина не может быть меньше, чем $\lambda/2$. В противном случае электромагнитные волны с длиной волны λ не могут распространяться по волноводу. Если размер волновода больше критического, то в нем может распространяться ТЕ-волну, но не любая. Угол распространения волны θ может принимать лишь дискретные значения θ_m , причем нетрудно показать, что

$$\sin \theta_m = \frac{(m+1)\lambda}{2d}, \quad (14)$$

где d — толщина металлического волновода, λ — длина волны каждой из двух парциальных волн, составляющих электромагнитное поле в волноводе, $m = 0, 1, 2, 3\dots$ m_{\max} — номер поперечной моды, m_{\max} — максимально возможный в данном волноводе номер поперечной моды $m_{\max} = \left[\frac{2d}{\lambda} \right] - 1$ (здесь $[x]$ — есть целая часть числа x).

Формула (14) есть следствие условия, что между верхней и нижней плоскостями должно уложиться целое число ($m+1$) величин

$$\frac{\lambda}{2 \sin \theta_m}$$
, которая является половиной пространственного периода по

оси x каждой из двух парциальных волн.

Формулу (14) можно вывести и иначе. Поставим вопрос: каковы должны быть условия, чтобы поле внутри металлического волновода можно было представить в виде суммы двух плоских волн? Это фазовое условие, т.е. на рис. 5 надо учесть то обстоятельство, что отраженные, например, от нижней металлической плоскости волны должны иметь не только одно направление (это выполняется безусловно), но и быть согласованы по фазе. Это означает, что все лучи, распространяющиеся в одном направлении, принадлежат одной и той же плоской

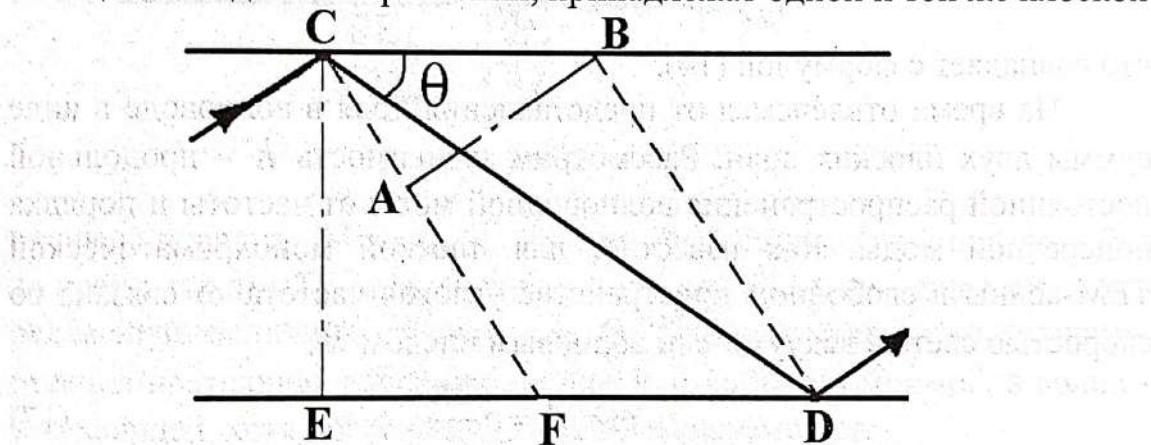


Рис. 6

волне, и всем точкам фазового фронта должна соответствовать одинаковая фаза колебаний. На рис. 6 изображены не только волновые векторы, но и фазовые фронты волн I (пунктиром). Видно, что исходящий из точки А в точку В луч не претерпевал отражения, так что соответствующий ему в точке В фазовый фронт направлен вдоль BD. Луч CD, исходящий из точки С в точку D, представляет собой отраженную волну, испытавшую отражение в точках С и D. В точке D ему должен соответствовать фазовый фронт, находящийся в фазе с фазовым фронтом луча AB в точке B. Для этого необходимо, чтобы длина оптического пути (сдвиг фазы) луча AB отличалась от длины оптического пути луча CD на число, кратное 2π . При отражении волны в точках С и D происходит сдвиг фазы на π , обеспечивающий равенство нулю

напряженности поля в этих точках, а суммарный сдвиг фаз равен 2π .

$$\text{Из рис. 6 следует: } CD = \frac{d}{\sin \theta}; DE = d \frac{1}{\tan \theta}$$

$$EF = d \tan \theta; BC = DE - EF; AB = BC \cos \theta = d \left(\frac{1}{\tan \theta} - \tan \theta \right) \cos \theta.$$

$$\text{Разность хода } CD - AB = 2d \sin \theta.$$

Полный фазовый сдвиг должен удовлетворять условию:

$$2d \sin \theta \cdot \frac{2\pi}{\lambda} = 2\pi(m+1). \quad (15)$$

Отсюда следует:

$$\sin \theta_m = \frac{(m+1)\lambda}{2d}, \quad (16)$$

что совпадает с формулой (14).

На время отвлечемся от представления поля в волноводе в виде суммы двух плоских волн. Рассмотрим зависимость h – продольной постоянной распространения волноводной моды от частоты и порядка поперечной моды. Как известно, для плоской монохроматической TEM-волны в свободном пространстве угловая частота ω связана со скоростью света в вакууме c и волновым числом k :

$$\omega = ck = c \frac{2\pi}{\lambda}, \quad (17)$$

где λ – длина волны в свободном пространстве.

При распространении плоской монохроматической TEM-волны в пространстве, заполненном диэлектриком с показателем преломления n_2 , в n_2 раз уменьшаются скорость распространения волны и ее длина. Постоянная распространения такой волны $h = kn_2 = \frac{\omega}{c} n_2$.

Из (16) и (17) следует, что продольная постоянная распространения TE-волны определяется условием:

$$h^2 = \left(\frac{\omega}{c} n_2 \right)^2 - \left[\frac{(m+1)\pi}{d} \right]^2, \quad (18)$$

где m – порядок поперечной моды, а член в квадратных скобках – это поперечное волновое число q , принимающее дискретные значения.

На рис. 7 изображена зависимость продольной постоянной распространения h от частоты ω для ТЕМ-волны и для ТЕ- и ТМ-волн

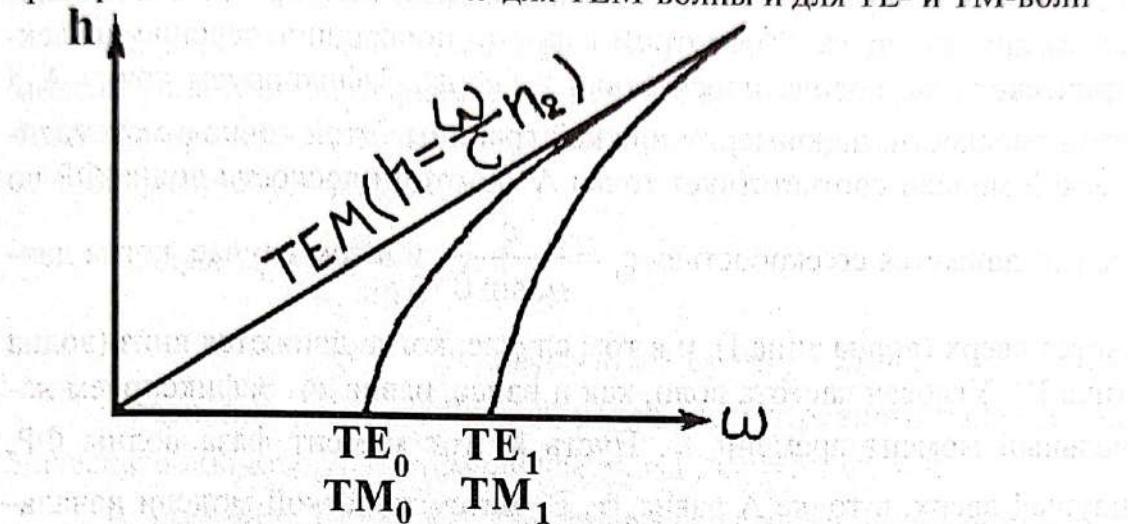


Рис. 7

нишего порядка. Мы видим, что для ТЕМ-волны $h = 0$ при $\omega = 0$, в то время как для ТЕ-волн (и аналогично для ТМ-волн) существует определенная величина отсечки для ω . При частотах ниже величины отсечки постоянная распространения h становится мнимой, а волна – затухающей, хотя поглощения энергии в системе нет.

Рассмотрим случай диэлектрического волновода с полным внутренним отражением электромагнитной волны на обеих границах. Будем считать, что реальная волна, волна физической реальности (далее ФР), существующая внутри центрального слоя волновода, при отражении от верхней и нижней границ раздела слоев испытывает сдвиг фаз не π (как для случая металлического волновода), а Φ_1 и Φ_3 соответственно. Математической моделью (далее ММ) этого случая будут две плоские монохроматические волны, распространяющиеся в однородной бесконечной среде (левая часть рис. 4). Но средой является не вакуум (как было в случае модели металлического волновода), а диэлектрик такой же, как в центральном слое диэлектрического волновода. Волны типа I и типа II ММ распространяются в этой среде со скоростью $\frac{c}{n_2}$, где n_2 – показатель преломления среднего слоя. В математической модели нет двух внешних слоев диэлектрика, нет границ

раздела, нет отражений электромагнитной волны от границ. Как же можно такой моделью описать, заменить процесс последовательного отражения волны ФР от границ раздела слоев в диэлектрическом волноводе? Можно представить для наглядности, что картина физической реальности и картина математической модели полупрозрачны и наложены друг на друга. Рассмотрим какое-то поперечное сечение диэлектрического волновода плоскостью $z = \text{const}$. Зафиксируем точку А в этой плоскости, например, у нижней границы. Этой точке в математической модели соответствует точка A^* . В этой плоскости волна ФР по

оси x движется со скоростью $c_x = \frac{c}{n_2 \sin \theta}$ и в том случае, когда дви-

жется вверх (волна типа I), и в том случае, когда движется вниз (волна типа II). Угловая частота волн, как и ранее, равна ω . Зафиксируем начальный момент времени t_1 . Пусть в этот момент фаза волны ФР, идущей вверх, в точке А равна ϕ . В математической модели начальную фазу идущей вверх волны ММ выберем так, чтобы в начальный момент времени t_1 в соответствующей точке A^* фаза волны ММ, идущей вверх, тоже равнялась ϕ , т.е. совпала с фазой волны ФР.

Рассмотрим теперь другой момент времени t_2 , когда волна ФР возвратится в точку А с измененной фазой. После прохода вдоль оси x ($z = \text{const}$) до верхней границы и отражения от нее фазой волны ФР станет $\phi + \Phi_1$ (за счет фазового сдвига Φ_1 при отражении). После достижения нижней границы и отражения от нее фазой волны станет $\phi + \Phi_1 + \Phi_3$ (за счет фазового сдвига Φ_3). В момент времени t_2 в точке А фаза волны ФР равна $\phi + \Phi_1 + \Phi_3$.

В математической модели в неизменной точке A^* за промежуток времени $t_2 - t_1$ фаза колебания изменится на $\omega(t_2 - t_1)$ и будет равна $\phi + \omega(t_2 - t_1)$. Если эта фаза совпадает с точностью до $2\pi m$ (где m – целое число) с фазой волны ФР, то волна ФР ничем не отличается от волны ММ типа I, распространяющейся во всем пространстве. Так как сечение волновода выбрано произвольно, как и точка А в этом сечении, то поле внутри центрального слоя диэлектрического волновода можно заменить суммой двух волн типа I и II, распространяющихся в однородном пространстве, заполненном диэлектриком с показателем преломления n_2 . Приведенное согласование математической модели и физической реальности объясняет, почему такое поле называется са-

самосогласованным. Условие самосогласованности поля или условие поперечного резонанса имеет вид

$$\omega \left(\frac{d}{c_x} + \frac{d}{c_x} \right) = \varphi_1 + \varphi_3 + 2\pi m . \quad (19)$$

Слева, в (19), изменение фазы колебания волны математической модели типа I (волна в фиксированной точке пространства – это колебание во времени) в точке А за время двойного прохода толщины d волной физической реальности.

Так как $c_x = \frac{c}{n_2 \sin \theta}$, а $k = \frac{\omega}{c}$, то

$$2kn_2 d \sin \theta - \varphi_1 - \varphi_3 = 2\pi m . \quad (20)$$

Уравнение (20) справедливо и для рассмотренного ранее металлического волновода. В этом случае $n_2 = 1$, $\varphi_1 = \pi$, $\varphi_3 = \pi$.

В диэлектрическом волноводе при отражении фазовый сдвиг Φ зависит от отношения величин показателя преломления среднего и внешнего слоя, от угла падения и от направления поляризации падающего света. Приведем выражение для Φ без вывода.

Для случая, когда вектор \vec{E} перпендикулярен плоскости падения (это соответствует случаю ТЕ-волны), сдвиг фазы равен

$$\Phi_{TE} = 2 \operatorname{arctg} \left\{ \frac{\sqrt{k^2(n_2^2 - n_1^2) - k^2 n_2^2 \sin^2 \theta_2}}{k n_2 \sin \theta_2} \right\} , \quad (21)$$

где $n_2 > n_1$, θ_2 – угол распространения волны в среднем слое.

Для случая, когда вектор \vec{H} перпендикулярен плоскости падения (это соответствует случаю ТМ-волны), сдвиг фазы

$$\Phi_{TM} = 2 \operatorname{arctg} \left\{ \frac{\sqrt{\cos^2 \theta_2 - (n_1/n_2)^2}}{(n_1/n_2)^2 \sin \theta_2} \right\} , \quad (22)$$

На рис. 8 изображена зависимость фазового сдвига Φ_{TE} от уг

ла распространения для ТЕ-моды. Видно, что при очень малых углах θ фазовый сдвиг Φ_{TE} стремится к π , т.е. отражение происходит как от

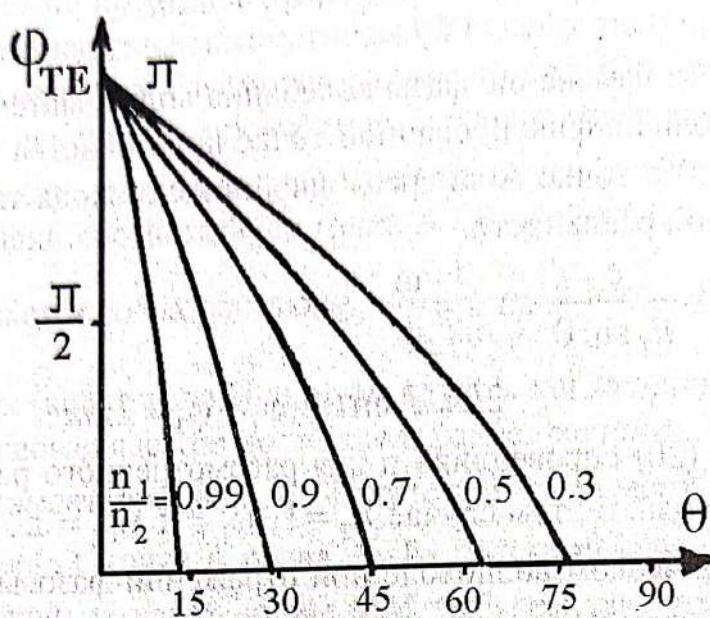


Рис. 8

металлической плоскости, и на границе раздела суммарная амплитуда падающей и отраженной волны равна нулю. В этом случае диэлектрический волновод ведет себя как металлический, заполненный диэлектриком с показателем преломления n_2 . Так как в общем случае Φ_1 и Φ_3 меньше π , то из условия поперечного резонанса (20) следует, что на толщине волновода d укладывается не целое число $(m + 1)$ величин $\frac{\lambda}{2n_2 \sin \theta}$, как в металлическом волноводе (см. (14)), а несколько

меньше, или, другими словами, целое число этих величин должно уложиться в несколько большем отрезке. На рис. 9 слева изображено распределение амплитуды поля нулевой моды в диэлектрическом волноводе толщины d , справа – в металлическом волноводе с толщиной $d_{\text{эфф}}$. Как видно на рис. 9, поле в диэлектрическом волноводе в среднем слое с показателем преломления совпадает с полем обозначенной пунктиром центральной части металлического волновода, заполненного диэлектриком с показателем преломления n_2 . Как будет показано в дальнейшем, отличие между этими волноводами заключается в том,

что в диэлектрическом волноводе в слоях 1 и 3 поле экспоненциально спадает с расстоянием от оси, а в металлическом волноводе поле обращается в ноль на границах металлического волновода толщины $d_{\text{эфф}}$. А вот поле в центральной части диэлектрического волновода совпадает с полем, существующим в той части металлического волновода, что ограничена пунктиром справа на рисунке. В зависимости от того, равны или нет друг другу показатели преломления слоев 1 и 3, эта часть располагается внутри металлического волновода симметричным или несимметричным образом. Для того чтобы найти распределение поля в диэлектрическом волноводе, сначала надо найти

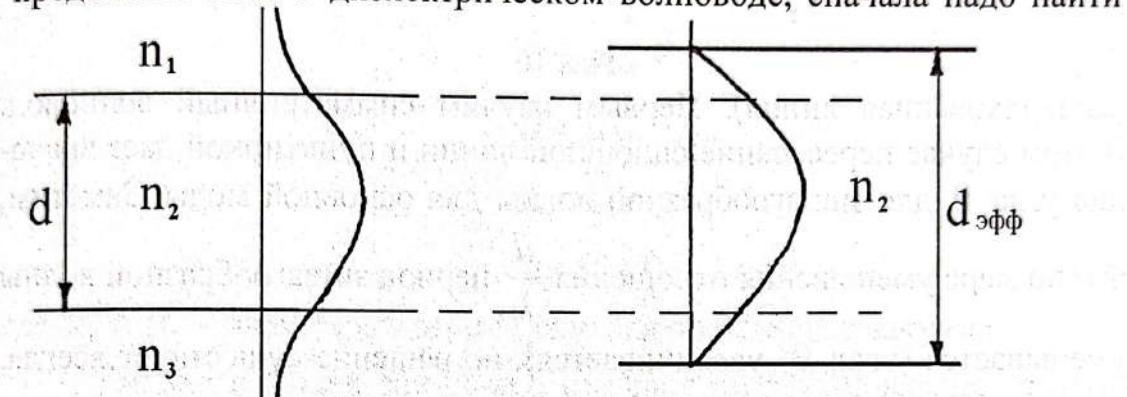


Рис. 9

угол θ_m – решение уравнения (20) при целых значениях параметра $m = 0, 1, 2\dots$ Если решение существует для $m = 0, 1\dots m_{\max}$, то максимальное число m_{\max} есть наивысший порядок поперечной моды, возможный в данном диэлектрическом волноводе.

Решение уравнения (20) можно провести графическим способом. Решения различны для симметричных ($n_1 = n_3$) и несимметричных диэлектрических волноводов. На рис. 10 представлено графическое решение уравнения поперечного резонанса для основной моды $m = 0$, которое дает дополнительную информацию о характеристиках диэлектрического волновода. На нем изображены зависимости от угла θ сдвига фаз за время прохода волны поперек центрального слоя вверх и вниз $2kn_2d \sin \theta$ (пунктирная кривая) и сумма фазовых сдвигов ($\phi_1 + \phi_3$) при отражении от границ слоя. Последнюю зависимость рассмотрим для двух типов волноводов: симметричного, когда $\phi_1 = \phi_3$, $n_1 = n_3$ (сплошная линия), и асимметричного волновода

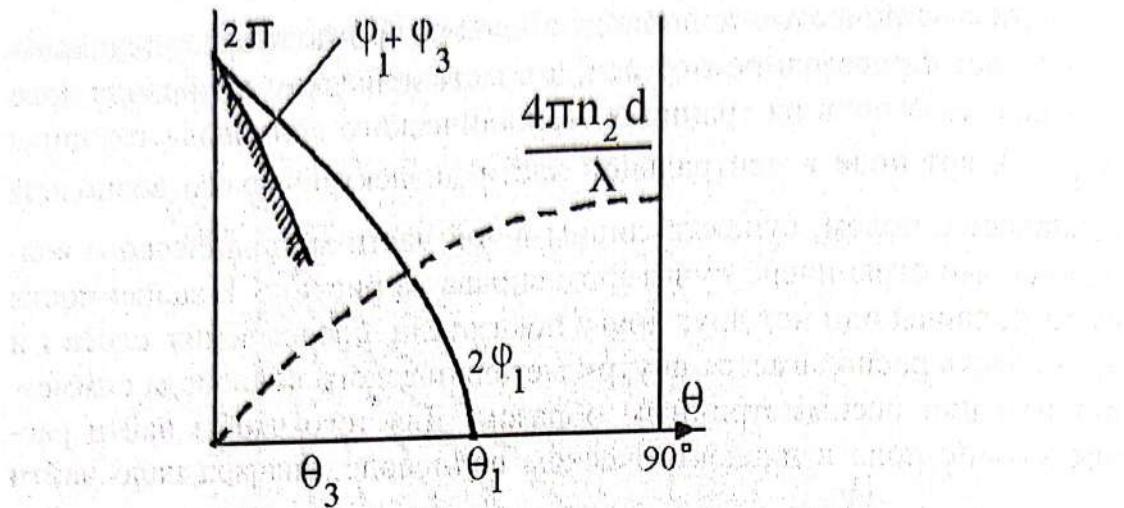


Рис. 10

(заштрихованная линия). Первым изучим симметричный волновод. В этом случае пересечение сплошной линии и пунктирной дает значение угла θ для зигзагообразной волны для основной моды. Заметим,

что по мере уменьшения отношения $\frac{d}{\lambda}$ период зигзагообразной волны

уменьшается (угол θ увеличивается), но решение существует всегда, даже когда толщина слоя становится очень малой. Это означает, что в симметричном диэлектрическом волноводе нет условий отсечки для основной моды в отличие от металлического волновода. Конечно, чем больше толщина волновода, тем больше волноводных мод будет в нем возбуждаться. При исследовании асимметричного волновода мы будем искать пересечение пунктирной и штриховой линий. Однако только часть кривой ($\Phi_1 + \Phi_3$), которая выделена штриховкой, лежит выше значения критического угла θ_1 для границы раздела 1–2. Для достаточно тонких слоев нет пересечения кривых выше границы отсечки, следовательно, в асимметричном волноводе не всегда может возбуждаться волноводная мода или, другими словами, даже для основной моды есть условие отсечки.

До сих пор решение задачи об электромагнитном поле в диэлектрическом волноводе искалось в виде математической модели – суммы двух более простых решений, плоских волн. При этом решение искалось только в центральном слое волновода, электромагнитные поля во внешних диэлектрических слоях не рассматривались.

Теперь воспользуемся другим подходом к решению задачи через решение уравнений Максвелла, удовлетворяющее граничным условиям на линиях раздела трех диэлектрических слоев.

4. Основы электромагнитной теории трехслойного диэлектрического волновода

Напомним вид уравнений Максвелла (в системе СИ):

$$\frac{\partial \vec{B}}{\partial t} = -\operatorname{rot} \vec{E}, \quad (23)$$

$$\frac{\partial \vec{D}}{\partial t} + \vec{j} = \operatorname{rot} \vec{H}, \quad (24)$$

$$\vec{D} = \epsilon \epsilon_0 \vec{E}, \quad (25)$$

$$\vec{B} = \mu \mu_0 \vec{H}, \quad (26)$$

$$\operatorname{div} \vec{D} = \rho, \quad (27)$$

$$\operatorname{div} \vec{B} = 0, \quad (28)$$

где ϵ_0 и μ_0 – электрическая и магнитная постоянные вакуума.

Так как в нашей задаче все три слоя диэлектрические, то плотность тока j , плотность электрического заряда ρ можно считать равными нулю. У немагнитных материалов μ очень близко к единице. Совокупность уравнений (23) – (28) полностью описывает электромагнитное поле в линейной изотропной среде.

Так как обычно на приемники электромагнитных волн действует электрическая составляющая электромагнитного поля, получим уравнение для вектора \vec{E} . Подставим величину $\vec{B} = \mu_0 \vec{H}$ в уравнение

$\frac{\partial \vec{B}}{\partial t} = -\nabla \times \vec{E}$ и возьмем ротор от обеих частей. В результате получим:

$$\nabla \times (\nabla \times \vec{E}) = -\mu_0 \frac{\partial}{\partial t} (\nabla \times \vec{H}). \quad (29)$$

При выводе уравнения (29) использовали предположение, что величина μ не зависит от координат. Подстановка $\nabla \times \vec{H} = \frac{\partial \vec{D}}{\partial t}$ и

$\vec{D} = \epsilon \epsilon_0 \vec{E}$ в правую часть (29) приводит к уравнению, содержащему только вектор \vec{E} :

$$\nabla \times (\nabla \times \vec{E}) + \epsilon \epsilon_0 \mu_0 \frac{\partial^2 \vec{E}}{\partial t^2} = 0. \quad (30)$$

Воспользовавшись известным векторным тождеством

$$\nabla \times (\nabla \times \vec{A}) = \nabla (\nabla \cdot \vec{A}) - \nabla^2 \vec{A}, \quad (31)$$

где \vec{A} – любой вектор, и, используя уравнение $\vec{E} = \frac{\vec{D}}{\epsilon \epsilon_0}$ и уравнение

$\operatorname{div} \vec{D} = 0$, получим

$$\nabla^2 \vec{E} = \epsilon \epsilon_0 \mu_0 \frac{\partial^2 \vec{E}}{\partial t^2}. \quad (32)$$

Здесь было использовано то обстоятельство, что величина ϵ постоянна в пространстве. Напомним, что скорость света в вакууме $c = 1/\sqrt{\epsilon_0 \mu_0}$; а величина $1/\sqrt{\epsilon \epsilon_0 \mu_0}$ – скорость света в диэлектрике с относительной диэлектрической проницаемостью ϵ . Показатель преломления диэлектрика $n = \sqrt{\epsilon}$. Таким образом, имеется три волновых уравнения. Для каждой из компонент вектора напряженности электрического поля \vec{E} – свое волновое уравнение:

$$\frac{\partial^2 E_x}{\partial t^2} = \frac{c^2}{n^2} \nabla^2 E_x; \quad \frac{\partial^2 E_y}{\partial t^2} = \frac{c^2}{n^2} \nabla^2 E_y; \quad \frac{\partial^2 E_z}{\partial t^2} = \frac{c^2}{n^2} \nabla^2 E_z. \quad (33)$$

Аналогично можно, исключив \vec{E} из уравнений Максвелла, получить три волновых уравнения для компонент \vec{B} .

Сформулируем те приближения, на основании которых будут получены все результаты:

1) волновод плоский, ось z направлена вдоль него, система однородна по координате y . В этом случае волновое уравнение зависит только от координат x и z ;

2) зависимость поля от времени имеет вид $\exp(-i\omega t)$, где $\frac{\omega}{c} = \frac{2\pi}{\lambda} = k$; c – скорость света в вакууме, λ – длина волны излучения

в вакууме, k – волновое число электромагнитной волны в вакууме частоты ω . Временную зависимость везде будем опускать;

3) считаем, что потери в волноводе малы и не могут изменить собственных значений волнового уравнения;

4) для определенности по-прежнему считаем $n_2 > n_1 \geq n_3$.

В электродинамике показано, что решению волнового уравнения для случая плоского диэлектрического волновода удовлетворяют два типа волн:

TE-волна, $\vec{E} = (0, E_y, 0)$; $\vec{H} = (H_x, 0, H_z)$,

TM-волна, $\vec{H} = (0, H_y, 0)$; $\vec{E} = (E_x, 0, E_z)$.

Начнем с изучения свойств TE-волн. Для них $E_x = 0$, $E_z = 0$. Подставив известную времененную зависимость в (23), получим

$$H_x = -\frac{i}{\omega\mu} \frac{\partial E_y}{\partial z}, \quad (34)$$

$$H_z = \frac{i}{\omega\mu} \frac{\partial E_y}{\partial x}, \quad (35)$$

$$H_y = 0. \quad (36)$$

Таким образом, зная только зависимость $E_y(x, z)$ можно однозначно определить все компоненты вектора \vec{H} . (Для TM-волн по зависимости $H_y(x, z)$ определяются компоненты вектора \vec{E}). Компонента E_y получается как решение приведенного волнового уравнения или уравнения Гельмгольца (используя известную времененную зависимость \vec{E}). Для каждого из трех диэлектрических слоев уравнение имеет вид

$$\frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial z^2} + n_i^2 k^2 E_y = 0, \quad i=1,2,3. \quad (37)$$

В каждом слое E_y будем искать в виде $E_y = E_y(x) E_y(z)$:

$$\frac{\partial^2 E_y(x)}{\partial x^2} + q^2 E_y(x) = 0, \quad (38)$$

$$\frac{\partial^2 E_y(z)}{\partial z^2} + h^2 E_y(z) = 0, \quad (39)$$

$$q^2 + h^2 = n_i^2 k^2 \quad i=1, 2, 3. \quad (40)$$

Нетрудно видеть, что $E_y(z) \sim \exp(ihz)$ и величина h является постоянной распространения волны вдоль волновода. Непрерывность тангенциальных составляющих поля на границах приводит к тому, что h одинакова во всех трех слоях. Поперечное волновое число q , определяющее конфигурацию поля во внутреннем и внешних слоях волновода, наоборот, зависит от толщины волновода и соотношений между показателями преломления слоев, образующих волновод.

Поскольку нас интересуют только те типы колебаний, которые удерживаются волноводом, т.е. испытывают полное внутреннее отражение на границах 1–2 и 2–3, то решение $E_y(x)$ ищем в виде

$$E_{y2}(x) = \cos(qx + \beta), \quad \left(-\frac{d}{2} < x < \frac{d}{2}\right); \quad (41)$$

$$E_{y1}(x) = A \exp(-px), \quad \left(x > \frac{d}{2}\right); \quad (42)$$

причем

$$E_{y3}(x) = B \exp(rx), \quad \left(x < -\frac{d}{2}\right); \quad (43)$$

$$q^2 + h^2 = n_2^2 k^2; \quad -p^2 + h^2 = n_1^2 k^2; \quad -r^2 + h^2 = n_3^2 k^2$$

или

$$q^2 + p^2 = (n_2^2 - n_1^2)k^2; \quad q^2 + r^2 = (n_2^2 - n_3^2)k^2. \quad (44)$$

Поля удовлетворяют следующим граничным условиям:

$$\begin{aligned} E_{y1} &= E_{y2}, \quad x = \frac{d}{2}; \\ H_{z1} &= H_{z2}, \quad x = \frac{d}{2}; \\ E_{y2} &= E_{y3}, \quad x = -\frac{d}{2}; \\ H_{z2} &= H_{z3}, \quad x = -\frac{d}{2}. \end{aligned} \quad (45)$$

Экспоненциальное спадание напряженности поля по x в слоях 1 и 3 можно интерпретировать физически, как образование распространяющейся вдоль границы волновода волны, чья амплитуда быстро затухает при удалении от оси волновода.

Поле внутри волновода можно интерпретировать как суперпозицию (сумму) двух парциальных плоских волн $\exp(-iqx+ihz)$ и $\exp(iqx+ihz)$, распространяющихся вдоль волновода под углом θ к оси волновода, определяемым с помощью соотношения

$$\sin \theta = \frac{q}{k n_2}. \quad (46)$$

Максимальный порядок поперечной моды и ему соответствующий максимальный угол распространения θ_{\max} , а следовательно, максимальное значение поперечного волнового числа q_{\max} определяется условием сохранения полного внутреннего отражения на границах волновода. Поскольку $n_2 > n_1 \geq n_3$, то в активном слое могут возникать моды с величиной поперечного волнового числа q_{\max} , удовлетворяющей неравенству

$$q_{\max} \leq k n_2 \sin \theta_{np}, \quad \text{где} \quad \theta_{np} = \frac{\pi}{2} - \Psi_{np},$$

Ψ_{np} – угол полного внутреннего отражения на границе 1–2 (см. рис. 3).

Нетрудно показать, что

$$\sin \theta_{np} = \frac{\sqrt{n_2^2 - n_1^2}}{n_2}, \quad q_{\max} \leq \sqrt{n_2^2 - n_1^2}. \quad (47)$$

Для дальнейших выкладок удобно ввести понятие предельного поперечного волнового числа $q_{np} = k \sqrt{n_2^2 - n_1^2}$. Если q_{\max} обозначает то максимальное поперечное волновое число, которое может быть в данном диэлектрическом волноводе с конкретной толщиной центрального слоя d , то q_{np} не зависит от толщины d и определяется только показателями преломления на границе 1–2. Можно показать, что $q_{\max} \leq q_{np}$ и q_{\max} стремится к q_{np} при неограниченном увеличении толщины d .

Выведем теперь соотношение, связывающее величину поперечного волнового числа q с параметрами волновода. Используя (41) и (45), получим:

$$\cos\left(\frac{qd}{2} + \beta\right) = A \exp\left(-\frac{pd}{2}\right), \quad (48)$$

$$-q \sin\left(\frac{qd}{2} + \beta\right) = -pA \exp\left(-\frac{pd}{2}\right), \quad (49)$$

$$\cos\left(-\frac{qd}{2} + \beta\right) = B \exp\left(-\frac{rd}{2}\right), \quad (50)$$

$$-q \sin\left(-\frac{qd}{2} + \beta\right) = -rB \exp\left(-\frac{rd}{2}\right) \quad (51)$$

или

$$\operatorname{tg}\left(\frac{qd}{2} + \beta\right) = \frac{p}{q}, \quad (52)$$

$$\operatorname{tg}\left(-\frac{qd}{2} + \beta\right) = -\frac{r}{q}. \quad (53)$$

Исключая из системы (52), (53) β , получим требуемое выражение

$$\operatorname{tg}(qd) = \frac{\frac{p}{q} + \frac{r}{q}}{1 - \frac{pr}{q^2}}. \quad (54)$$

Обозначая $\frac{p}{q} = \operatorname{tg} \alpha$; $\frac{r}{q} = \operatorname{tg} \psi$ получим

$$\operatorname{tg}(qd) = \operatorname{tg}(\alpha + \psi) \quad (55)$$

или

$$qd = \alpha + \psi + \pi m = \operatorname{arctg}\left(\frac{p}{q}\right) + \operatorname{arctg}\left(\frac{r}{q}\right) + \pi m, \quad (56)$$

Воспользуемся выражением (44) и связями между p , r , q :

$$qd = \operatorname{arctg}\left(\frac{\sqrt{(n_2^2 - n_1^2)k^2 - q^2}}{q}\right) + \operatorname{arctg}\left(\frac{\sqrt{(n_2^2 - n_3^2)k^2 - q^2}}{q}\right) + \pi m.$$

Если в полученном уравнении воспользоваться соотношением $q = kn_2 \sin \theta$, то это уравнение переходит к виду

$$dk n_2 \sin \theta = \operatorname{arctg} \left(\frac{\sqrt{(n_2^2 - n_1^2) k^2 - k^2 n_2^2 \sin^2 \theta}}{kn_2 \sin \theta} \right) + \\ + \operatorname{arctg} \left(\frac{\sqrt{(n_2^2 - n_3^2) k^2 - k^2 n_2^2 \sin^2 \theta}}{kn_2 \sin \theta} \right) + \pi m. \quad (57)$$

Легко видеть, что уравнение (57) есть ранее полученное уравнение (20) – условие поперечного резонанса или условие самосогласованности поля в волноводе для ТЕ-волны, т.е. при подстановке в (20) формулы (21).

Уравнение (57) можно упростить, введя безразмерную переменную $Q = \frac{q}{q_{np}} \leq 1$ и обозначение $\gamma^2 = \frac{n_2^2 - n_1^2}{n_2^2 - n_3^2} \ll 1$.

В этом случае условие нарушения волноводности на границе с минимальным скачком показателя преломления означает, что $Q=1$.

Уравнение (57) приобретает вид

$$Q dq_{np} = \operatorname{arctg} \left(\sqrt{\frac{1}{Q^2} - 1} \right) + \operatorname{arctg} \left(\sqrt{\frac{1}{Q^2 \gamma^2} - 1} \right) + \pi m.$$

Если воспользоваться соотношением $x = \operatorname{arctg} \left(\sqrt{\frac{1}{\cos^2 x} - 1} \right)$,

то получим окончательное выражение

$$Q d q_{np} = \arccos Q + \arccos(\gamma Q) + \pi m \quad (58)$$

или

$$Q \frac{2\pi}{\lambda} d \sqrt{n_2^2 - n_1^2} = \arccos Q + \arccos(\gamma Q) + \pi m. \quad (59)$$

Каждому порядку поперечной моды $m = 0, 1, 2, \dots m_{\max}$ соответствует определенное Q_m – решение уравнения (59). Число корней этого уравнения определяет число поперечных мод, которые могут распространяться в волноводе. Максимальный порядок моды определяется из условия $Q=1$.

На рис. 11 изображено графическое решение уравнения (59) для несимметричного волновода, найдены Q_0 и Q_1 для основной (нулевой) и первой поперечной моды.

Рассмотрим приближенное решение уравнения (59) и определим критерий его применения.

Пусть $Q_m \ll 1$. Тогда $Q_m d q_{np} = \frac{\pi}{2} - Q_m + \frac{\pi}{2} - \gamma Q_m + \pi m$

или $Q_m = \frac{\pi(m+1)}{d q_{np} + 1 + \gamma}$, т.е.

$$q_m = q_{np} Q = \frac{\pi(m+1)}{d \left(1 + \frac{1+\gamma}{dq_{np}} \right)} = \frac{\pi(m+1)}{d(1+\delta)}, \quad (60)$$

где $\delta = \frac{1+\gamma}{dq_{np}} = \frac{\lambda(1+\gamma)}{2d\pi\sqrt{n_2^2 - n_1^2}}$.

При условии $\delta \ll 1$ $q_m \approx \frac{\pi(m+1)}{d}$, т.е. поперечные волновые числа

в диэлектрическом волноводе близки по величине к поперечным волновым числам в металлическом волноводе, если

$$\frac{d}{\lambda} \gg \frac{1+\gamma}{2\pi\sqrt{n_2^2 - n_1^2}}. \quad (61)$$

Параметр δ показывает, насколько эффективная толщина волновода $d_{eff} = d(1+\delta)$ (см. рис. 9) больше действительной толщины d .

Как правило, критерий (61) довольно часто выполняется на практике, что позволяет, как мы увидим ниже, достаточно просто рассчитывать диаграмму направленности ПИЛ,

Нами был рассмотрен случай ТЕ-поляризации. Нетрудно получить аналогичные выражения и для ТМ-поляризации. Однако, как правило, в ПИЛ возбуждаются волны с ТЕ-поляризацией.

Рассмотрим основные факторы, определяющие порядок возбуждаемой поперечной моды. С увеличением порядка m поперечной моды увеличивается поперечное волновое число q_m . Из (44) видно, что

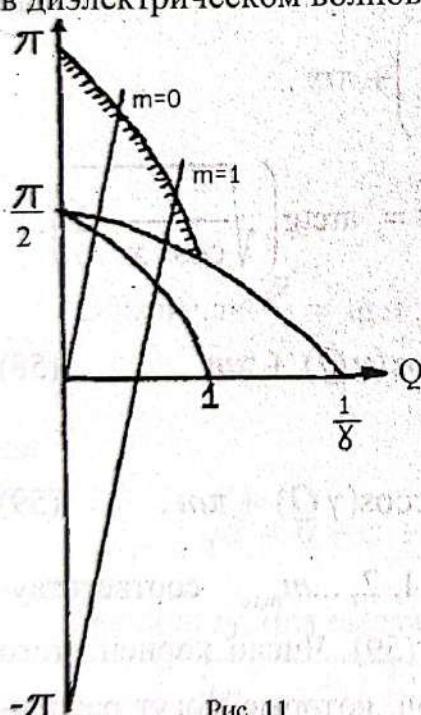


Рис. 11

возрастание q_m приводит к уменьшению поперечных волновых чисел p и r в слоях, прилегающих к волноводному слою, т.е. с увеличением m возрастает глубина проникновения светового поля в пассивные слои. Поскольку коэффициент поглощения в этих слоях не равен нулю, то увеличение m будет приводить к возрастанию внутрирезонаторных потерь.

С другой стороны, с увеличением m и q_m возрастает угол распространения парциальной плоской волны $\sin \theta_m = \frac{q_m}{k n_2}$ и увеличивается угол падения волны на выходную грань резонатора.

На рис. 12 приведена зависимость коэффициента пропускания T выходного зеркала при падении плоской волны из среды с показателем преломления 3,6 (показатель преломления GaAs) на границе раздела среды с вакуумом. Предполагается, что $d \gg \frac{\lambda}{2n_2}$. Как следует

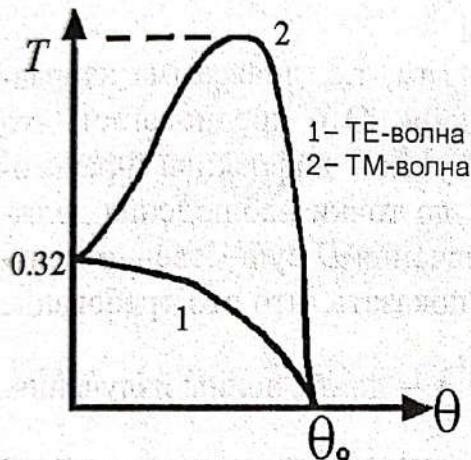


Рис. 12

из рис. 12, для TE-поляризации с увеличением угла падения θ коэффициент пропускания падает (т.е. уменьшаются потери на выход), а для TM-поляризации коэффициент пропускания растет.

Поэтому для TM-волн наименьшие потери имеет поперечная мода с $m = 0$, распространяющаяся под минимальным углом к оси волновода. В

случае TE-волн минимальные потери определяются конкуренцией между возрастанием внутрирезонаторных потерь при увеличении m и уменьшением при этом потерь на излучение через зеркало лазера. Вследствие этого минимум потерь реализуется при некотором угле θ_{opt} , величина которого определяется поглощением в слоях, прилегающих к волноводному слою, и пропусканием зеркала.

В лазере будет возбуждаться та поперечная мода, угол распространения которой θ_m максимально близок к оптимальному значению θ_{opt} . Из результатов расчетов в этом случае следует, что порог возбуждения TE-волны оказывается меньше, чем порог возбуждения нулевой

ТМ-моды. Экспериментальные исследования показали, что излучение полупроводниковых инжекционных лазеров обычно имеет ТЕ-поляризацию (т.е. возникают ТЕ-волны), что подтверждает правильность результатов теоретического анализа.

5. Диаграмма направленности излучения ПИЛ

Рассмотрим теперь, как связана поперечная структура поля в активном слое с диаграммой направленности излучения ПИЛ.

Если толщина активного слоя $d \gg \frac{\lambda}{2n_2}$ много больше длины

волны излучения в активном слое, то можно считать, что коэффициент отражения выходного зеркала близок к френелевскому, а распределение поля на зеркальной грани резонатора такое же, как и внутри волновода. Распределение интенсивности излучения на зеркальной грани принято называть картиной ближней зоны.

Распределение интенсивности по углам, т.е. диаграмма направленности, определяет картину дальней зоны. Это предполагает, что имеет место дифракция в параллельных лучах – дифракция Фраунгофера. Отсюда следует, что расстояние R до точки наблюдения должно выбираться из условия: размер тела свечения D существенно меньше, чем первая зона Френеля. Нетрудно показать, что это требование

приводит к соотношению $R \gg \frac{D^2}{\lambda}$; где λ – длина волны излучения.

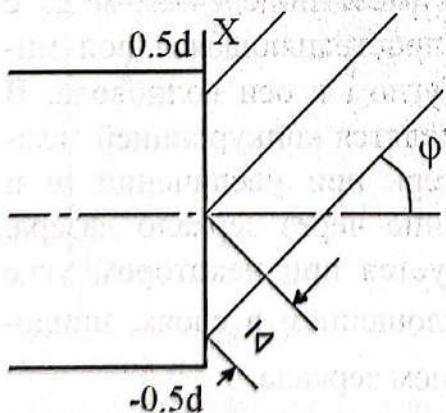


Рис. 13

Рассчитаем диаграмму направленности излучения. Волны, исходящие от центральной точки среднего слоя волновода и от точки, удаленной от центра на расстояние x , идущие от выходного зеркала под углом φ , отличаются по фазе на величину $\Delta = ik x \sin \varphi = ik \bar{\Delta}$ (см. рис. 13).

Амплитуда волны S , излучаемой под углом φ , дается выражением

$$dS(\varphi) \sim E(x) \exp(i k x \sin \varphi) dx;$$

$$S(\phi) \sim \int_{-\frac{d}{2}}^{\frac{d}{2}} E(x) \exp(i k x \sin \phi) dx. \quad (62)$$

Конкретизируем вид функции $E(x)$. Рассмотрим два случая, когда $E(x) \sim \cos(qx)$ и когда $E(x) \sim \sin(qx)$.

Для случая, когда $E(x) \sim \cos(qx)$, непосредственное интегрирование дает

$$\begin{aligned} S(\phi) &\sim \int_{-\frac{d}{2}}^{\frac{d}{2}} \cos(qx) \exp(i k x \sin \phi) dx = \\ &= \frac{1}{2} \int_{-\frac{d}{2}}^{\frac{d}{2}} [\exp(iqx) + \exp(-iqx)] \cos(qx) \exp(i k x \sin \phi) dx = \\ &= \frac{\sin\left(q \frac{d}{2} + k \frac{d}{2} \sin \phi\right)}{q + k \sin \phi} + \frac{\sin\left(q \frac{d}{2} - k \frac{d}{2} \sin \phi\right)}{q - k \sin \phi}. \end{aligned} \quad (63)$$

Для случаев, когда $E(x) \sim \sin(qx)$ – это интегрирование дает

$$\begin{aligned} S(\phi) &\sim \int_{-\frac{d}{2}}^{\frac{d}{2}} \sin(qx) \exp(i k x \sin \phi) dx = \\ &= \frac{1}{2i} \int_{-\frac{d}{2}}^{\frac{d}{2}} [\exp(iqx) - \exp(-iqx)] \exp(i k x \sin \phi) dx = \\ &= \frac{\sin\left(k \frac{d}{2} \sin \phi + q \frac{d}{2}\right)}{k \sin \phi + q} - \frac{\sin\left(k \frac{d}{2} \sin \phi - q \frac{d}{2}\right)}{k \sin \phi - q}. \end{aligned} \quad (64)$$

Напомним, что дифракция электромагнитной волны той же частоты на щели шириной d , зависимость амплитуды от угла описывается функцией

$$A(\phi) \sim \frac{\sin\left(k \frac{d}{2} \sin \phi\right)}{k \sin \phi}. \quad (65)$$

Формулы (63) и (64) содержат по два аналогичных слагаемых, отличающихся от выражения (65) тем, что центральный дифракционный максимум этого слагаемого приходится не на нулевой угол (как указывает (65)), а на угол, определяемый соотношением

$$k \sin \phi = \pm q. \quad (66)$$

Нетрудно показать, что условие (66) выражает обычный закон преломления парциальной плоской волны, распространяющейся в волноводе. Действительно,

$$k \sin \phi = q = k n_2 \sin \theta,$$

$$\frac{\sin \phi}{\sin \theta} = n_2.$$

Из выражений (63), (64) ясно, что диаграмма направленности излучения ПИЛ имеет многолепестковый характер.

Для часто встречающихся случаев, когда поле внутри активного слоя диэлектрического волновода близко к полю внутри металлического волновода, т.е. выполнено неравенство (61), можно считать, что

$$q_m \approx \frac{\pi(m+1)}{d}, \text{ т.е. на толщине волновода } d \text{ укладывается целое число } (m+1) \text{ полупериодов функции } E(x).$$

При этом возможно два случая:

$$m = 0, 2, 4, \dots \quad \beta_m = 0, \quad E(x) \sim \cos(q_m x),$$

$$E\left(\pm \frac{d}{2}\right) \sim \cos\left(q_m \frac{d}{2}\right) = 0; \quad \sin\left(q_m \frac{d}{2}\right) = \pm 1$$

– четные поперечные моды;

$$m = 1, 3, 5, \dots \quad \beta_m = \frac{\pi}{2}, \quad E(x) \sim \sin(q_m x)$$

– нечетные поперечные моды.

В этих случаях выражения (63) и (64) упрощаются:

$$S(\phi) \sim \frac{\cos\left(\frac{\pi d}{\lambda} \sin \phi\right)}{\left(\frac{\lambda(m+1)}{2d}\right)^2 - \sin^2 \phi} \quad - \text{четные моды}; \quad (67)$$

$$S(\phi) \sim \frac{\sin\left(\frac{\pi d}{\lambda} \sin \phi\right)}{\left(\frac{\lambda(m+1)}{2d}\right)^2 - \sin^2 \phi} \quad - \text{нечетные моды}. \quad (68)$$

Анализ выражений (63), (64), (67), (68) показывает, что дифракционная картина дальнего поля для случая $m \neq 0$ имеет два главных максимума, причем направления на главные максимумы близки к углам, даваемым выражением

$$\sin \phi_{\max} = \frac{q_m}{k} = \frac{\lambda(m+1)}{2d}. \quad (69)$$

Полуширина главного максимума равна $\frac{\lambda}{2d}$. Между главными максимумами находятся $m-1$ промежуточных максимумов. Для основной моды – $m=0$ диаграмма направленности имеет один главный максимум; $\phi_{\max} = 0$.

На рис. 14 приведены распределения амплитуды напряженности электрического поля излучаемой волны в волноводе, распределения по оси x интенсивности излучения на зеркальной грани резонатора (ближнее поле излучения) и диаграммы направленности (далнее поле излучения) для поперечных мод разных порядков.

Из (69) легко определить толщину активного слоя, зная порядок моды m и направление на главный максимум ϕ_{\max} :

$$d = \frac{\lambda(m+1)}{2 \sin \phi_{\max}}. \quad (70)$$

Выражение (70) справедливо, если возбуждается мода с $m \neq 0$. При $m=0$ диаграмма направленности имеет один главный максимум. Нетрудно показать, что угол преломления парциальной плоской волны в этом случае соответствует углу, при котором интенсивность лепест-

ка уменьшается до уровня $0.6 I_{\max}$. При этом $d = \frac{\lambda}{2 \sin \phi_{0.6}}$, где $\phi_{0.6}$ отсчитывается по уровню $I(\phi) = 0.6 I(0)$.

Подводя итог вышесказанному, можно сделать выводы:

- 1) активный слой ПИЛ должен обладать хорошими волноводными характеристиками;
- 2) волноводные свойства активного слоя ПИЛ хорошо описываются в рамках модели плоского диэлектрического волновода;
- 3) с увеличением толщины активного слоя в ПИЛ происходит возбуждение поперечных мод с порядками, отличными от нулевого;
- 4) диаграмма направленности излучения ПИЛ, в котором возбуждаются поперечные моды с порядками, отличными от нулевого, имеет два главных максимума, направления на которые задаются углом преломления парциальных плоских волн, соответствующих возбужденной моде;
- 5) число промежуточных максимумов в диаграмме направленности равно $m - 1$, где m — порядок поперечной моды;
- 6) возбуждение поперечных мод высоких порядков существенно увеличивает расходимость излучения ПИЛ.

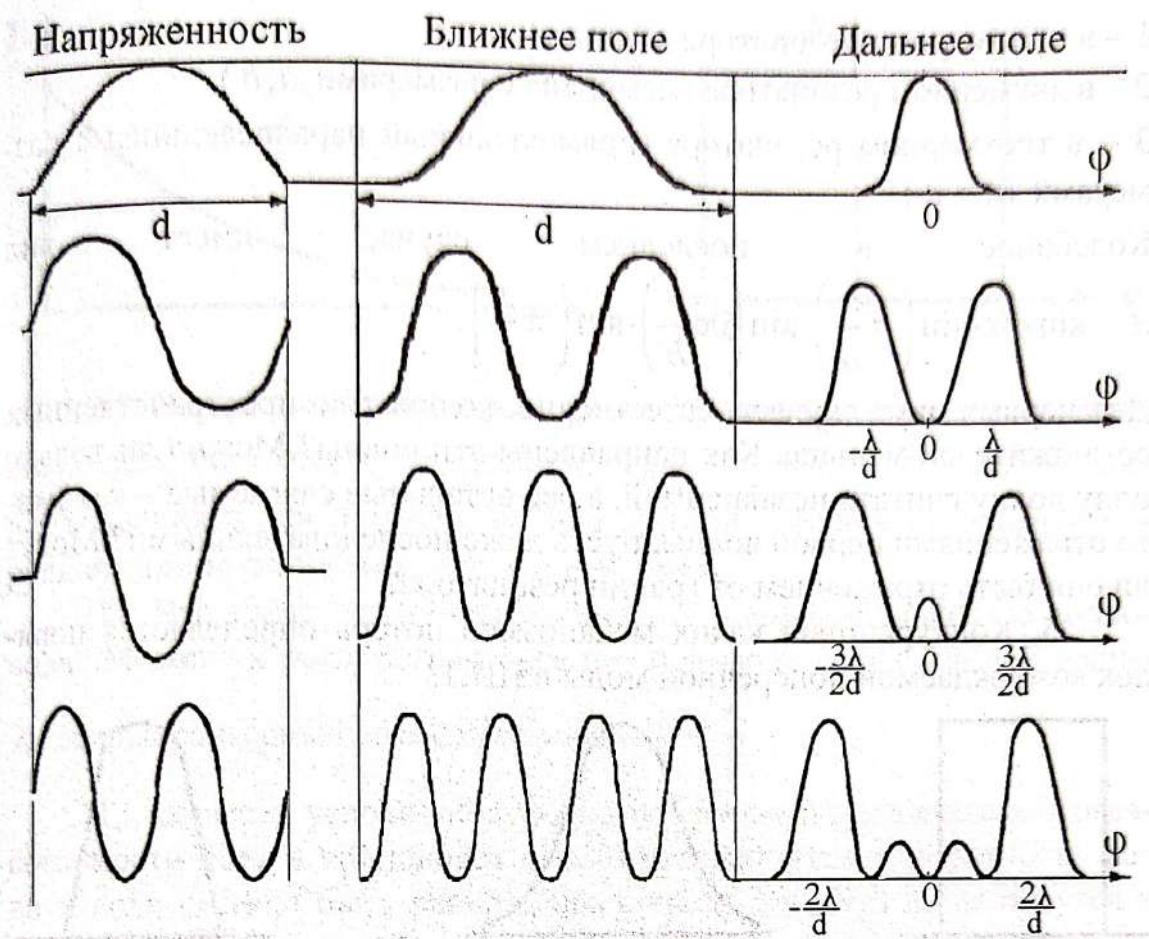


Рис. 14

Контрольные вопросы

- Под углом 2α интерферируют две плоские монохроматические волны. Какая картина возникнет во всем пространстве? Написать зависимость поля от координат в виде произведения соответствующих функций. Описать, какова картина поля на произвольно ориентированном экране, и нарисовать ее.
- Рассчитать максимальное число поперечных мод в волноводе $\epsilon_2 > \epsilon_3 \geq \epsilon_1$ в приближении металлического волновода, заполненного диэлектриком. Когда справедливо это приближение?
- На каком расстоянии нужно поставить фотоприемник, чтобы правильно измерить диаграмму направленности источника с размером тела свечения D ?
- Сколько независимых плоских волн нужно для описания монохроматического колебания (для простоты – акустического) в трех случаях:

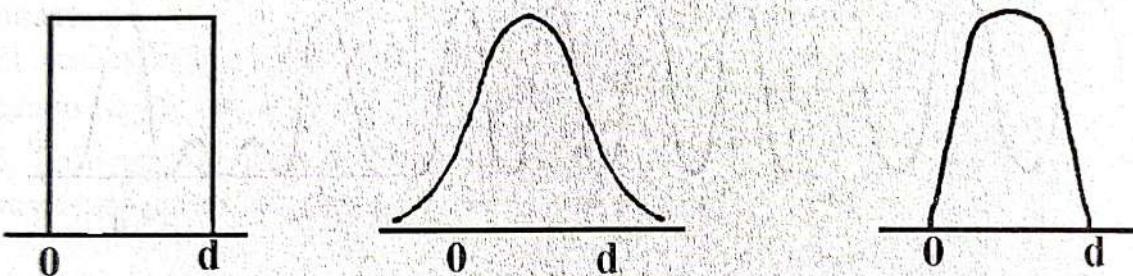
- 1 – в одномерном резонаторе длины a ,
- 2 – в двумерном резонаторе (мембрана с размерами a, b),
- 3 – в трехмерном резонаторе (прямоугольный параллелепипед с размерами a, b, c).

Колебание в последнем случае имеет вид

$$A \sim \sin \omega t \cdot \sin \left(\pi \frac{x}{a} \right) \cdot \sin \left(\pi \frac{y}{b} \right) \cdot \sin \left(\pi \frac{z}{c} \right).$$

Для первых двух случаев, естественно, количество пространственных сомножителей меньше. Как направлены эти волны? Можно ли только одну волну считать независимой, а все остальные слагаемые – какими-то отражениями первой волны, пусть даже последовательными? Могут ли они быть отражением от границ резонатора?

5. Конкуренцией каких механизмов потерь определяется порядок возбуждаемой поперечной моды в ПИЛ?



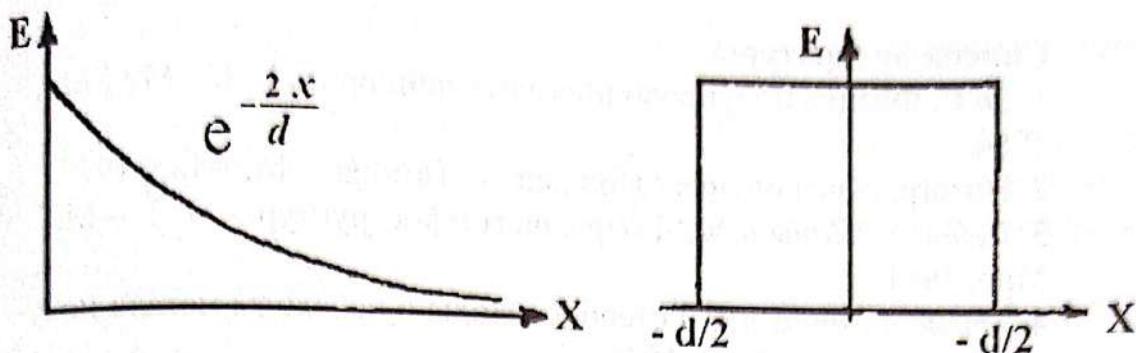
6. Даны три источника с распределениями световых полей на излучающей поверхности. Как будут отличаться их диаграммы направленности?

7. Как будет меняться полуширина диаграммы направленности, если мы будем увеличивать толщину волноводного слоя ПИЛ?

8. На щель шириной D падает плоская волна. На расстоянии R от щели расположен фотоприемник (расстояние измеряется по нормали к плоскости щели). Как будет меняться сигнал от фотоприемника при изменении R ?

9. В чем отличие диэлектрического волновода от металлического волновода?

10. Сравнить полуширины диаграмм направленности от таких двух источников.



11. Вычислить критический размер (толщину) диэлектрического волновода, состоящего из слоев $n_1 = 3.4$, $n_2 = 3.6$, $n_3 = 3.2$ – для возбуждения в ней поперечных мод нулевого, первого и второго порядков. Длина излучения лазера $\lambda = 0,9$ мкм.

12. Что такое "критическое сечение" для металлического волновода? Может ли распространяться по волноводу свет с длиной волны λ , если характерный размер волновода $d < \frac{\lambda}{2}$?

13. Вывести условие поперечного резонанса (условие самосогласованности поля в волноводе). В (20) утверждается, что разность фаз двух волн должна быть равна целому числу 2π . Эти фазы берутся в один момент времени? В одной точке пространства? Нужно ли складывать эти две волны?

14. Провести графическое решение уравнения поперечного резонанса в диэлектрическом несимметричном волноводе для основной и неосновной мод.

15. Как будет меняться полуширина диаграммы направленности, если будет увеличиваться толщина волноводного слоя ПИЛ?

16. Как будет меняться диаграмма направленности с увеличением порядка m поперечной моды? Чем определяется направление на главные максимумы в диаграмме направленности?

Список литературы

1. Зи С. Физика полупроводниковых приборов. Т. 2. – М.: Мир, 1984.
2. Интегральная оптика / Под ред. Т. Тамира. – М.: Мир, 1978.
3. Кейси Х., Паныш М. Лазеры на гетероструктурах. Т. 1. – М.: Мир, 1981.
4. Милнес А., Фойхт Д. Гетеропереходы и переходы металл-полупроводник. – М.: Мир, 1975.
5. Елисеев П.Г. Введение в физику инжекционных лазеров. – М.: Наука, 1983.
6. Грибковский В.П. Полупроводниковые лазеры. – Минск.: Университетское, 1988.
7. Пейн Г. Физика колебаний и волн. – М.: Мир, 1979.
8. Крауфорд Ф. Волны. – М.: Наука, 1976.

Экспериментальная часть

Цель работы

Исследование диаграмм направленности излучения полупроводникового инжекционного лазера (ПИЛ) в плоскости, перпендикулярной активному слою, и в плоскости активного слоя.

Методика эксперимента

Блок-схема установки приведена на рис. 15. Образец 1 закрепляется во вторичной обмотке импульсного трансформатора 2, вокруг которого вращается фотоприемник 5. На входе фотоприемника ФЭУ-28 перпендикулярно плоскости движения фотоприемника имеется узкая цель. Питание на ФЭУ подается от блока питания 7, сигнал с ФЭУ подается на вход селективного микровольтметра 6 (У2-6).

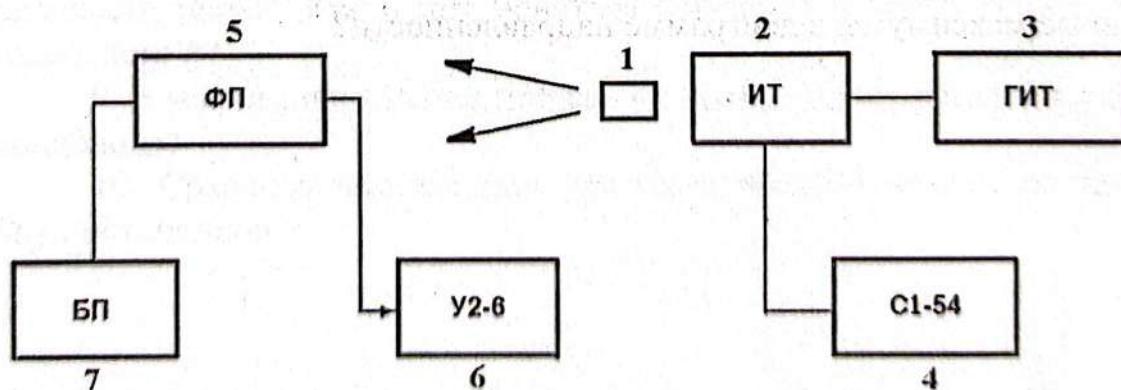


Рис. 15

Контроль импульсов тока накачки ПИЛ осуществляется осциллографом по схеме, приведенной на рис. 16.

Сигнал с измерительного сопротивления $R = 0,1 \text{ Ом}$, включенного во вторичную обмотку импульсного трансформатора последовательно с исследуемым образцом ПИЛ, поступает на вход Y осциллографа С1-54. Амплитуда тока накачки $I_{\text{нак}} = \frac{V_{\text{изм}}}{R}$, где $V_{\text{изм}}$ – амплитуда напряжения на экране осциллографа.

Угол поворота фотоприемника отсчитывается визуально. Сигнал с фотоприемника поступает на вход селективного микровольтметра У2-6. Измеряя зависимость сигнала фотоприемника от угла поворота фотоприемника, можно определить характер пространственного распределения излучения ПИЛ.

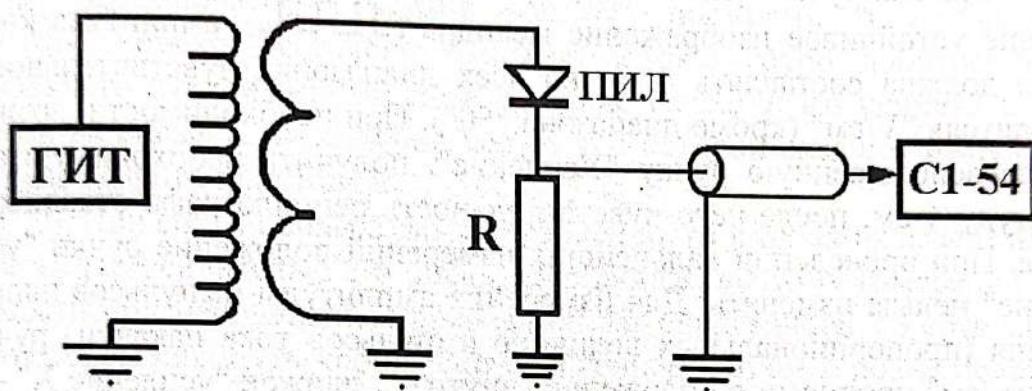


Рис. 16

Порядок проведения работы

1. Включение установки.

1.1. Включить средний рубильник лабораторного щитка.

1.2. Включить рубильник в нижней части стенда.

1.3. Нажать кнопку "сеть" осциллографа С1-54.

1.4. Включить тумблер "сеть" генератора импульсов ГИТ 6-2.

Тумблер "пуск" должен быть выключен, крайний правый тумблер должен быть в положении "6 кГц". Ручка "Ампл. тока" должна быть повернута до упора против часовой стрелки.

1.5. Включить тумблер "сеть" блока питания электромотора редуктора.

1.6. У высоковольтного стабилизированного выпрямителя ВС-22 тумблер "напряжение выхода" поставить в положение "выкл", после чего включить тумблер "сеть 220В".

1.7. Включить тумблер "сеть" селективного усилителя У2-6.

1.8. Дать прогреться приборам 10–15 минут.

2. Произвести калибровку осциллографа С1-54 для измерения амплитуды входного сигнала, пропорционального току накачки лазера. Для этого ручку "синхронизация" переключить в "калибр", ручку "длит./см" в положение "100 μ S", движок "калибраторы" поставить в положение "амплитуды", движок "Вход Y" поставить в левое положение. Маленькую ручку над этим движком переключить в положение "широкая". Выход "калибраторы амплитуды" соединить со входом Y осциллографа (нижний левый вход).

Вращая ручки "режим запуска" и "уровень запуска", получить на экране устойчивое изображение меандра (, амплитуда которого должна составлять 4 см на всех диапазонах чувствительности усилителя "V/см" (кроме диапазона "50"). При необходимости, вращая вышерасположенную ручку "Усиление", получить необходимую амплитуду 4 см, после чего чувствительность осциллографа откалибрована. При проведении дальнейших измерений положение ручки "усление" нельзя изменять. Для измерения амплитуды импульсов напряжения (пропорциональных величине импульсов тока накачки) ручку "длит./см" поставить в положение "внутр.~"; движок "усление X" – в положение "В.Ч." .

3. В измерительную головку, в которой находятся импульсный трансформатор, посеребренный держатель образца лазера и измерительное сопротивление $R = 0,1$ Ома, вставить образец лазера так, чтобы черная точка или пометка "+", нарисованная на торце лазера, оказалась бы в углублении держателя и была бы видна.

Шнур от выхода измерительной головки "контроль тока" подать на вход "Y" осциллографа.

4. Повернуть ручку "ампл. тока" генератора импульсов тока ГИТ6-2 на два деления. На экране осциллографа должно появиться изображение импульса тока накачки. Вращением ручек "режим запуска" и "усление X" добиться устойчивого изображения импульса тока, ручками "фокус" и "астигматизм" добиться наилучшей видимости, измерять величину импульса.

5. На высоковольтном выпрямителе ВС-22 переключить тумблер "Напряжение выхода" в положение "ВКЛ". Установить напряжение 1.2 кВ.

6. Повернуть измерительную головку относительно горизонтальной оси так, чтобы плоскость $p-n$ -перехода была бы параллельна щели фотоприемника.

7. Кронштейн с фотоприемником установить строго против нулевого значения лимба.

8. Выход сигнала с фотоприемника подать на вход усилителя У2-6. Ручку "диапазон частоты" усилителя поставить в положение (2,8–10 кГц). Настроить усилитель на максимальную чувствительность. Для этого ручку переключателя "режим работы" оставить в положение "первая узкая", после чего, вращая ручку "установка частоты узкая–грубо", добиться максимального отклонения стрелки прибора. Затем переключатель "режим работы" поставить в положение "2 узкая" и ручкой "установка частоты – 2 узкая–грубо" добиться максимального отклонения стрелки прибора. После этого усилитель настроен на сигналы именно той частоты, с которой следуют импульсы сигналов с фотоприемника.

9. При снятии диаграммы направленности излучения лазера надо учесть, что усилитель У2-6 очень инерционен, поэтому измерения надо вести при неподвижном положении фотоприемника. То есть электромотором сдвинуть кронштейн с фотоприемником в новое положение, остановить его и измерить новое значение сигнала фотоприемника. Угол поворота отсчитывается по лимбу, расположенному под кронштейном.

Снять зависимость сигнала ФЭУ-28 от угла поворота при 2–3 значениях тока накачки. После этого повернуть измерительную головку относительно горизонтальной оси и снять зависимость сигнала фотоприемника от угла для случая, когда плоскость $p-n$ -перехода перпендикулярна щели фотоприемника при 2–3 значениях тока накачки.

10. Проделать такие измерения для 2–3 образцов.

11. По результатам, приведенным в теоретическом разделе, и на основании полученных экспериментальных данных оценить толщину активного слоя исследуемых образцов.

12. После окончания измерений установить ручку "ампл. тока" генератора импульсов в нулевое положение, выключить тумблер "пуск". На высоковольтном выпрямителе переключатель "выс. напр." поставить в положение "выкл", затем выключить все приборы, рубильник внизу стенда, рубильник лабораторного щитка.