

VarRCWA：一种自适应高阶严格耦合波分析方法

朱子伟¹ 和 郑昌希^{*1}

¹哥伦比亚大学计算机科学系，美国纽约州纽约市 10027

2022年2月1日

Abstract

半解析方法，如严格耦合波分析，对于光子结构的数值分析至关重要。与其他方法相比，它们提供了更快的计算速度，尤其适用于具有恒定横截面形状的结构（如超表面单元）。然而，当横截面形状即使轻微变化（如锥形）时，现有的半解析方法会面临高计算成本。我们表明，现有方法可以被视为相对于结构横截面变化的零阶近似。我们则推导出相对于横截面变化的高阶微扰展开。基于这一展开，我们提出了一种新的半解析方法，即便在横截面形状变化较大的情况下，也能实现快速计算。此外，我们设计了一种算法，可以自动离散化结构，从而在实现用户指定精度水平的同时，减少计算量

1 Introduction

数值模拟是理解光子结构的基本工具。在众多流行的方法中，半解析方法，例如严格耦合波分析（RCWA）[15]，已被广泛用于分析如超表面 [6]、光栅 [14] 和波导 [24] 等器件。与其他方法相比，如有限差分时间域（FDTD）方法，半解析方法通常具有更低的计算成本。

这一优势源于半解析方法对麦克斯韦方程的离散化方式。与其他方法（例如 FDTD 方法）完全离散化空间域（即在所有三个维度上）[23] 相比，半解析方法仅部分离散化空间域（例如，仅在 x 维和 y 维，而不是 z 维）。这是可能的，因为许多光子结构具有主要的光传播方向（本文称为 z 方向；见图 2）。在某些情况下，沿光传播方向，结构的截面形状保持不变（例如，一个超表面单元）。因此，我们不必沿 z 方向离散化结构；相反，结构中的光传播可以视为各个传播模式经历相位变化的叠加。这是使半解析方法能够降低计算成本的基本观点（见 2.1）。

*cxz@cs.columbia.edu

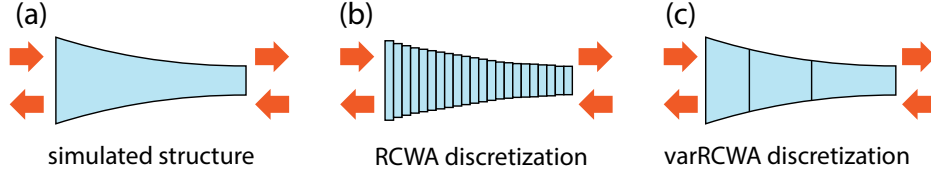


图 1: **Discretization in conventional and our method.** 提供了一个具有不同横截面形状的光子结构 (a), 传统的半解析方法 (如 RCWA) 必须沿波传播方向将结构离散化为许多薄片段 (b), 导致计算成本高。(c) 我们的方法通过使用高阶扰动展开, 自适应地将结构离散为更少的片段。箭头表示光波的入射和出射方向。

然而, 对于许多光子结构来说, 这种观点是不成立的, 因为沿着主要光传播方向, 结构的横截面会发生变化 [13, 17]。一个常见的例子是光子波导 (例如锥形波导; 见图 1-a)。要使用半解析方法模拟这些光子结构, 就必须沿 z 方向将结构进一步离散为一系列薄层 [9, 11] (见图 1-b)。在每一层中, 假设横截面形状保持不变, 从而可以使用半解析方法对该层进行模拟。然而, 这种方法需要大量的层, 这反过来会破坏半解析方法的计算优势。除了计算成本之外, 通常不清楚需要多少个离散层才能达到某种精度。在实际操作中, 需要依靠反复试验来选择适当的分辨率以确保足够的精度。通常, 为了获得令人满意的结果, 需要多次运行模拟。

在这项工作中, 我们克服了这些限制。我们的方法不需要反复试验, 因此使用起来更简单: 只需提供一个光子结构和用户指定的精度水平 (例如一个实数), 我们的方法就会自动决定如何在 z 方向上离散化结构, 旨在在达到所需精度的同时降低总体计算成本。要获得用户指定精度的模拟结果, 只需运行一次即可。

为此, 我们的核心开发有两个方面: 1) 我们表明传统的半解析方法 (如 RCWA) 仅是关于结构横截面变化的零阶近似。通过一种新颖的变量变换, 我们提出了一种高阶半解析方法, 该方法允许结构的横截面在 z 方向上变化, 而无需将其离散成薄片。2) 借助这种高阶方法, 我们引入了一种算法, 能够自动且自适应地对结构进行离散化, 以达到用户指定的精度水平。对于横截面在 z 方向变化快速的区域, 我们的算法会以高分辨率切分结构以确保模拟精度; 对于横截面变化较小的区域, 则会粗略离散以节省计算成本。

我们使用我们的方法来分析各种光子结构, 并将其与传统的半解析方法 (如 RCWA) 进行比较。我们展示了我们的方法作为一种高阶方法, 确实收敛更快。因此, 为了获得相同的精度水平, 我们的方法需要的计算时间要少得多, 而且完全不需要调整分辨率。

2 Method

我们现在介绍我们的核心开发内容。为了理解我们开发背后的原理，我们首先简要回顾广泛使用的半解析方法的局限性。

2.1 Established Semi-analytical Methods and Their Limitations

半解析方法将麦克斯韦方程在 x 和 y 方向上进行空间离散，但在 z 方向上不离散。在频率域中，麦克斯韦方程变为

$$\frac{\partial \mathbf{e}}{\partial z} = \frac{j}{k_0} \mathbf{P} \mathbf{h} \quad \text{and} \quad \frac{\partial \mathbf{h}}{\partial z} = \frac{j}{k_0} \mathbf{Q} \mathbf{e}, \quad (1)$$

其中向量 \mathbf{e} 和 \mathbf{h} 是 xy 平面上电场和磁场的离散表示，位于 z 位置，而矩阵 \mathbf{P} 和 \mathbf{Q} 编码了 xy 平面上材料的磁导率和介电常数在相同 z 位置的分布。根据 \mathbf{e} 和 \mathbf{h} 的具体表示方式，会出现不同的半解析方法。最广泛使用的方法（例如，用于超表面[6]、光栅[14]和波导[24]的分析）是严格耦合波分析（RCWA）方法，其中 \mathbf{e} 和 \mathbf{h} 在 xy 平面上使用二维傅里叶基进行离散化。在本文中，我们的发展可以应用于一般的半解析方法（例如线法[18]），尽管我们的实现和数值实验特别集中于 RCWA 方法。

当结构的横截面形状沿 z 方向固定时，(1) 式中的 \mathbf{P} 和 \mathbf{Q} 都是常矩阵，并且 (1) 式可以通过特征值分解来求解，即 $\mathbf{PQ} = \mathbf{W} \mathbf{\Lambda}^2 \mathbf{W}^{-1}$ 。由此得到的 \mathbf{W} 和 $\mathbf{\Lambda}$ 使我们能够将 (1) 式的解表示为

$$\mathbf{e} = \frac{1}{2} \mathbf{W} (e^{\mathbf{\Lambda} z} \mathbf{a}_L + e^{-\mathbf{\Lambda} z} \mathbf{b}_L) \quad \text{and} \quad \mathbf{h} = \frac{1}{2} \mathbf{V} (e^{\mathbf{\Lambda} z} \mathbf{a}_L - e^{-\mathbf{\Lambda} z} \mathbf{b}_L). \quad (2)$$

这里 \mathbf{V} 是描述横截面磁场的基底，与特征向量 \mathbf{W} （相关，即通过 $\mathbf{V} = \mathbf{QW} \mathbf{\Lambda}^{-1}$ 得到电场的基底）； \mathbf{a}_L 和 \mathbf{b}_L 是在结构左端堆叠前向和后向光波系数的向量（见图 2）。此外，依据方程 (2)，我们可以定义该结构的 *scattering matrix*，它将光波通过结构传播后的输出状态与其输入状态联系起来，即

$$\begin{bmatrix} \mathbf{a}_R \\ \mathbf{b}_L \end{bmatrix} = \mathbf{S} \begin{bmatrix} \mathbf{a}_L \\ \mathbf{b}_R \end{bmatrix}, \quad \text{where} \quad \mathbf{S} = \begin{bmatrix} \mathbf{T}_{LR} & \mathbf{R}_R \\ \mathbf{R}_L & \mathbf{T}_{RL} \end{bmatrix}, \quad (3)$$

其中，与 \mathbf{a}_L 和 \mathbf{b}_L 相对应， \mathbf{a}_R 和 \mathbf{b}_R 描述了右端的向前波和向后波（参见图 2）。一旦已知散射矩阵，就可以直接计算结构的光学性能（例如波导的模式转换效率和相位变化）。

当结构的横截面形状沿 z 方向变化时，必须将结构划分为一系列小截面，以便每个截面都可以近似为具有固定横截面（图 1-b）。然后通过上述过程分析每个截面 i ，从中可以计算其散射矩阵 \mathbf{S}_i 。最后，通过使用 Redheffer 星形乘积 [19] 组合所有散射矩阵，可以得到整个结构的散射矩阵。

Limitations. 在半解析方法中，假设截面固定可以被视为结构的零阶近似（如第 2.2 节所述）。为了达到足够的精度，这种粗略的近似必须通过较小的结构长度来补偿。因此，即使是轻微的截面变化也需要大量离散截面（如图 1-b 所示）。对于每个截面，都需要进行特征值分解，因此其计算成本很高。

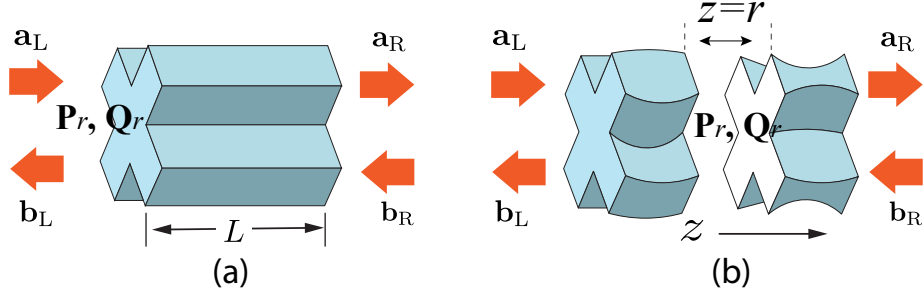


图 2: **Notation.** 提供了一个光子结构截面，我们在四个部分建模光波传播：左端和右端的入射波与出射波，如橙色箭头所示。(a) 截面具有恒定的横截面形状，其材料矩阵 P_r 和 Q_r 是固定的。(b) 当截面横截面形状变化时，我们选择一个特定位置 $z = r$ (，称为参考位置)，并使用 $z = r$ 处的材料矩阵 P_r 和 Q_r 构建描述前向波和后向波的基底。我们的方法将这样的截面视为 (a) 中恒定横截面截面的微扰，并利用微扰展开进行数值模拟。为了可视化，这里我们在 $z = r$ 处切开截面，以显示 $z = r$ 处的横截面，该横截面可能与其两端的横截面形状不同。

进一步按总段数进行缩放。为了减少计算成本，我们需要在保持模拟精度的同时减少离散段的总数（图 1-c）。这促使我们寻求一种高阶半解析方法，该方法考虑了长段中的横截面变化，从而减少总段数。

2.2 High-order Semi-analytical Methods

考虑沿 z 方向的光子结构截面。假设它的横截面形状有变化，也就是说，在方程 (1) 中， P 和 Q 不是恒定的矩阵；它们会随 z 改变。在这种情况下，(1) 的解不会像 (2) 那样简单。现在，我们的目标是将 (1) 的解表示为关于横截面变化的微扰展开（图 2），这个展开将作为我们方法的核心数值方案。

Naïve solution. 为了解我们发展的见解，我们从解的一个天真的（但不切实际的）展开形式开始。首先，受公式(2)的启发，我们使用一组基向量 \mathbf{W} 和 \mathbf{V} 来分别描述横截面的电场和磁场——在横截面变化的情况下， \mathbf{W} 和 \mathbf{V} 的具体选择将在稍后说明。横截面的电场和磁场 \mathbf{e} 和 \mathbf{h} 由在结构中对向前和向后传播的光波形成，其关系为：

$$\mathbf{a}(z) = \mathbf{W}^{-1}\mathbf{e}(z) + \mathbf{V}^{-1}\mathbf{h}(z) \text{ and } \mathbf{b}(z) = \mathbf{W}^{-1}\mathbf{e}(z) - \mathbf{V}^{-1}\mathbf{h}(z), \quad (4)$$

其中 \mathbf{a} 和 \mathbf{b} 分别是在所选基底中描述前向波和后向波的系数，它们在 z 上变化。接下来，为了建立 \mathbf{a} (和 \mathbf{b}) 的微分方程，我们对 (4) 的两边进行微分，然后使用麦克斯韦方程组 (1)，我们得到

$$\frac{\partial \mathbf{a}}{\partial z} = \mathbf{W}^{-1} \frac{\partial \mathbf{e}}{\partial z} + \mathbf{V}^{-1} \frac{\partial \mathbf{h}}{\partial z} = \frac{j}{2k_0} [\mathbf{W}^{-1} \mathbf{P} \mathbf{V} (\mathbf{a} - \mathbf{b}) + \mathbf{V}^{-1} \mathbf{Q} \mathbf{W} (\mathbf{a} + \mathbf{b})]. \quad (5)$$

Equivalently, we have the integral equation

$$\mathbf{a}(z) = \mathbf{a}(z_L) + \frac{j}{2k_0} \int_{z_L}^z [(\mathbf{W}^{-1}\mathbf{P}\mathbf{V} + \mathbf{V}^{-1}\mathbf{Q}\mathbf{W}) \mathbf{a}(z') + (\mathbf{V}^{-1}\mathbf{Q}\mathbf{W} - \mathbf{W}^{-1}\mathbf{P}\mathbf{V}) \mathbf{b}(z')] dz', \quad (6)$$

其中 z_L 是所考虑的结构截面的起始 z 位置。还要注意, \mathbf{P} 和 \mathbf{Q} 都在 z 上变化, 并且可以对 $\mathbf{b}(z)$ 得到类似的积分方程。理论上, 方程(6) 允许我们将 $\mathbf{a}(z)$ 表示为扰动展开。这是通过递归地将积分中的 $\mathbf{a}(z')$ 用方程(6) 本身替代到一定阶数来实现的 ($\mathbf{b}(z)$ 也类似)。例如, 要获得一阶展开, 可以在(6)中将 $\mathbf{a}(z')$ 和 $\mathbf{b}(z')$ 替换为它们的零阶近似 $\mathbf{a}_L \approx \mathbf{a}(z_L)$ 和 $\mathbf{b}_R \approx \mathbf{b}(z_R)$ 。

为了使用该展开式分析较长的部分 (从而减少总的部分数量), $\mathbf{W}^{-1}\mathbf{P}\mathbf{V}$ 和 $\mathbf{V}^{-1}\mathbf{Q}\mathbf{W}$ 的范数必须足够小——对此要求的直观解释见补充材料 1。然而, 这一要求在实际中几乎难以满足, 因为 \mathbf{P} 和 \mathbf{Q} 都依赖于横截面材料分布, 其范数可能变得任意大。尽管如此, 该展开式的发展仍激发了一种可行的策略: 为了获得稳定的微扰展开, 我们需要避免在类似公式 (6) 的积分方程中使用 \mathbf{P} 和 \mathbf{Q} ; 相反, 我们寻求一种仅涉及 \mathbf{P} 和 \mathbf{Q} 在 z 上变化的展开式。

Preconditioned solution. 我们提出的扩展从变量变换开始。我们引入两个变量:

$$\tilde{\mathbf{a}} = e^{-\frac{j}{k_0}\Lambda z} \mathbf{a} \quad \text{and} \quad \tilde{\mathbf{b}} = e^{\frac{j}{k_0}\Lambda z} \mathbf{b}, \quad (7)$$

Λ 是通过特征分解 $\mathbf{P}_r \mathbf{Q}_r = \mathbf{W} \Lambda^2 \mathbf{W}^{-1}$ 得到的特征值矩阵。在这里, \mathbf{P}_r 和 \mathbf{Q}_r 是固定矩阵, 用于编码在特定 $z = r$ 位置的材料渗透率和介电常数的分布。理想情况下, $z = r$ 处的截面被选择以代表整个截面的“平均”截面, 从而可用于构建基向量 \mathbf{W} 和 \mathbf{V} 。因此, 我们将该位置称为截面的 *reference position* (见图 2-b)。虽然理论上可以选择任意 r 位置, 但在实际操作中, 我们仅使用截面的中点。通过 $\mathbf{V} = \mathbf{Q}\mathbf{W}\Lambda^{-1}$ 得到的 \mathbf{W} (和 \mathbf{V} 用作描述前向波和后向波的基础 (回顾公式 (4))。

这种变量的变化是引入 \mathbf{P} 和 \mathbf{Q} 在类似于 (6) 的积分方程中变动的关键。通过对 (7) 进行微分并使用方程 (1), 我们得到以下微分方程 (推导见补充材料 2) :

$$\begin{aligned} \frac{\partial \tilde{\mathbf{a}}}{\partial z} &= \frac{j}{2k_0} e^{-\frac{j}{k_0}\Lambda z} \delta \mathbf{A} e^{\frac{j}{k_0}\Lambda z} \tilde{\mathbf{a}} - \frac{j}{2k_0} e^{-\frac{j}{k_0}\Lambda z} \delta \mathbf{B} e^{\frac{j}{k_0}\Lambda z} \tilde{\mathbf{b}}, \\ \frac{\partial \tilde{\mathbf{b}}}{\partial z} &= \frac{j}{2k_0} e^{\frac{j}{k_0}\Lambda z} \delta \mathbf{B} e^{-\frac{j}{k_0}\Lambda z} \tilde{\mathbf{a}} - \frac{j}{2k_0} e^{\frac{j}{k_0}\Lambda z} \delta \mathbf{A} e^{-\frac{j}{k_0}\Lambda z} \tilde{\mathbf{b}}, \end{aligned} \quad (8)$$

其中 $\delta \mathbf{A}$ 和 $\delta \mathbf{B}$ 与沿 z 方向的材料变化有关, 即,

$$\begin{aligned} \delta \mathbf{A}(z) &= \mathbf{W}^{-1} (\mathbf{P}(z) - \mathbf{P}_r) \mathbf{V} + \mathbf{V}^{-1} (\mathbf{Q}(z) - \mathbf{Q}_r) \mathbf{W}, \\ \delta \mathbf{B}(z) &= \mathbf{W}^{-1} (\mathbf{P}(z) - \mathbf{P}_r) \mathbf{V} - \mathbf{V}^{-1} (\mathbf{Q}(z) - \mathbf{Q}_r) \mathbf{W}. \end{aligned} \quad (9)$$

接下来, 我们将方程 (8) 重写为积分形式, 并使用方程 (7) 替换 $\tilde{\mathbf{a}}$ 和 $\tilde{\mathbf{b}}$ 。这将得到一个

关于 $\mathbf{a}(z)$ 和 $\mathbf{b}(z)$ 的一组新的积分方程，它们与 (6) 不同：

$$\mathbf{a}(z) = e^{\frac{j}{k_0}\Lambda(z-z_L)}\mathbf{a}_L + \frac{j}{2k_0} \int_{z_L}^z e^{\frac{j}{k_0}\Lambda(z-z')}\delta\mathbf{A}(z')\mathbf{a}(z')dz' - \frac{j}{2k_0} \int_{z_L}^z e^{\frac{j}{k_0}\Lambda(z-z')}\delta\mathbf{B}(z')\mathbf{b}(z')dz' \quad (10)$$

$$\mathbf{b}(z) = e^{\frac{j}{k_0}\Lambda(z_R-z)}\mathbf{b}_R - \frac{j}{2k_0} \int_z^{z_R} e^{\frac{j}{k_0}\Lambda(z'-z)}\delta\mathbf{B}(z')\mathbf{a}(z')dz' + \frac{j}{2k_0} \int_z^{z_R} e^{\frac{j}{k_0}\Lambda(z'-z)}\delta\mathbf{A}(z')\mathbf{b}(z')dz'. \quad (11)$$

请注意，这里的两个方程是从不同方向进行积分的。在公式 (10) 中，我们使用截面左端的 \mathbf{a}_L 作为初始值，积分沿正向进行。相反，公式 (11) 沿反向积分，使用 \mathbf{b}_R 作为初始值。通过这种方式，所有的传播相位项（即 $e^{j\Lambda\Delta z/k_0}$ ）都变得小于 1，从而这些积分的数值计算保持稳定。

现在，我们可以构建出射波 \mathbf{a}_R 和 \mathbf{b}_L 的微扰展开式。类似于上述天真的展开式的构建方法，我们将积分项中的 $\mathbf{a}(z')$ 和 $\mathbf{b}(z')$ 代入方程 (10) 和 (11) 本身，并且这种代入递归进行到某一阶。这些展开式为分析结构与传播波的相互作用提供了一种数值方法。例如，已知输入到该截面的波，包括左端的正向波（由 \mathbf{a}_L 描述；见图 2）和右端的反向波（由 \mathbf{b}_R 描述），可以使用以下一阶展开式估算右端和左端的出射波：

$$\begin{aligned} \mathbf{a}_R &\approx e^{\frac{j}{k_0}\Lambda(z_R-z_L)}\mathbf{a}_L + \frac{j}{2k_0} \int_{z_L}^{z_R} e^{\frac{j}{k_0}\Lambda(z_R-z')}\delta\mathbf{A}(z')e^{\frac{j}{k_0}\Lambda(z'-z_L)}dz'\mathbf{a}_L \\ &\quad - \frac{j}{2k_0} \int_{z_L}^{z_R} e^{\frac{j}{k_0}\Lambda(z_R-z')}\delta\mathbf{B}(z')e^{\frac{j}{k_0}\Lambda(z_R-z')}dz'\mathbf{b}_R, \end{aligned} \quad (12)$$

$$\begin{aligned} \mathbf{b}_L &\approx e^{\frac{j}{k_0}\Lambda(z_R-z_L)}\mathbf{b}_R + \frac{j}{2k_0} \int_{z_L}^{z_R} e^{\frac{j}{k_0}\Lambda(z'-z_L)}\delta\mathbf{A}(z')e^{\frac{j}{k_0}\Lambda(z_R-z')}dz'\mathbf{b}_R \\ &\quad - \frac{j}{2k_0} \int_{z_L}^{z_R} e^{\frac{j}{k_0}\Lambda(z'-z_L)}\delta\mathbf{B}(z')e^{\frac{j}{k_0}\Lambda(z'-z_L)}dz'\mathbf{a}_L. \end{aligned} \quad (13)$$

值得注意的是，这些展开式并不涉及 $\mathbf{W}^{-1}\mathbf{P}\mathbf{V}$ 或 $\mathbf{V}^{-1}\mathbf{Q}\mathbf{W}$ （与 (6) 不同）。相反，它们的被积函数依赖于 $\delta\mathbf{A}$ 和 $\delta\mathbf{B}$ 。因此，方程 (12) 和 (13) 可以被视为出射波的一阶扰动解。如果截面保持不变， $\delta\mathbf{A} = \delta\mathbf{B} = \mathbf{0}$ ，并且方程 (12) 和 (13) 中的积分项消失。在这种情况下，这些表达式与传统半解析方法完全相同，表明在存在截面变化的情况下，传统的半解析方法是零阶扰动。如果截面在 z 上缓慢变化， $\delta\mathbf{A}$ 和 $\delta\mathbf{B}$ 的范数较小，因此即使对于较长的截面，一阶扰动也会收敛。这与基于方程 (6) 的简单展开形成鲜明对比，因为它们对 \mathbf{P} 和 \mathbf{Q} 的依赖极大限制了截面长度。如果截面变化快速， $\delta\mathbf{A}$ 和 $\delta\mathbf{B}$ 的范数可能会很大，因此必须使用较短的截面长度以确保一致性。

Discussion: alternative approaches. 麦克斯韦方程 (1) 也可以看作是一个初值问题，给定在 $z = z_L$ 的电场和磁场。这个初值问题的解可以表示为涉及矩阵指数的乘积积分 [7, 21]：

$$\begin{bmatrix} \mathbf{e}(z_R) \\ \mathbf{h}(z_R) \end{bmatrix} = \lim_{N \rightarrow \infty} \prod_{i=N}^1 \exp \left(\frac{j(z_i - z_{i-1})}{k_0} \begin{bmatrix} \mathbf{0} & \mathbf{P}_r(z_{i-1} \rightarrow z_i) \\ \mathbf{Q}(z_{i-1} \rightarrow z_i) & \mathbf{0} \end{bmatrix} \right) \begin{bmatrix} \mathbf{e}(z_L) \\ \mathbf{h}(z_L) \end{bmatrix}. \quad (14)$$

在这个表达式中，结构也被离散化为 z_i 处的一系列截面，其中 $z_L = z_0 < z_1 < \dots < z_N = z_R$ ；而 P_r 和 Q_r 是在每个截面的参考位置处评估的渗透率和介电常数矩阵。

然而，对式 (14) 进行数值求解相当具有挑战性。已有的研究使用矩阵指数的低阶泰勒展开来求解式 (14) [10, 20]。但要使用这种展开，区段长度必须非常短（即，约为0.1个波长），且需要大量区段。

另一种方法是将矩阵指数的乘积转换为矩阵求和的指数，类似于标量值 x_i 的 $\prod \exp(x_i) = \exp(\sum x_i)$ 。然而，这并不简单，因为矩阵乘法是不可交换的。因此，人们使用 Magnus 展开式 [1, 2, 8] 和 Fer 展开式 [22] 来修正矩阵求和的指数的使用。在这方面，RCWA 可以被看作 Magnus 展开的一个近似，如 [4] 所示。虽然在 [4] 中引入了 Magnus 展开的高阶近似，但即使横截面变化较小，它仍然要求每个离散部分足够短。

这一类方法的一个更根本的问题源于对初值问题的处理。在光子结构中传播的光几乎总是会发生散射，产生向前和向后的传播。但是将其视为初值问题意味着只考虑单一方向，从而导致数值积分不稳定。正因为这个原因，我们的方法像传统的RCWA一样，将光波分为向前和向后的分量，并沿两个独立方向进行积分（回想公式(10)和(11)）。

2.3 Numerical Implementation with Adaptive Discretization

Numerical integration. 方程 (12) 和 (13) 中的积分可以使用求积法进行数值计算。在实际操作中，我们使用的求积法在 z_L 、 $\frac{z_L+z_R}{2}$ 和 z_R 三个位置进行采样，并通过对采样位置处的被积函数数值进行加权求和来估计积分值。该求积法仅涉及矩阵乘法，因此计算成本较低；在实际操作中，我们在图形处理单元 (GPU) 上实现该方法。因此，在单个截面上计算波传播（使用 (12) 和 (13)）的主要成本来自于参考位置处计算 W 、 Λ 和 V 的特征分解。对于单个结构截面，传统半解析方法也需要进行特征分解。我们的微扰展开在增加少量数值积分的成本下精度更高。但在模拟整个光子结构时，我们的展开方法可以实现自适应离散化。

Adaptive discretization. 我们现在考虑整个光子结构的仿真，而不仅仅是单个部分。我们提出了一种算法，该算法能够在 z 方向上自适应地离散化结构并模拟波传播。输入到我们算法的是一个光子结构和一个期望的精度水平（即一个实数）。这里的目标是在实现所需精度的同时降低整体计算成本。

我们提出的算法在算法1中进行了概述。该算法的关键思想如下：从整个结构作为一个单独的部分开始，它递归地将结构部分细分为 M 个子部分（算法1第5行）。当当前部分的估计模拟误差大于用户指定的精度水平 α （算法1第2行）时，就会进行细分。之后，所有子部分将被单独模拟，并且它们可能以递归方式进一步细分（第9行）。对于每个子部分，我们计算其如何

Algorithm 1: Adaptive-Variant-RCWA($z_{\min}, z_{\max}, \alpha$)

Input: Simulated region along z axis: $[z_{\min}, z_{\max}]$, error bound α
Output: Scattering matrix from z_{\min} to z_{\max}
Data: Simulated Geometry \mathcal{G}

- 1 Evaluate high order scattering matrix S according to Supplement 3 by sampling on \mathcal{G}
- 2 **if** *estimated error (see Supplement 6)* $< \alpha$ **then**
- 3 **return** S
- 4 **else**
- 5 Subdivide this section into M subsections $(z_0, z_1), (z_1, z_2), \dots, (z_{M-1}, z_M)$
- 6 where $z_{\min} = z_0 < z_1 < \dots < z_M = z_{\max}$
- 7 Find a reference point for each subsection.
- 8 **for** $i = 1, \dots, M$ **do**
- 9 scattering matrix $S_i(z_{i-1} \rightarrow z_i) \leftarrow \text{Adaptive-Variant-RCWA}(z_{i-1}, z_i, \alpha)$
- 10 Project the basis of S_i to match the basis of S_{i-1} (see Supplement 4)
- 11 **return** $S \leftarrow$ Redheffer star product of S_1, \dots, S_M .

波使用 (12) 和 (13) 传播, 并进一步计算本小节的散射矩阵 (第1行; 详细公式见补充材料3)。最后, 各小节的散射矩阵通过Redheffer星积[19]组合, 形成父部分的散射矩阵 (第11行)。

我们通过计算连续阶次的渐近展开之间的差异 (例如, 零阶与一阶之间) 来估计误差, 详见补充材料6——这种误差估计方法在其他渐近展开中也有使用 [3, 5]。在实际操作中, 当一个区段被细分时, 我们将其细分为三个子区段 (即, $M = 3$)。通过这种方式, 父区段所需的特征分解可以被重用于模拟中央子区段 (见图3-b)。另一种方法是使用二分细分 (即, $M = 2$), 我们将在 3.2 中的数值实验中进行比较。

通过递归细分, 在 z 方向上自然出现自适应离散: 横截面变化较快的区域会被更多地细分, 从而算法自动使用更精细的部分以确保精度; 同时, 在变化平缓的区域, 它使用低分辨率以节省计算量。

3 Results

我们进行数值实验来验证我们方法的准确性, 并将其性能与传统半解析方法进行比较。由于半解析方法的实现方式不同 (取决于(1)中的 e 和 h 的表示方式), 我们采用最广泛使用的表示方法, 即 RCWA 方法 [15]。

我们使用 C++ 编程语言实现了我们的方法和传统的 RCWA 方法。两种方法中的数值计算, 例如矩阵乘法和特征值分解, 都可以从并行计算中受益。因此, 我们利用 Nvidia 图形处理单元 (GPU) 上的 CUDA [16] 来加速两种方法的计算。

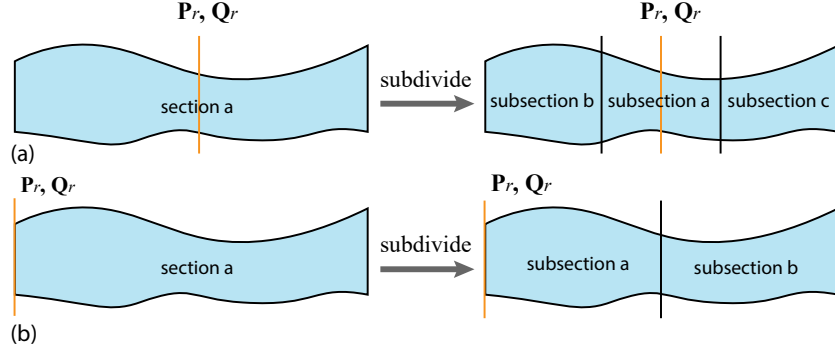


图 3: (a) 我们选择每个部分的中点作为其参考点来评估 P_r 和 Q_r 。在对子部分进行细分时, 我们将其均匀地细分为三个子部分。通过这种方式, 中间子部分的中点与父部分的中点相同, 因此可以重复使用 P_r 和 Q_r 。(b) 我们还尝试将一个部分均匀地细分为两个子部分 (即 $M = 2$)。当我们使用部分的端点作为其参考点时, 这种策略是合适的。

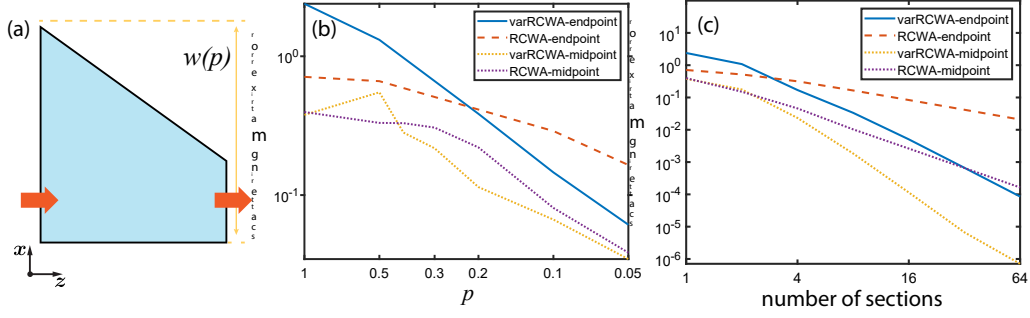


图 4: (a) 我们使用 VarRCWA 分析一个横截面形状线性变化的波导。波导的长度为 $1 \mu\text{m}$ 。其右端的宽度固定为 $2.6 \mu\text{m}$, 我们使用参数 p 来指定左端的宽度。当 $p = 1$ 时, 左端宽度为 $3.7 \mu\text{m}$; 当 $p = 0$ 时, 左端宽度与右端相同 (即 $2.6 \mu\text{m}$)。 (b) 我们模拟该波导而不进行 z 方向的离散化 (即将其视为单一段落), 并测量随 p 变化的散射矩阵误差。 (c) 然后我们增加 z 方向的分辨率, 并测量随 p 变化的散射矩阵误差。在图例中, “中点” 表示使用中点位置作为参考位置来构建 P_r 和 Q_r (参见图 3-a), 而 “端点” 表示使用段落的末端位置作为参考位置 (参见图 3-b)。有关基础值, 请参见 Data File 1-2。

3.1 Validation

为了验证我们方法的准确性, 我们考虑了如图4-a所示的梯形波导。该波导厚度固定为 220nm , 我们使用参数 p 来控制其从左端到右端的横截面变化 (见图4说明)。

首先, 我们验证我们的方法比传统方法具有更高的准确性。为了评估准确性, 我们首先将波导离散为256个部分, 并使用传统RCWA计算整个波导的散射矩阵 S^* 。然后将此 S^* 用作

为了测量最大范数误差 $\|S - S^*\|_{\max}$ 的真实值，其中 S 是由我们的方法或传统 RCWA 得到的散射矩阵。在两种情况下，我们都使用单个截面来模拟整个波导，以保证公平比较。我们在横截面变化（由 p 控制）变化时测量误差。如图 4-b 所示，当横截面变化较大（即 $p \rightarrow 1$ ）时，由于高阶近似的优势，我们的方法比传统 RCWA 更准确。当变化变小时（即 $p \rightarrow 0$ ），两种方法的精度相当。

此外，我们评估了我们方法的收敛性。我们再次考虑图4中所示的波导（其中 $p = 1$ ），并在 z 方向分辨率（两种方法中）增加时评估我们方法和传统 RCWA 的散射矩阵误差。结果如图4-c 所示：我们的方法收敛更快；随着分辨率的提高，我们的方法准确性更高。

3.2 Performance for Various Waveguides

接下来，我们测试我们的方法在分析各种光子结构时的性能。这些结构的形状如图5所示。对于每个结构，基于高分辨率的传统RCWA方法计算其真实散射矩阵（ z 方向上的分辨率为 $N = 1024$ ）。在获得真实散射矩阵后，我们测量我们的方法和传统RCWA的性能-精度曲线：对于我们的方法，我们逐步降低误差阈值（即算法1中的 α ）；对于每个误差阈值，我们测量得到的散射矩阵误差和计算成本。这样，我们可以绘制出显示精度随计算时间变化的曲线。同样地，对于传统RCWA，我们逐步提高 z 方向的分辨率，并测量散射矩阵误差和计算时间。所有时间都是在配备8个Intel Xeon(R) E5-1620 CPU（3.60GHz）和NVIDIA GeForce GTX 1080 GPU的工作站上测量的。

所得的性能-精度曲线如图5所示。给定一个固定的精度水平（图5中的绿色水平线所示），我们的方法明显更快。例如，对于计算误差约为 10^{-4} 的散射矩阵，我们的方法在所有测试案例中至少比传统RCWA快 $3.5\times$ 倍。在某些情况下，如图5-e，我们的方法甚至快一个数量级。在所有案例中，我们的方法的收敛速度都快于传统RCWA，证明了我们的方法是高阶方法。这意味着随着所需精度水平的提高，我们方法的加速效果将更加显著。

我们还注意到，当使用传统的RCWA分析特定光子结构时，对于某个精度水平，没有指南可以确定 z 方向的分辨率。因此，在实际操作中，需要通过多次试验来选择合适的分辨率，从而花费的时间比图5中报告的更多。

最后，我们在图6中展示的两个真实案例上评估了我们的方法，即Y型波导和频率分频器。前者是我们基于90/10分频器设计[25]用贝塞尔曲线平滑处理而得到的，后者则在文献[12]中报告。对于这两种情况，其真实散射矩阵都是通过传统RCWA获得的，其中Y型波导使用了8192个离散截面，频率分频器使用了32768个截面，因为后者的几何形状更复杂。如图6所示，我们的方法收敛更快，并且在达到足够高的精度时，我们的方法计算成本要低得多。

4 Conclusion

总之，我们提出了一种高阶半解析方法。我们的方法最适合模拟截面形状沿传播方向变化的光子结构。

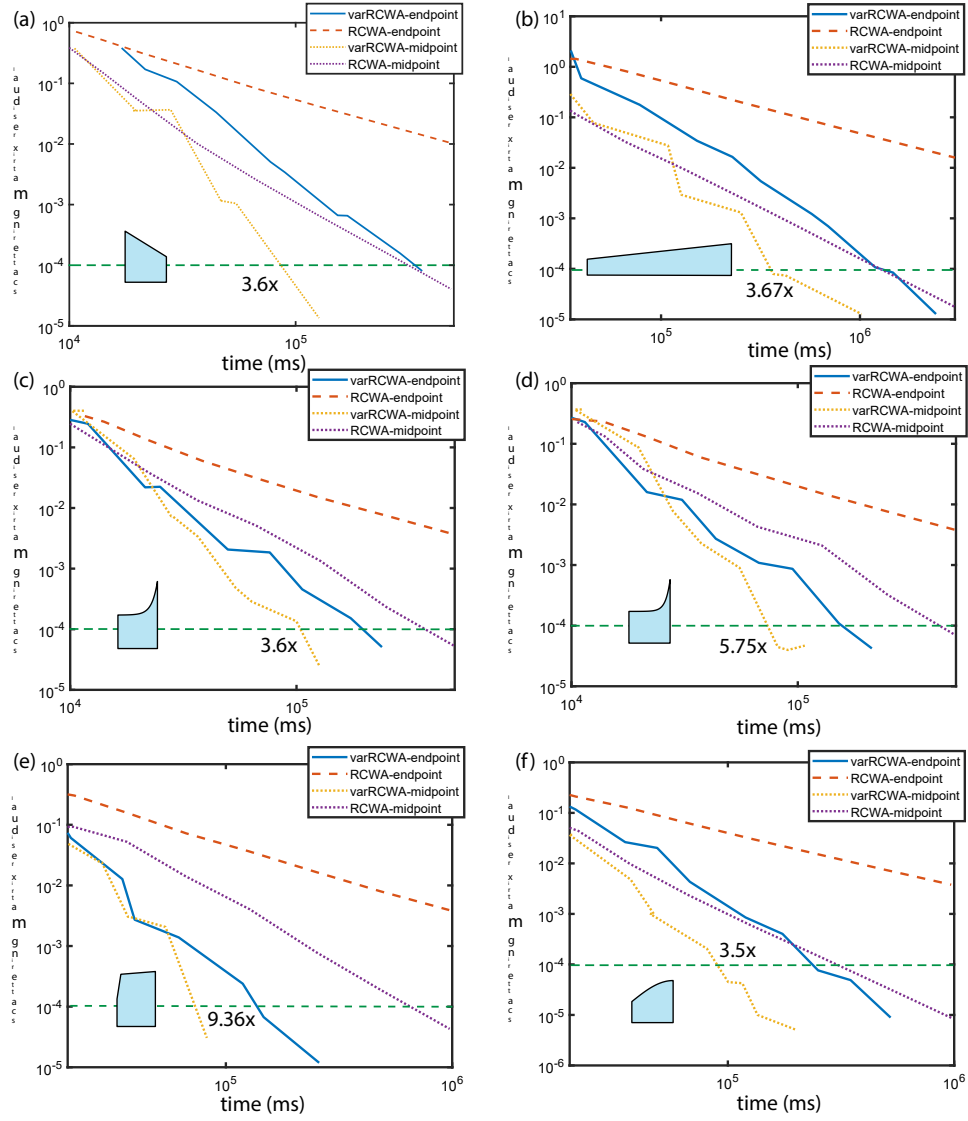


图5: Tests on various waveguide geometries. (a) 一个梯形波导。(b) 一个长度为 $10\mu\text{m}$ 的长波导。(c-d) 单侧呈指数形状变化的波导。(e) 一段分段线性波导。(f) 侧面呈正弦变化的波导。除(b)之外, 所有波导的长度均为 $1\mu\text{m}$ 。它们的厚度固定为 $0.22\mu\text{m}$ 。方法相对于传统RCWA在相同精度水平(即 10^{-4})下的加速比用黄色标出。相关数值见Data File 3-8。

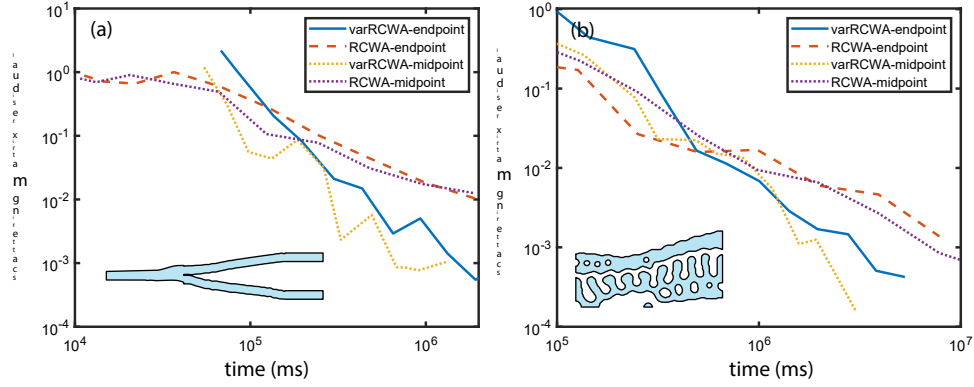


图6: Tests on complex photonic structures. (a) 一个Y型波导。(b) 文献[12]中报道的一个频率分离器。有关基础数值, 请参见Data File 9-12。

与传统的半解析方法相比, 我们的方法更快且更精确。此外, 我们的方法允许用户指定精度水平, 并沿 z 方向自适应地离散化结构以实现所需的精度。

将来, 这项工作可以向多个方向扩展。目前, 当某个截面需要细分时, 该截面会被均匀地划分为三个子截面。也可以根据结构的局部横截面形状进行非均匀划分。此外, 我们的实现使用了我们推导的一级展开式, 虽然我们可以通过将低阶展开递归代入公式(10)和(11)来很容易地扩展到更高阶的展开式。在实践中, 我们基于一级展开的方法已经优于传统的RCWA方法。

Funding. 国家科学基金会 (1910839)。

Acknowledgments. 我们感谢 Utsav Dave 和 Michal Lipson 提出的宝贵建议。

Disclosures. 作者声明不存在利益冲突。支持内容请参见补充材料1。

References

- [1] BLANES, S., CASAS, F., OTEO, J., AND ROS, J. Magnus 及矩阵微分方程的 Fer 展开: 收敛性问题. *Journal of Physics A: Mathematical and General* 31, 1 (1998), 259. [2]
- BLANES, S., CASAS, F., OTEO, J.-A., AND ROS, J. Magnus 展开及其一些应用. *Physics reports* 470, 5-6 (2009), 151–238. [3]
- CHEN, X., ZHENG, C., XU, W., AND ZHOU, K. 一种用于逆弹性形状设计的渐近数值方法。 *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2014)* 33, 4 (2014年8月).

- [4] CHU, H. 光栅光散射的有限差分方法。发表于 *Advanced Characterization Techniques for Optics, Semiconductors, and Nanotechnologies* (2003), 第 5188 卷, 国际光学与光子学学会, 页 358–370。 [5] COCHELIN, B. 通过渐近-数值方法的路径跟踪技术。 *Computers & structures* 53, 5 (1994), 1181–1192。 [6] DIVITT, S., ZHU, W., ZHANG, C., LEZEC, H. J., AND AGRAWAL, A. 使用介电超材料表面实现超快光脉冲整形。 *Science* 364, 6443 (2019), 890–894。 [7] HELTON, J., AND STUCKWISCH, S. 积分乘积的数值近似。 *Journal of Mathematical Analysis and Applications* 56, 2 (1976), 410–437。 [8] ISERLES, A., AND NØRSETT, S. P. 李群中线性微分方程的解法。 *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 357, 1754 (1999), 983–1019。
- [9] JING, X., JIN, S., TIAN, Y., LIANG, P., DONG, Q., AND WANG, L. 正弦纳米图案光栅结构分析。 *Optics & Laser Technology* 48 (2013), 160–166。
- [10] LI, J., SHI, L. H., MA, Y., RAN, Y., LIU, Y., AND WANG, J. 超薄多层光栅的 RCWA 高效稳定实现: 无需求解特征值的 T 矩阵方法。 *IEEE Antennas and Wireless Propagation Letters* (2020)。
- [11] LIU, V., AND FAN, S. S4: 一种用于分层周期结构的自由电磁求解器。 *Computer Physics Communications* 183, 10 (2012), 2233–2244。
- [12] LUMERICAL. 使用轮廓方法的逆向设计器件GDS图案提取, 2021年。
- [13] MILLER, S. A., CHANG, Y.-C., PHARE, C. T., SHIN, M. C., ZADKA, M., ROBERTS, S. P., STERN, B., JI, X., MOHANTY, A., GORDILLO, O. A. J., ET AL. 使用低功率多通道硅光子平台的大规模光学相控阵。 *Optica* 7, 1 (2020), 3–6。 [14] MOHAMAD, H., ESSAIDI, S., BLAIZE, S., MACIAS, D., BENECH, P., AND MORAND, A. 用于微分法和RCWA的快速傅里叶分解: 非层状金属衍射光栅建模的强大工具。 *Optical and Quantum Electronics* 52, 2 (2020), 1–13。 [15] MOHARAM, M., AND GAYLORD, T. 平面光栅衍射的严格耦合波分析。 *JOSA* 71, 7 (1981), 811–818。 [16] NVIDIA, VINGELMANN, P., AND FITZEK, F. H. Cuda, 版本: 10.2.89, 2020年。 [17] PIGGOTT, A. Y., LU, J., LAGOUDAKIS, K. G., PETYKIEWICZ, J., BABINEC, T. M., AND VUČKOVIĆ, J. 紧凑型宽带片上波长解复用器的逆向设计与演示。 *Nature Photonics* 9, 6 (2015), 374–377。 [18] PREGLA, R., AND PASCHER, W. 线路法。 *Numerical techniques for microwave and millimeter wave passive structures* 1 (1989), 381–446。 [19] REDHEFFER, R. 矩阵李卡提方程的不等式。 *Journal of Mathematics and Mechanics* (1959), 349–367。

- [20] ROBERTS, C. M., AND PODOLSKIY, V. A. 严格的衍射界面理论。 *Applied Physics Letters* 110, 17 (2017), 171108. [21]
- SLAVÍK, A. *Product integration, its history and applications*. 布拉格数学物理出版社, 2007. [22]
-] TAKEGOSHI, K., MIYAZAWA, N., SHARMA, K., AND MADHU, P. 固态核磁中 Magnus/Floquet/Fer 展开方案的比较。 *The Journal of chemical physics* 142, 13 (2015), 134201. [23] YEE, K. 各向同性介质中涉及麦克斯韦方程的初边值问题的数值解。 *IEEE Transactions on antennas and propagation* 14, 3 (1966), 302–307. [24]
- ZHU, Z., DAVE, U. D., LIPSON, M., AND ZHENG, C. 制作稳健纳米光子波导的逆几何设计。在 *2020 Conference on Lasers and Electro- Optics (CLEO)* (2020), IEEE, 第 1–2 页. [25]
- ZHU, Z., DAVE, U. D., LIPSON, M., AND ZHENG, C. 通过自适应逆设计实现超宽带纳米光子学。在 *2021 Conference on Lasers and Electro-Optics (CLEO)* (2021), IEEE, 第 1–2 页.

VarRCWA：一种自适应高阶严格耦合波分析方法 ：补充文件

2022年2月1日

1 Explanation of the requirement of the integrand in Eqn. (6)

要在长段落中使用公式 (6)，系数 $\mathbf{W}^{-1}\mathbf{P}\mathbf{V}$ 和 $\mathbf{V}^{-1}\mathbf{Q}\mathbf{W}$ 应该足够小。我们通过一个简单的例子来解释这一要求的直觉，这个例子是一个类似于公式 (5) 形式的标量微分方程。

$$\frac{d}{dz}y(z) = ay(z), \quad (1)$$

其中 a 是一个常数，类似于方程 (5) 中的 $\mathbf{W}^{-1}\mathbf{P}\mathbf{V}$ 和 $\mathbf{V}^{-1}\mathbf{Q}\mathbf{W}$ ，而 $y(z)$ 类似于其中的向量 $\mathbf{a}(z')$ 和 $\mathbf{b}(z')$ 。这个方程可以写成类似于方程 (6) 的积分形式：

$$y(z) = y(z_L) + \int_{z_L}^z ay(z')dz'. \quad (2)$$

通过将 (2) 递归地代入 $y(z')$ ，我们得到

$$y(z) = \left[1 + \int adz' + \frac{1}{2} \left(\int adz' \right)^2 + \dots \right] y(z_L), \quad (3)$$

右边是 $\exp(\int adz')$ 的泰勒展开式：

$$\exp\left(\int adz'\right) = 1 + \int adz' + \frac{1}{2} \left(\int adz' \right)^2 + \dots \quad (4)$$

尽管指数函数的泰勒展开在所有 $\int adz'$ 情况下总是收敛的，但如果我们将该级数截断到一阶，则残差是

$$\text{Re} = \left| \exp\left(\int adz'\right) - 1 - \int adz' \right| = \left| \left(\int adz' \right)^2 \right| \left| \frac{1}{2} + o\left(\int adz'\right) \right|, \quad (5)$$

随 $\int a dz' = a(z - z_L) = aL$ 的大小而缩放。为了在保持积分值 aL ((从而保持准确性)) 的同时增加此积分的长度 L , a 应该足够小。同样地, $\mathbf{W}^{-1}\mathbf{P}\mathbf{V}$ 和 $\mathbf{V}^{-1}\mathbf{Q}\mathbf{W}$ 也应足够小, 以确保一阶截断的合理准确性。 \mathbf{W} 和 \mathbf{V} 不随截面长度缩放, 因为它们是特征向量。因此, \mathbf{P} 和 \mathbf{Q} 必须具有足够小的范数, 但遗憾的是, 这无法得到保证。

2 Derivation of Eqn. (8)

请注意, 系数 \mathbf{P} 和 \mathbf{Q} 取决于截面材料分布, 即使截面是固定的, 它们也可能非常大。因此, 我们希望用只依赖于截面变化的 $\delta\mathbf{P}$ 和 $\delta\mathbf{Q}$ 来替代它们。为此, 我们不是对 \mathbf{a} 和 \mathbf{b} 进行微分, 而是先对 \mathbf{a} 和 \mathbf{b} 进行缩放, 以消掉某些项并构造 $\delta\mathbf{P}$ 和 $\delta\mathbf{Q}$ 。特别地, 令

$$\tilde{\mathbf{a}} = \gamma_1(z)\mathbf{a}(z) \quad (6)$$

$$\tilde{\mathbf{b}} = \gamma_2(z)\mathbf{b}(z) \quad (7)$$

对于某些特定的 $\gamma_1(z)$ 和 $\gamma_2(z)$ 选择。受到 [1] 和 RCWA 推导的启发, 经过几次尝试, 我们在方程 (7) 中得到了 $\gamma_1(z)$ 和 $\gamma_2(z)$ 。

首先, 我们在公式 (7) 的第一部分对 $\tilde{\mathbf{a}}$ 进行微分如下

$$\frac{\partial \tilde{\mathbf{a}}}{\partial z} = -\frac{j}{k_0}\mathbf{\Lambda}\tilde{\mathbf{a}} + \exp\left(-\frac{j}{k_0}\mathbf{\Lambda}z\right)\frac{\partial \mathbf{a}}{\partial z}. \quad (8)$$

第一个项 $-\frac{j}{k_0}\mathbf{\Lambda}\tilde{\mathbf{a}}$ 用于构建 $\delta\mathbf{P}$ 和 $\delta\mathbf{Q}$ 。现在将方程 (5) 代入 (8):

$$\frac{\partial \tilde{\mathbf{a}}}{\partial z} = \frac{j}{2k_0}\exp\left(-\frac{j}{k_0}\mathbf{\Lambda}z\right)\left(-2\mathbf{\Lambda}\mathbf{a} + \mathbf{W}^{-1}\mathbf{P}\mathbf{V}(\mathbf{a} - \mathbf{b}) + \mathbf{V}^{-1}\mathbf{Q}\mathbf{W}(\mathbf{a} + \mathbf{b})\right). \quad (9)$$

由 $\mathbf{P}_r\mathbf{Q}_r = \mathbf{W}\mathbf{\Lambda}^2\mathbf{W}^{-1}$ 和 $\mathbf{V} = \mathbf{Q}_r\mathbf{W}\mathbf{\Lambda}^{-1}$, 我们得到以下关系:

$$\mathbf{\Lambda} = \mathbf{V}^{-1}\mathbf{Q}_r\mathbf{W} = \mathbf{W}^{-1}\mathbf{P}_r\mathbf{V}. \quad (10)$$

接下来, 通过将方程 (10) 代入 (9), 我们得到

$$\begin{aligned} \frac{\partial \tilde{\mathbf{a}}}{\partial z} &= \frac{j}{2k_0}\exp\left(-\frac{j}{k_0}\mathbf{\Lambda}z\right)\left(\mathbf{W}^{-1}(\mathbf{P} - \mathbf{P}_r)\mathbf{V}\mathbf{a} + \mathbf{V}^{-1}(\mathbf{Q} - \mathbf{Q}_r)\mathbf{W}\mathbf{a} - \mathbf{W}^{-1}\mathbf{P}\mathbf{V}\mathbf{b} + \mathbf{V}^{-1}\mathbf{Q}\mathbf{W}\mathbf{b}\right) \\ &= \frac{j}{2k_0}\exp\left(-\frac{j}{k_0}\mathbf{\Lambda}z\right)\left(\mathbf{W}^{-1}\delta\mathbf{P}\mathbf{V}\mathbf{a} + \mathbf{V}^{-1}\delta\mathbf{Q}\mathbf{W}\mathbf{a} - \mathbf{W}^{-1}\mathbf{P}\mathbf{V}\mathbf{b} + \mathbf{V}^{-1}\mathbf{Q}\mathbf{W}\mathbf{b}\right) \end{aligned}$$

\mathbf{b} 的系数中仍然有 \mathbf{P} 和 \mathbf{Q} 。但是这些系数的符号不同, 一个是正数, 另一个是负数。因此, 我们可以加上一个额外的零项

$(\Lambda - \Lambda)b$ 来协助取消：

$$\begin{aligned}
\frac{\partial \tilde{\mathbf{a}}}{\partial z} &= \frac{j}{2k_0} \exp\left(-\frac{j}{k_0} \Lambda z\right) (\delta \mathbf{A} \mathbf{a} - \mathbf{W}^{-1} \mathbf{P} \mathbf{V} \mathbf{b} + \mathbf{V}^{-1} \mathbf{Q} \mathbf{W} \mathbf{b} + (\Lambda - \Lambda) \mathbf{b}) \\
&= \frac{j}{2k_0} \exp\left(-\frac{j}{k_0} \Lambda z\right) (\delta \mathbf{A} \mathbf{a} - \mathbf{W}^{-1} \mathbf{P} \mathbf{V} \mathbf{b} + \mathbf{V}^{-1} \mathbf{Q} \mathbf{W} \mathbf{b} + (\mathbf{W}^{-1} \mathbf{P}_r \mathbf{V} - \mathbf{V}^{-1} \mathbf{Q}_r \mathbf{W}) \mathbf{b}) \\
&= \frac{j}{2k_0} \exp\left(-\frac{j}{k_0} \Lambda z\right) (\delta \mathbf{A} \mathbf{a} - \delta \mathbf{B} \mathbf{b}) \\
&= \frac{j}{2k_0} \exp\left(-\frac{j}{k_0} \Lambda z\right) \delta \mathbf{A} \exp\left(\frac{j}{k_0} \Lambda z\right) \tilde{\mathbf{a}} - \frac{j}{2k_0} \exp\left(-\frac{j}{k_0} \Lambda z\right) \delta \mathbf{B} \exp\left(-\frac{j}{k_0} \Lambda z\right) \tilde{\mathbf{b}}.
\end{aligned} \tag{11}$$

由于 $\tilde{\mathbf{a}}$ 和 $\tilde{\mathbf{b}}$ 的对称性，我们可以用同样的方法对 $\tilde{\mathbf{b}}$ 进行微分。

3 Construction of the high order scattering matrix

方程(3)中的矩阵元可以从方程(12)和(13)中按如下方式提取：

$$\mathbf{T}_{LR} = e^{\frac{j}{k_0} \Lambda(z_R - z_L)} + \frac{j}{2k_0} \int_{z_L}^{z_R} e^{\frac{j}{k_0} \Lambda(z_R - z')} \delta \mathbf{A}(z') e^{\frac{j}{k_0} \Lambda(z' - z_L)} dz' \tag{12}$$

$$\mathbf{R}_R = -\frac{j}{2k_0} \int_{z_L}^{z_R} e^{\frac{j}{k_0} \Lambda(z_R - z')} \delta \mathbf{B}(z') e^{\frac{j}{k_0} \Lambda(z_R - z')} dz' \tag{13}$$

$$\mathbf{R}_L = -\frac{j}{2k_0} \int_{z_L}^{z_R} e^{\frac{j}{k_0} \Lambda(z' - z_L)} \delta \mathbf{B}(z') e^{\frac{j}{k_0} \Lambda(z' - z_L)} dz' \tag{14}$$

$$\mathbf{T}_{RL} = e^{\frac{j}{k_0} \Lambda(z_R - z_L)} + \frac{j}{2k_0} \int_{z_L}^{z_R} e^{\frac{j}{k_0} \Lambda(z' - z_L)} \delta \mathbf{A}(z') e^{\frac{j}{k_0} \Lambda(z_R - z')} dz'. \tag{15}$$

4 Projection of scattering matrix

因为截面 i 处的散射矩阵（在公式 (23-26) 中给出）是以截面 i 参考位置的本征模为基底表示的。如果这些本征模与截面 $(i-1)$ 的本征模不同，我们需要在使用 Redheffer 星乘法将它们相乘之前，将截面 i 处的散射矩阵重新投影，以匹配截面 $(i-1)$ 的本征模。

根据两个截面之间切向场的连续性，如果我们将截面 i 处的本征模表示为 \mathbf{W}_i 和 \mathbf{V}_i ，并且使用上标 $(i-1)$ 表示矢量在截面 $(i-1)$ 的基底中表示，而没有上标的矢量是在截面 i 的基底中表示的。从公式 (2) 可得

$$\begin{aligned}
\mathbf{W}_{i-1} \left(\mathbf{a}_L^{(i-1)} + \mathbf{b}_L^{(i-1)} \right) &= \mathbf{W}_i (\mathbf{a}_L + \mathbf{b}_L) \\
\mathbf{V}_{i-1} \left(\mathbf{a}_L^{(i-1)} - \mathbf{b}_L^{(i-1)} \right) &= \mathbf{V}_i (\mathbf{a}_L - \mathbf{b}_L).
\end{aligned} \tag{16}$$

现在我们表示

$$\begin{aligned}\mathbf{X} &= \frac{1}{2} (\mathbf{W}_i^{-1} \mathbf{W}_{i-1} + \mathbf{V}_i^{-1} \mathbf{V}_{i-1}) \\ \mathbf{Y} &= \frac{1}{2} (\mathbf{W}_i^{-1} \mathbf{W}_{i-1} - \mathbf{V}_i^{-1} \mathbf{V}_{i-1}),\end{aligned}\tag{17}$$

这使我们能够使用 (16) 表达 \mathbf{a}_L 和 \mathbf{b}_L , 即,

$$\begin{aligned}\mathbf{a}_L &= \mathbf{X} \mathbf{a}_L^{(i-1)} + \mathbf{Y} \mathbf{b}_L^{(i-1)} \\ \mathbf{b}_L &= \mathbf{Y} \mathbf{a}_L^{(i-1)} + \mathbf{X} \mathbf{b}_L^{(i-1)}.\end{aligned}\tag{18}$$

与此同时, 根据公式(3)中散射矩阵的定义, 我们有

$$\begin{aligned}\mathbf{b}_L &= \mathbf{R}_L \mathbf{a}_L + \mathbf{T}_{RL} \mathbf{b}_R \\ (\mathbf{X} - \mathbf{R}_L \mathbf{Y}) \mathbf{b}_L^{(i-1)} &= -(\mathbf{Y} - \mathbf{R}_L \mathbf{X}) \mathbf{a}_L^{(i-1)} + \mathbf{T}_{RL} \mathbf{b}_R.\end{aligned}\tag{19}$$

通过将(18)与(19)进行比较, 我们得到散射矩阵重新投影的方法:

$$\begin{aligned}\tilde{\mathbf{R}}_L &= -(\mathbf{X} - \mathbf{R}_L \mathbf{Y})^{-1} (\mathbf{Y} - \mathbf{R}_L \mathbf{X}) \\ \tilde{\mathbf{T}}_{RL} &= (\mathbf{X} - \mathbf{R}_L \mathbf{Y})^{-1} \mathbf{T}_{RL}.\end{aligned}\tag{20}$$

类似地, 从

$$\mathbf{a}_R = \mathbf{T}_{LR} \mathbf{X} \mathbf{a}_L^{(i-1)} + \mathbf{T}_{LR} \mathbf{Y} (\tilde{\mathbf{R}}_L \mathbf{a}_L^{(i-1)} + \tilde{\mathbf{T}}_{RL} \mathbf{b}_R) + \mathbf{R}_R \mathbf{b}_R,\tag{21}$$

我们有

$$\begin{aligned}\tilde{\mathbf{T}}_{LR} &= \mathbf{T}_{LR} \mathbf{X} + \mathbf{T}_{LR} \mathbf{Y} \tilde{\mathbf{R}}_L \\ \tilde{\mathbf{R}}_R &= \mathbf{T}_{LR} \mathbf{Y} \tilde{\mathbf{T}}_{RL} + \mathbf{R}_R.\end{aligned}\tag{22}$$

投影散射矩阵元 $\tilde{\mathbf{R}}_L$ 、 $\tilde{\mathbf{T}}_{RL}$ 、 $\tilde{\mathbf{T}}_{LR}$ 、 $\tilde{\mathbf{R}}_R$ 可以左乘第 $(i-1)$ 节的散射矩阵。

5 CPU Implementation

我们也在CPU上实现了我们的方法。我们选择与图5-a相同的结构来运行实验。如图1所示, 无论我们选择中点还是端点作为参考点, 我们的方法性能都更好。然而, 由于程序是顺序运行的, CPU实现无法并行计算矩阵乘积, 因此效率不高。因此, 与特征值分解相比, 矩阵乘积仍然需要较多的时间 (每次操作大约1-20)。尽管运行在CPU上的RCWA和我们的方法都较慢, 但与传统的RCWA相比, 我们的方法仍然更快, 尽管在GPU上的加速效果没有那么显著。

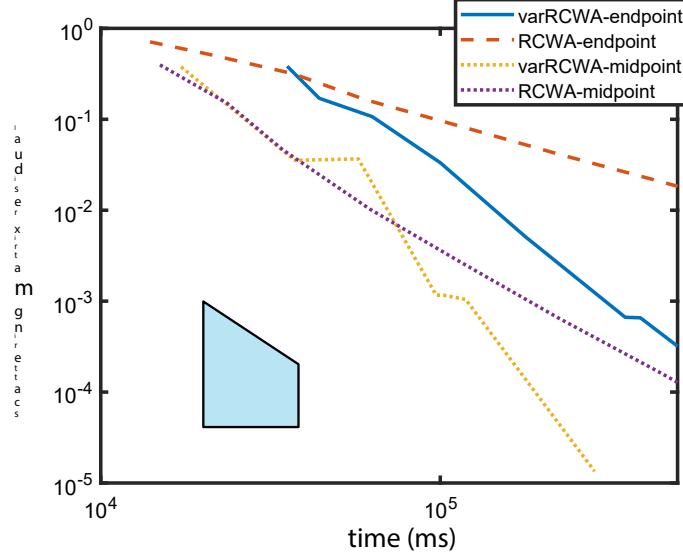


图 1: 图 5(a) 实验的 CPU 实现。底层数值见 **Data File 13**。

6 Error Estimation

在对波导进行细分时，我们的算法会检查估计误差是否达到用户提供的精度水平 α 。为了估计误差，我们将第 p 阶和第 $(p-1)$ 阶的散射矩阵进行比较。在我们的例子中，我们将标准 RCWA ($p=0$) 的结果从公式 (23-26) ($p=1$) 的结果中减去，并得到以下四个差异矩阵

$$\Delta T_{LR} = \frac{j}{2k_0} \int_{z_L}^{z_R} e^{\frac{j}{k_0} \Lambda(z_R - z')} \delta \mathbf{A}(z') e^{\frac{j}{k_0} \Lambda(z' - z_L)} dz' \quad (23)$$

$$\Delta \mathbf{R}_R = -\frac{j}{2k_0} \int_{z_L}^{z_R} e^{\frac{j}{k_0} \Lambda(z_R - z')} \delta \mathbf{B}(z') e^{\frac{j}{k_0} \Lambda(z_R - z')} dz' \quad (24)$$

$$\Delta \mathbf{R}_L = -\frac{j}{2k_0} \int_{z_L}^{z_R} e^{\frac{j}{k_0} \Lambda(z' - z_L)} \delta \mathbf{B}(z') e^{\frac{j}{k_0} \Lambda(z' - z_L)} dz' \quad (25)$$

$$\Delta T_{RL} = \frac{j}{2k_0} \int_{z_L}^{z_R} e^{\frac{j}{k_0} \Lambda(z' - z_L)} \delta \mathbf{A}(z') e^{\frac{j}{k_0} \Lambda(z_R - z')} dz'. \quad (26)$$

我们使用上述四个矩阵的最大范数的最大值 ε 作为估计误差，并将 ε 与用户提供的 α 进行比较。

尽管 ε 是估算值，由高阶扰动引起的任何误差应小于 ε 。我们还将这个估算误差与散射中的最终误差进行了比较

图 5 中每个实验的矩阵。如图 2 所示，对于每个用户指定的 α ，实际误差总是小于 α ，这表明它是一个很好的高估值。

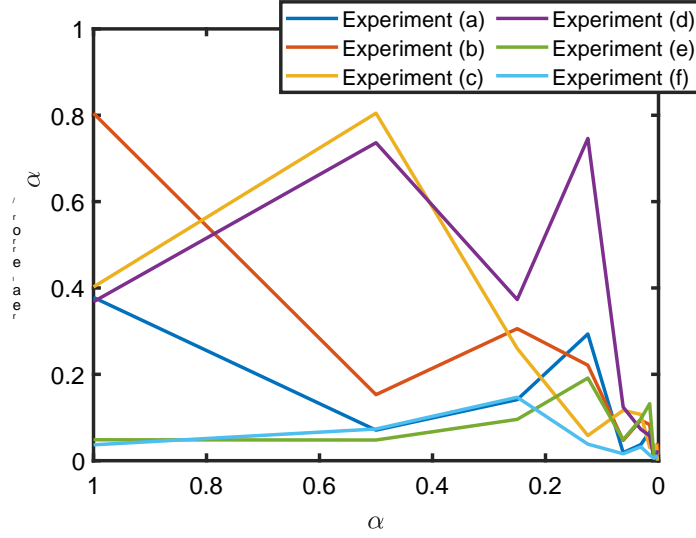


图 2: **Real errors versus different α s for experiments in Fig. 5.** 我们使用用户提供的不同 α (范围从 0 到 1) 重新运行每个实验。当算法终止时，我们将每个散射矩阵与相应的真实散射矩阵 (以大量截面 $N = 1024$ 运行) 进行比较，以得到两矩阵差异的 *real error* (最大范数)。结果显示，散射矩阵的实际误差总是被 α 所限制。有关基础数值，请参见 **Data File 14**。

References

- [1] ISERLES, A. 关于高振荡方程的诺伊曼级数方法。 *Bit Numerical Mathematics* 44, 3 (2004), 473–488.