

一种新的基于最小平方逼近的广角光束传播方法^{*}

夏金松 余金中

(中国科学院半导体研究所, 集成光电子国家重点实验室, 北京 100083)

(2002 年 7 月 2 日收到, 2002 年 8 月 16 日收到修改稿)

用最小平方逼近展开传播算子, 实现了一种新的半矢量显式高阶有限差分光束传播方法. 这种方法中不需要选择参考折射率, 并在整个传播常数(包括辐射模的传播常数)分布区域进行逼近, 解决了在泰勒展开和庞德逼近中存在的参考折射率选择和远离展开点误差增大等问题. 用这种方法对几种典型波导结构进行了数值模拟, 模拟结果验证了算法的正确性和可靠性.

关键词: 光束传播方法, 有限差分, 集成光学, 数值方法

PACC: 0340K, 0260, 4280L

1. 引 言

随着集成光子器件研究的不断发展, 器件的结构越来越复杂, 用解析方法^[1]对系统进行分析变得不可能, 于是对复杂光子器件数值模拟工具的需求不断加强, 这使得各种模拟方法得到快速发展. 光束传播方法(beam propagation method)便是其中最主要的方法之一. 简单地说, 光束传播方法认为光场在传播方向上以很小的步长传播, 下一步的光场分布完全由上一步的光场决定, 这样三维的波动方程可以简化为二维方程, 便于快速求解. 光束传播方法出现后便受到了广泛的重视, 得到了快速的发展, 到现在已经有了 20 多年的历史. 开始时主要是用快速傅里叶变换(FFT)求解标量亥姆霍兹方程^[2], 有限差分方法的引入使得性能和应用范围得到了扩展^[3-7]. 但是这些方法有一个明显的缺点: 标量近似完全没有考虑光波内在的矢量特性, 所以这种方法不能获得光在非均匀介质中传播时的极化特性^[8], 也不能用来模拟光在各向异性介质中的传播. 半矢量方法可以部分解决这个问题. 在半矢量方法中, 将光场分解为两种没有耦合的模式 TE, TM, 两种模式可以分别求解^[3,5]. 全矢量方法的出现完全解决了这个问题, 它完全考虑了模式间的耦合^[6]. 为了求解矢量亥姆霍兹方程发展了很多方法, 其中大部分隐式方法需

要矩阵求逆或者求矩阵特征值, 而对于大型稀疏矩阵, 求逆和求特征值需要极长的计算时间, 除非应用特殊的数值技巧, 这些方法在时间上不经济^[4]. 显式方法则避免了这个问题, 因为它只需要作矩阵的相乘运算^[5]. 为了简化求解过程, 在求解二阶亥姆霍兹方程时, 往往采用缓变包络近似和傍轴近似, 可以去掉二阶项^[7]. 这种做法虽然使求解过程简化, 但限制了光束传播方法的适用范围, 不能在大角度传播问题中使用, 而且不能有效地模拟多模波导和辐射模. 泰勒展开和庞德逼近可以较好地解决这个问题^[5,9], 但是在这些逼近中存在参考折射率的选择问题, 另外在远离展开点处逼近程度减弱. 为了避免这个问题, 本文采用最小平方逼近在整个区域上对传播算子进行逼近, 实现了一种半矢量的显式广角光束传播方法. 该方法可以容易地推广到全矢量形式.

2. 基本原理

在无电荷、无电流的介质中由麦克斯韦方程组可以得到亥姆赫兹方程:

$$\nabla^2 \mathbf{E}(x, y, z) + k_0^2 n^2 \mathbf{E}(x, y, z) = \nabla(\nabla \cdot \ln n^2(x, y, z)) \cdot \mathbf{E}(x, y, z), \quad (1)$$

式中 k_0 为真空中的波数, n 为介质的折射率, \mathbf{E} 为已经消去时谐部分的电矢量复振幅. 将电矢量展开为标量形式, 并认为折射率沿 z 轴不变化或者变化

^{*} 国家自然科学基金(批准号 69896260 和 69990540)和国家重点基础研究发展规划项目(批准号 G2000-03-66)资助的课题.

很缓慢,忽略折射率对 Z 的偏微分,可以得到

$$\begin{aligned} & \frac{\partial^2}{\partial z^2} \begin{pmatrix} E_x \\ E_y \end{pmatrix} + \begin{pmatrix} A_{xx} & A_{xy} \\ A_{yx} & A_{yy} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \\ & = 0 \Leftrightarrow \frac{\partial^2}{\partial z^2} \bar{E} + A \bar{E} = 0, \end{aligned} \quad (2)$$

式中

$$\begin{aligned} A_{xx} E_x &= \frac{\partial}{\partial x} \left(\frac{1}{n^2} \frac{\partial}{\partial x} (n^2 E_x) \right) + \frac{\partial^2}{\partial y^2} E_x + n^2 k_0^2 E_x, \\ A_{xy} E_y &= \frac{\partial}{\partial x} \left(E_y \frac{1}{n^2} \frac{\partial n^2}{\partial y} \right), \\ A_{yy} E_y &= \frac{\partial}{\partial y} \left(\frac{1}{n^2} \frac{\partial}{\partial y} (n^2 E_y) \right) + \frac{\partial^2}{\partial x^2} E_y + n^2 k_0^2 E_y, \\ A_{yx} E_x &= \frac{\partial}{\partial y} \left(E_x \frac{1}{n^2} \frac{\partial n^2}{\partial x} \right). \end{aligned} \quad (3)$$

如果只考虑前向传播,方程(2)的形式解表示为

$$\bar{E}(Z_0 + \Delta z) = \exp(-j\sqrt{A}\Delta z)\bar{E}(Z_0). \quad (4)$$

从(4)式可以看出,只要将等号右边的传播算符用已知矩阵 A 展开,便可以重复于运用(4)式得到整个 Z 轴上的光场分布图.展开传播算子分两步进行,首先将 \sqrt{A} 展开成 A 的级数,传统的做法是用泰勒展开,但是如前所述,泰勒展开存在参考点选择和远离展开点误差增大等问题,所以本文采用最小平方逼近展开.展开后 A 的表达式为

$$R_M(A) = \sum_{i=0}^M c_i A^i, \quad (5)$$

平方意义下的误差为

$$\Delta(c) = \int_a^b (R_M(x) - \sqrt{x})^2 u(x) dx. \quad (6)$$

(6)式很容易理解为整个逼近区域上的平方误差,区间 $[a, b]$ 为逼近区间,取为传播矩阵的特征值(传播常数)平方的分布范围, $u(x)$ 为权函数,可以用来控制误差分布特点和逼近特性,下标 M 为展开阶数, M 越大则逼近精度越好.在最小平方逼近中,系数 c_i 必须使得 $\Delta(c)$ 取最小值,于是(6)式对 c_i 求偏导数为零,由此可计算得到系数 c_i ,得到最小平方逼近的展开式.在具体计算展开系数时,可以用格拉姆-施米特方法求解,系数很容易得到.显然,在前述的展开方式中不需要选择展开点,而且逼近是在整个传播常数分布范围内进行,在整个区域里都可以获得很好的逼近精度.

在获得了 \sqrt{A} 的展开式 $R_M(A)$ 后,下一步便是对(4)式等号右边的传播算子中指数部分进行泰勒展开:

$$\begin{aligned} \exp(-j\sqrt{A}\Delta z) &\cong \exp(-j\Delta z R_M(A)) \\ &\cong \sum_{i=0}^N \frac{(-j\Delta z R_M(A))^i}{i!} \\ &= \sum_{i=0}^N \frac{(-j\Delta z (\sum_{l=0}^M c_l A^l))^i}{i!} = P_N(A), \end{aligned} \quad (7)$$

$P_N(A)$ 为展开表达式, N 为泰勒展开阶数, N 越大精度越高,但是 N 不能一味地增大,一是会增加计算时间,二是要和第一步最小平方逼近配合,如果平方逼近的精度不高,泰勒展开的精度再高,整个算法的精度也不会很高.

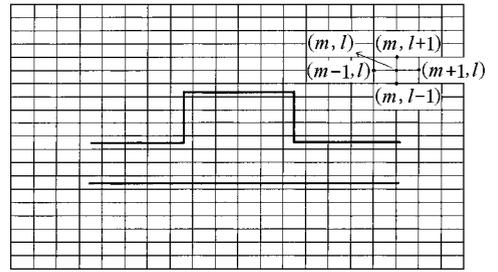


图1 计算窗口的离散

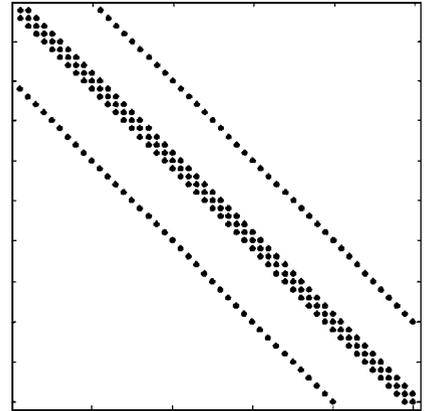


图2 稀疏矩阵 A_{xx} 零元素分布图

3. 算法的具体实现、模拟结果及讨论

3.1. 方程的离散

为了使计算过程执行起来简便,又要考虑极化效应,我们采用了半矢量方法,即忽略(2)式中耦合项 A_{xy} , A_{yx} , 方程变为

$$\frac{\partial^2}{\partial z^2} E_x + A_{xx} E_x = 0, \quad (8)$$

$$\frac{\partial^2}{\partial z^2} E_y + A_{yy} E_y = 0. \tag{9}$$

显然, E_x, E_y 之间没有相互耦合, 可以分别求解. 在下面的过程只针对 E_x 进行. 首先将计算窗口离散化 (N_x, N_y) 分别表示 x 和 y 方向上的离散点数, 如图 1. A_{xx} 的二阶精度的 Crank-Nicholson 离散格式为

$$E_x(m, l) = E_x(m\Delta x, l\Delta y),$$

$$n(m, l) = n(m\Delta x, l\Delta y),$$

$$A_{xx} E_x(m, l) = \frac{2}{(\Delta x)^2} \times \frac{n(m+1, l)^2 E_x(m+1, l) - n(m, l)^2 E_x(m, l)}{n(m+1, l)^2 + n(m, l)^2} - \frac{2}{(\Delta x)^2} \times \frac{n(m, l)^2 E_x(m, l) - n(m-1, l)^2 E_x(m-1, l)}{n(m, l)^2 + n(m-1, l)^2} + \frac{1}{(\Delta y)^2} (E_x(m, l+1) + E_x(m, l-1) - 2E_x(m, l)) + n(m, l)^2 k_0^2 E_x(m, l). \tag{10}$$

从 (10) 式可以看出, A_{xx} 的离散表达式中涉及点 (m, l) 及其上下左右共 5 点, 当离散点先按行后按列进行排列时, 矩阵 A_{xx} 成为在 $(-N_x, -1, 0, 1, N_x)$ 对角线上有非零元素的大型稀疏矩阵, 见图 2. 采用隐式方法对 A_{xx} 求解或者求逆都非常耗费机时, 采用显式方法则避免了这个问题, 所以本文中的算法采用显式方法.

3.2. 最小平方逼近

从矩阵理论可以得到: A_{xx} 特征值 λ_A 只能存在于区间 $(n_{\min}^2 k_0^2 - \frac{4}{\Delta x^2} - \frac{4}{\Delta y^2}, n_{\max}^2 k_0^2)$, 所以最小平方逼近区间应取此区间, 但是, 当 $n_{\min}^2 k_0^2 - \frac{4}{\Delta x^2} - \frac{4}{\Delta y^2}$

< 0 时, $\sqrt{\lambda_A}$ 可能为虚数, 不可能用一个实函数来逼近一个虚函数, 所以我们在逼近时, 区间取为 $(0, n_{\min}^2 k_0^2)$. 至于由此带来的影响, 还有待于进一步研究. 当传输波长 $\lambda = 1.55 \mu\text{m}$, $n_{\max} = 3.45$ 时, 这时逼近区间取为 $(0, 196)$. 从 (6) 式很容易看出权函数 $w(x)$ 的意义: 在最小逼近的情况下, 权函数大的地方, 逼近误差会相对更加重要, 综合考虑后, 该处的误差会相对较小, 所以可以用 $w(x)$ 来控制逼近误差的分布. 对于一般的波导结构, 如图 1 所示的脊形波导, 本征模的传播常数一般都 very 靠近 $n_{\max} k_0$, 所以在逼近区间中, 越靠近上界越显得重要, $w(x)$ 应该是单调上升函数. 在本文的模拟中 $w(x)$ 取如下形式的函数:

$$w(x) = x^H, \quad H \text{ 为正整数.} \tag{11}$$

由于对整个区域进行逼近, 所以最小平方逼近的阶数 M 一般要比泰勒展开和庞德逼近要高, 但是由此带来的计算问题可以由本文中的有效算法得以解决. 有了如上的这些设置, 便可以通过上一节中的方法求得最小平方逼近的展开系数 c_i .

表 1 所列数据为展开阶数 $M = 6$ 时, 不同权函数所对应的展开系数. 图 3 为 $H = 10$ 的曲线和理论值及泰勒展开曲线, 从图 3 可以明显地看出, 平方逼近的效果要比泰勒展开更好. 这种改善可以更明显地在精度曲线图 4 中看出. 图 4 纵轴表示 \sqrt{x} 的实际值与逼近值之间的差别, 可以看到, 当 H 逐渐增大时, 最小平方逼近的误差分布逐渐向左边集中, 而右边 (及 $n_{\max} k_0$ 附近) 的逼近精度不断地提高, 虽然达到固定精度的有效区域在减小, 但是和图中同阶泰勒展开达到有效精度的区域相比还是要大许多, 以 $M = 6, H = 10$, 误差 $< 10^{-4}$ 为例, 最小平方展开达到

表 1 最小平方展开系数表

H	C_0	$C_1/10^{-1}$	$C_2/10^{-4}$	$C_3/10^{-6}$	$C_4/10^{-8}$	$C_5/10^{-11}$	$C_6/10^{-14}$
0	1.0051	2.4615	-47.095	64.075	-48.161	183.47	-276.57
3	1.8671	1.5718	-16.038	14.183	-7.7536	23.076	-28.542
6	2.2133	1.3551	-10.802	7.8388	-3.6423	9.4606	-10.421
7	2.2897	1.3142	-9.9337	6.8928	-3.0829	7.7479	-8.2930e
8	2.3543	1.2813	-9.2608	6.1853	-2.6768	6.5367	-6.8217
10	2.4580	1.2313	-8.2888	5.2049	-2.1339	4.9679	-4.9685
12	2.5377	1.1951	-7.6220	4.5629	-1.7928	4.0174	-3.8820
16	2.6523	1.1463	-6.7689	3.7807	-1.3950	2.9524	-2.7077

要求的区域为[85 ,196],而泰勒展开为[135 ,196],扩大了接近一倍.虽然泰勒展开在接近 $n_{\max} k_0$ 时精度会很高,但在离散精度不可能非常高的情况下,这种过高的精度在实际计算中没有多大用处.从表 1 可以看出系数上的变化, H 增大时,高次项系数的绝对值在减小,而低次项系数的绝对值在增大,逼近函数振荡幅度的减小使得误差分布逐渐向左端靠拢,也就是更加接近泰勒展开.相信在 H 不断增大时,最小平方逼近会不断地向泰勒展开接近.图 5 为 $H = 10$ 时不同 M 对应的精度曲线.显然,当 M 增大时,可以获得更高的逼近,但是 M 增大会同比例地增加计算量,所以一般在和离散精度对应的情况下取 $M = 10, H = 10$ 已经足够,此时在 [50 ,196]上绝对误差达到 10^{-4} ,而在 [95 ,196]上则为 10^{-6} .同理 $P_N(A)$ 中的泰勒展开的阶数也不用过高,一般 $N < 40$ 即可.

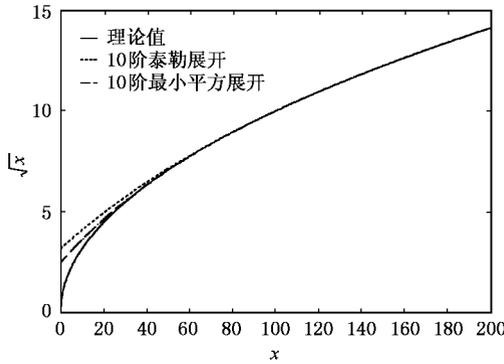


图 3 理论值、泰勒展开和最小平方展开的比较

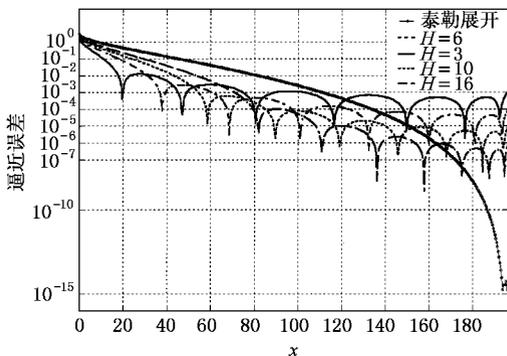


图 4 泰勒展开和不同权函数下最小平方展开的逼近误差比较

3.3. 算法

计算中需要执行下式描述的运算：

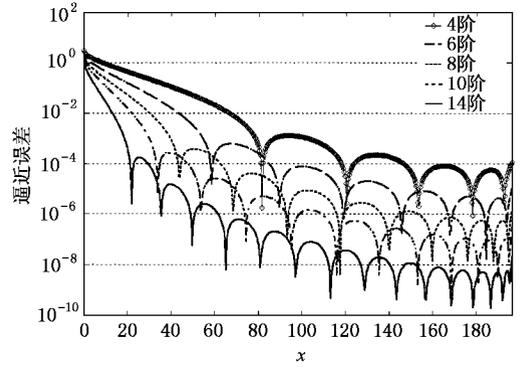


图 5 不同展开阶数时最小平方展开的逼近误差

$$\begin{aligned} \bar{E}(Z_0 + \Delta z) &= \sum_{i=0}^N \frac{(-j\Delta z R_M(A))^i}{i!} \bar{E}(Z_0) \\ &= \sum_{i=0}^N \frac{(-j\Delta z (\sum_{l=0}^M c_l A^l))^i}{i!} \bar{E}(Z_0), \end{aligned} \quad (12)$$

式中 A 为 $(N_x N_y) \times (N_x N_y)$ 稀疏方阵, $\bar{E}(Z_0)$ 为包含 $(N_x N_y)$ 个元素的向量.在文献[5]中,采用先计算 $R_M(A)$,然后代入(12)式中进行循环运算,对低阶光束传播方法($M < 3$),这种算法还可以工作,但是对高阶光束传播方法,这种算法对内存的需求将很大,以 $N_x = N_y = 100$ 为例, A 为 10000×10000 的稀疏矩阵,每一行有 5 个非零元素,对称分布在对角线两边.当 A 作幂次方运算时,所得矩阵每行的非零元素数量将急剧增长,矩阵 A^l 非零元素的数量为 $N_x N_y (2l^2 + 2l + 1)$,当 $l = 10$ 时,以每个元素用 8 个字节存储,在没有考虑中间变量和其他变量的情况下,单独存储 $R_M(A)$ 需要 17.6M 的内存,内存需要量已经很大.本文采用 2 级嵌套的循环算法,只需要对 A 和中间变量进行存储,内存需求量大大减小,大概为 $O(10^5)$ 字节,计算总量没有增加,传输一步的典型计算时间为 5s (Matlab 上运行).

3.4. 边界条件

为了计算方便,我们采用了执行起来最简单但有效的透明边界条件(TBC)^[10],透明边界条件认为在边界处,波的传播和平面波一样,这样向外传播的模式,将不受阻挠地穿透计算窗口,不会返回形成非物理解.在实际计算中,下一步边界处的情况由上一步靠近边界的几个点处决定,反映在矩阵 A 上,只需在计算每一步时,根据上一步的值改变 A 中对应

边界处的 $(N_x + N_y)$ 个元素,几乎没有增加计算量.在实际计算中还应该对边界处两点的比值作动态截断处理,因为当边界处的场很小时,数值误差对解的影响可能很大,相对误差很大,两点之间比值已经完全失真,如果不截去过高的比值,可能导致算法发散.

3.5. 结果与讨论

应用上述方法,本文对几种典型结果进行了模拟.波导为文献[5]中典型的脊形波导结构,参数为:芯层厚度 $H = 6\mu\text{m}$,脊高 $h = 2.5\mu\text{m}$,脊宽 $W = 4\mu\text{m}$, $n_1 = 3.44$, $n_2 = 3.435$, $n_3 = 1$.计算窗口为 $20\mu\text{m} \times 30\mu\text{m}$ 的长方形,模拟时 $N_x = N_y = 100$,对应 $D_x = 0.2\mu\text{m}$, $D_y = 0.3\mu\text{m}$,传播方向上步长 $D_z = 0.4\mu\text{m}$.输

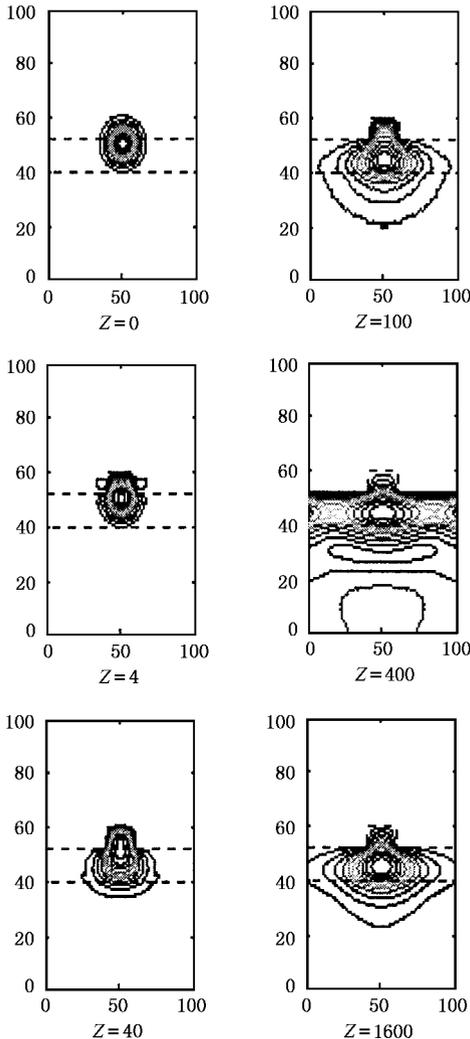


图6 脊形直波导的模拟结果 $n_1 = 3.44$, $n_2 = 3.435$, $n_3 = 1$, $D_z = 0.4\mu\text{m}$, $M = 8$, $N = 3$

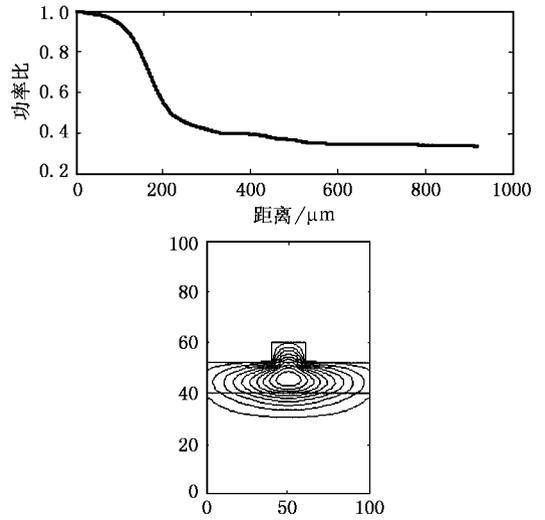


图7 高斯光束进入直波导后的能量曲线和本征模分布 $W = 4\mu\text{m}$, $H = 6\mu\text{m}$, $h = 3.5\mu\text{m}$, 其余参数同图6

入 $1/e$ 半径为 $2\mu\text{m}$ 的高斯光束.图6为脊形直波导的模拟结果.从图6可以明显地看到光场的逐步变化、辐射的产生、能量的外逸、单模的形成等整个波传播过程.开始时,处在波导芯区外的入射光场很快地溢出计算窗口.处在芯区内的光场逐渐横向扩散,光斑尺寸变大,受制于空气-芯区的强折射率变化,辐射模携带的能量主要向两侧和下限制层中扩散,传播距离为 $100\mu\text{m}$ 时,已有相当一部分能量扩散到了下限制层中.当传播距离达到 $400\mu\text{m}$ 时,两侧和向下的辐射模清晰可见,它们携带着相当一部分能量向外扩散.传播距离达到 $1600\mu\text{m}$ 时,辐射模基本上都逃逸出计算窗口,光场趋于稳定.上述过程的出现说明本文的方法可以正确地模拟光传播过程,并能较好地模拟辐射模的传播过程,与前面的分析一致.图7为波传播过程中计算窗口中的能量和初始能量的比值及通过虚轴传播法^[11]得到的本征单模场分布.计算得到的模式传播常数为 $3.3890k_0$ 和文献[11]中求得的传播常数差别在 10^{-4} 量级.从能量曲线中可以看到,当传播距离超过 $1500\mu\text{m}$ 时,波导中的总能量趋于稳定,与图6中看到的一致.最后留在波导中的能量比例较小,原因是高斯光束和本征模的失配很大,大部分能量激发辐射模损失掉了,这与波导结构、输入光场的中心位置、分布有很大关系,通过调整输入光场的中心位置,可以使这种模失配式导致的能量损失降到最小.图8为脊型多模波导的模拟,由于上下的限制都很强,所以辐射模只能从两侧的限制通道中溢出.从图8可以很清楚地看

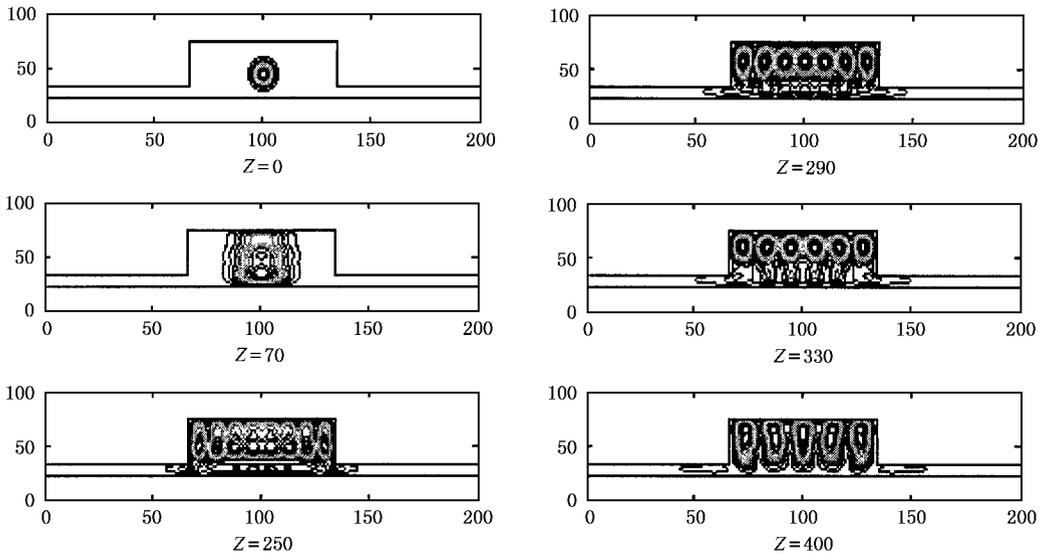


图 8 脊形多模波导的模拟结果

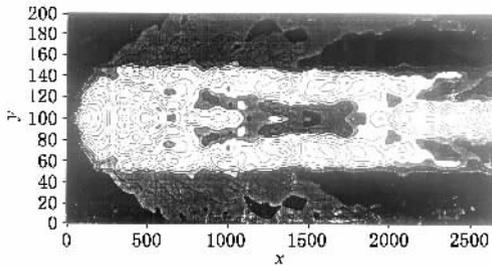


图 9 多模调形波导模拟结果在 XZ 方向上的投影

到多模波导的自映像现象,说明本方法对多模波导的模拟也可以获得较好的结果.用这种方法可以精确确定三维多模波导的自映像长度,给多模波导器件的设计制作提供帮助.图 9 为多模波导模拟的顶

视图,可以看到明显的自映像和辐射模向外溢出的过程.采用此方法可为多模波导干涉器的设计提供理论分析.

4. 结 论

本文实现了一种新的基于最小平方逼近的高阶有限差分束传播方法,由于采用了最小平方逼近,在整个解区间上较好地逼近传播算子,避免了传统方法中参考折射率的选择问题,从而改善了传统光束传播方法对辐射模、大角度模拟和多模模拟的精度.文中还应用该方法对几种波导结构进行了数值模拟,模拟结果为该方法的正确性提供了有力的证据.

[1] Liu S D , Fu Z T , Liu S K and Zhao Q 2002 *Acta Phys. Sin.* **51** 718 [in Chinese] 刘式达、傅遵涛、刘式适、赵强 2002 *物理学报* **51** 718]

[2] Feit M D and Fleck J A Jr 1978 *Appl. Opt.* **17** 3990

[3] Kim C M and Ramaswamyeling R V 1989 *J. Lightwave Technol.* **7** 1581

[4] Lee P C , Schulz D and Voges E 1992 *J. Lightwave Technol.* **10** 1832

[5] Lee P C and Voges E 1994 *J. Lightwave Technol.* **12** 215

[6] Hsueh Y L , Yang M C and Chang H C 1999 *J. Lightwave Technol.* **17** 2389

[7] Fogli F , Bellanca G , Bassi P , Madden I and Johnstone W 1999 *J. Lightwave Technol.* **17** 136

[8] Liu X Y 2000 *Acta Phys. Sin.* **49** 186 [in Chinese] [刘新芽 2000 *物理学报* **49** 186]

[9] Hadley G R 1992 *Opt. Lett.* **17** 1426

[10] Hadley G R 1992 *IEEE J. Quantum Electron.* **28** 363

[11] Scarmozzino R , Gopinath A and Helfert S 2000 *IEEE J. Selected Topics Quantum Electron.* **6** 150



A new beam propagation method based on least-squares expansion approximating^{*}

Xia Jin-Song Yu Jin-Zhong

(*State Key Laboratory on Integrated Optics , Institute of Semiconductors , Chinese Academy of Sciences ,Beijing 100083 ,China*)

(Received 2 July 2002 ; revised manuscript received 16 August 2002)

Abstract

A new finite difference wide-angle beam propagation method is developed by introducing the least-squares expansion approximant in the propagator expansion. In this new method it is not necessary to select the reference index point because of the whole region approaching the least-square expansion. This method avoids the problems induced by error selection of the reference index in the old methods based on Taylor or Padè expansion. Several typical structures are simulated by the new method and the results prove the validity of it.

Keywords : beam propagation method , finite difference , integrated optics , numeric method

PACC : 0340K , 0260 , 4280L

^{*} Project supported by the National Natural Science Foundation of China(Grant Nos. 69896260 and 69990540) , and the State Key Development Program for Basic Research of China(Grant No. G2000-03-66).