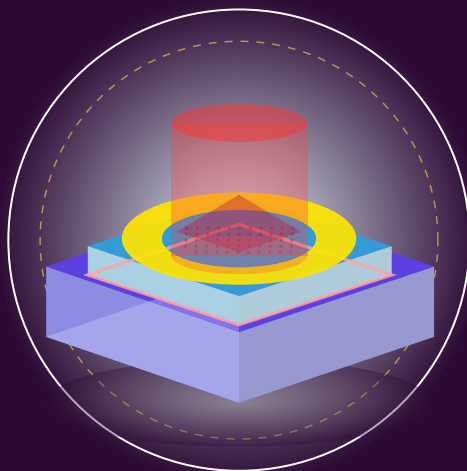


PHOTONIC DEVICES MONOGRAPH SERIES

PCSEL理论、仿真 与器件物理导论

从 Maxwell 方程、Bloch 模与能带分析，
到耦合波、全波数值方法、半导体外延层、
电注入与热-电-光耦合的系统导论



Rigorous · Multiscale · Device-Oriented

严谨 · 多尺度 · 器件导向

Feiyang Wu

2026 年 3 月

前言

一枚出光口径约 3 mm 的芯片，怎样在连续波下输出 50 W，同时把发散角压到约 0.05° ？这组数字把 PCSEL 最吸引人的潜力和最难的物理同时摆到了桌面上：大面积、高功率与窄远场必须在同一个开放周期腔里成立。本书沿一条尽量连续、少跳步的路径，把这个结果拆回物理图像、数学模型、数值方法和器件实现，使读者知道每个模型能回答什么，也知道它不能替代什么。

PCSEL 值得单独讨论，是因为它同时涉及多个通常分开处理的知识模块。把它仅仅视为一个光子晶体共振器，会忽略它作为半导体激光器的增益、损耗、阈值、电注入和热效应；把它仅仅视为一个半导体激光器，又会错过二维周期开放电磁系统、Bloch 模、Gamma 点、辐射耦合和远场工程这些核心特征。对 PCSEL 的分析至少需要同时包含三个层面：它是一个周期开放电磁系统；它是一个具有增益与损耗竞争的有源光学腔；它还是一个依赖外延层、载流子输运、电流扩展和热分布的真实半导体器件。若只保留其中一个层面，结论容易失去适用边界，并可能导致设计判断错误。

因此，本书的组织原则不是“把所有相关名词列一遍”，而是尽量把它们压缩成一条可连续阅读的主线。前几章从 Maxwell 方程、周期介质和 Bloch 理论出发，先建立读者理解 PCSEL 所需的电磁学骨架；随后讨论 Gamma 点、带边、辐射耦合和面内反馈，把“为什么能面发射、为什么可能大面积单模、为什么不能把被动高 Q 直接等价成低阈值”这些最容易被误解的问题讲清楚；再往后进入耦合波理论、平面波展开法、RCWA、FDTD 与 FEM，使读者能够从不同保真度层次理解和计算 PCSEL；后半部分则把重点转向半导体外延层、量子阱增益、电注入、漂移-扩散、热效应和多物理场耦合，说明为什么真实器件设计不能停留在“冷腔 + 均匀增益”的理想模型上。

本书坚持三条写作原则：重要结论给出推导或来源；常用近似说明适用条件；仿真结果标明它属于单胞、无限周期、有限腔、被动开放系统还是有源器件。贯穿全书的三组区分是阈值增益与阈值电流、被动共振与实际激射模、能带候选模与真实工作模。

本书不把商用软件界面写成逐步手册。软件版本会变化，控制方程、边界条件、离散思想、收敛标准和数据接口更稳定。因此数值章节先解释方法在数学上解什么，再说明几何、材料和边界如何进入，最后把输出接回 PCSEL 设计；Lumerical 与 COMSOL 只作为工作流实例出现。

如果你是第一次接触 PCSEL，最稳妥的阅读方式是顺着本书章节次序向前推进，不要一开始就跳到器件级阈值或多物理场仿真章节。因为那些章节虽然看起来更接近工程结果，但它们的输入量，例如候选模、约束因子、材料增益、损耗分布、热反馈灵敏度，本质上都建立在前面的电磁与模式分析之上。如果基础骨架没有搭起来，后面的器件结果只会变成一堆参数

和软件设置。如果你已经具备一定背景，则可把本书当作一份交叉索引式参考：围绕某个具体问题，例如“Gamma 点模的辐射损耗如何调控”“为什么高 Q 模不一定是最低阈值模”“电流扩展怎样影响模式竞争”，沿着书中的交叉引用回溯相关章节。

为减少阅读中的难度跳变，全书“严格表述”框统一给出难度标记：（本科可读，建议通读）、（研究生基础，建议带着问题读）、（研究专项，可按需选读）。若你当前目标是先建立主线，可优先保证正文与 框读通，再回看更高星级框。

本书不是某套实验数据的汇编，也不是特定材料体系的工艺手册。需要材料数据库、精确工艺窗口或软件版本支持之处，只给出建模框架、数量级和验证路径；严格结论、近似结果与待校准量会分别标明。这个边界是让书稿成为可靠研究起点的必要条件。

本书的目标，是帮助读者判断 PCSEL 模型是否合理、仿真流程是否闭合，以及器件结论是否超出模型边界。阅读时应始终注意：PCSEL 不是抽象的光场图样，也不是孤立的能带曲线，而是一个把周期电磁结构、有源增益介质和真实半导体器件约束同时绑定在一起的问题。只有在这三个层面上同时保持清醒，后续的设计、仿真与研究判断才会真正可靠。

Feiyang Wu

2026 年 3 月

目录

前言	i
插图索引	xvii
符号与记号	xix
证据标签规范	xxi
缩略语表	xxii
第一部分 总览与电磁基础	1
第一章 PCSEL 是什么，为什么重要	2
1.1 从传统激光器到光子晶体激光器：技术背景	2
1.1.1 边发射激光器的成就与局限	3
1.1.2 VCSEL 的突破与新挑战	3
1.1.3 DFB 激光器的作用与限制	4
1.1.4 光子晶体概念的引入与 PCSEL 的定位	4
1.2 为什么还需要一种新的面发射激光器	5
1.3 PCSEL 的基本图像：从有源导模到法向出光的完整链条	7
1.3.1 第一步：有源区产生导模	8
1.3.2 第二步：二维布拉格散射与 $\Gamma^{(2)}$ 点共振	8
1.3.3 第三步：辐射通道与法向出光	8
1.3.4 第四步：模式选择——从单晶格到双晶格再到 Hermitian/非 Hermitian 联合调控	9
1.3.5 小结：一条连贯的出光链	9
1.4 定义 PCSEL：什么必须成立，什么不应混淆	9
1.5 与相邻器件逐一比较：通过对比限定边界	10
1.5.1 PCSEL、VCSEL 与 RCLED 的定量对照（代表性报道）	10
1.6 为什么大面积单模困难，而 PCSEL 可能有优势	12
1.7 PCSEL 必须同时用三种视角来理解	13
1.7.1 第一种视角：开放周期电磁系统	13

1.7.2 第二种视角：半导体有源光学腔	13
1.7.3 第三种视角：真实电注入器件	13
1.8 关键观测量：本书后面所有章节都在回答哪些量	15
1.9 本书的主线：从 Maxwell 到工作器件	15
1.10 2024–2026 进展速览：从“能激射”走向“系统可用”	16
第二章 全书路线图、符号、坐标与基本约定	18
2.1 为什么在第二章先停下来统一语言	18
2.2 本书默认的 benchmark 教学器件	18
2.3 前置工具页：全书最高频符号先一次看全	19
2.4 空间坐标与结构坐标：先分清什么是纵向，什么是面内	19
2.5 时间因子与复频率：为什么符号约定会影响物理解释	21
2.5.1 角频率、普通频率与波长	21
2.6 本书的模型层级：为什么同一个问题会有不同答案	21
2.7 几个最容易混淆的术语：先立边界，细节由首用章展开	24
2.8 如何阅读后续章节：每一章都在回答链条上的哪一环	24
2.9 全书模型依赖图：每章向下一章交付什么	24
2.10 全书需要反复区分的六类边界	25
第三章 周期介质中的 Maxwell 方程	27
3.1 从时域方程出发：我们到底在求解什么	27
3.2 转到频域：为什么教材几乎总在频域讨论模式	28
3.3 算符观点：为什么有时本征频率是实数，有时必须是复数	29
3.4 开放 Maxwell 问题的函数空间直觉：为什么它不再是普通闭腔本征值问题	31
3.5 QNM / resonant states 的最小定义	32
3.6 为什么 PCSEL 作为开放系统	32
3.7 复频率、品质因子与光子寿命	33
3.8 开放边界的实现：数学边界条件与数值边界条件不是一回事	34
3.9 这一章和后文的关系：为什么先讲 Maxwell，再讲 Bloch	36
第四章 Bloch 定理与二维光子晶体基础	37
4.1 为什么周期性如此重要	37
4.2 Bloch 定理的来源：平移算符与共同本征态	38
4.3 晶格平移与倒格矢记号	40
4.4 为什么波矢只需取第一布里渊区	41
4.4.1 约化区方案与扩展区方案	41
4.4.2 区折叠、Umklapp 过程与 Γ 点候选模	42
4.5 二维晶格、倒晶格与布里渊区	42
4.6 从平面波展开理解能带为什么会出现	45
4.7 二维光子晶体的常见晶格：方格与三角晶格	45

4.7.1	三角晶格速览：后面为什么会频繁出现 K 点与 C_{6v}	46
4.7.2	初读高对称路径时应看什么	46
4.8	把二维理想图像放回真实 PCSEL: slab 结构会带来什么变化	47
4.9	TE/TM 分类为什么常常只是近似	47
4.9.1	一个对初学者更安全的学习顺序	47
4.10	Bloch 理论能告诉我们什么，又不能告诉我们什么	48
4.11	通向下一章：为什么 Gamma 点成为重点	48
第五章	光子能带基础、Gamma 点物理与模选择	50
5.1	为什么在 Bloch 之后必须讲能带	50
5.2	先把两个“能带”分开：材料电子能带与光子晶体能带	51
5.3	什么是光子晶体能带	52
5.4	一个一维玩具模型：带隙为什么会打开	53
5.5	Γ 点四带与 ABCD 模式分类	54
5.5.1	Γ 点四带的起源与 ABCD 分类	56
5.5.2	能带计算的收敛性测试：一个可复现的最小例子	57
5.5.3	为什么对称圆孔常给出中心发暗的远场	57
5.6	一个最小例子：正方晶格圆孔二维光子晶体	60
5.7	怎么看一张能带图	60
5.8	从能带图中我们究竟能得到什么	61
5.8.1	带边模式、缺陷模式与 PCSEL 候选模的区别	62
5.8.2	有限尺寸下带边模式的品质因子为什么会随群速度降低而增大	63
5.8.3	光锥、引导共振与为什么不是所有带边都适合面发射	64
5.8.4	用三角晶格 I-IV 的布拉格图像理解为什么只有某些带边能成为面发射候选模	64
5.9	为什么 Γ 点会成为 PCSEL 的重点	65
5.10	带边模、 Γ 点模、被动共振模和真实激射模的区别	65
5.11	为什么能带图好看不等于器件就会好用	66
5.12	从本章到 PWEM：为什么还需要平面波展开法	66
第二部分	对称性、BIC 与降阶理论	69
第六章	面发射机制与面内耦合	70
6.1	先从最朴素的问题开始：如果没有光子晶体，会发生什么	70
6.2	面内反馈从哪里来：从倒格矢散射看布拉格回路	71
6.3	为什么说 PCSEL 的反馈不是一维 DFB 的简单二维版	73
6.4	垂直辐射为什么会出现：从光锥条件理解面发射	73
6.5	最小四波图像：为什么它能抓住 PCSEL 的主导物理	75
6.6	晶格傅里叶分量为什么如此关键	76

6.7	反馈与提取之间的张力：为什么 PCSEL 设计不能只追一个指标	77
6.8	模式竞争在本章图像下如何理解	77
6.9	有限尺寸为什么会改变本章图像	77
6.10	2024–2026 前沿补充：反馈网络的“调制-功率”双目标约束	78
6.11	与后续章节的关系	78
第七章	阈值、增益、损耗与约束因子	80
7.1	阈值首先是净衰减被补偿的条件	80
7.2	从材料增益到模增益：约束因子为什么必须出现	81
7.3	内部损耗的组成与标定	83
7.4	为什么高 Q 有利，但不等于最低阈值	84
7.5	从速率方程再看一遍阈值	85
7.6	非均匀增益为什么会改写模式排序	86
7.7	为什么本章只能给出阈值增益，不能给出阈值电流	86
7.8	为什么本章故意停在静态阈值，而不在这里直接讨论动态学	87
7.9	从冷腔候选模到主动激光模：本书采用的理论层级	87
7.10	本章应输出的结果	88
7.10.1	从注入电流到阈值条件：为什么需要一张 Sankey 图	88
第八章	偏振、对称性破缺与远场整形	91
8.1	偏振不是输出端的装饰，而是模式本身的一部分	91
8.2	为什么对称性会控制偏振	92
8.3	Γ 点小群与不可约表示：最小分类工具	92
8.4	方格晶格 Γ 点： C_{4v} 分类	94
8.5	三角晶格 Γ 点： C_{6v} 分类	95
8.6	法向辐射通道本身也有表示：这是 BIC 判断的最小规则	95
8.6.1	把选择定则写成真正可复用的张量积条件	95
8.6.2	如何从数值模场中真正给出 irrep 标签	96
8.7	从对称操作看模式：为什么有些辐射能出，有些出不来	97
8.8	从近场到远场：为什么傅里叶角谱是理解 PCSEL 远场的核心	97
8.9	偏振选择为什么常常和模式分裂绑在一起	98
8.9.1	频率分裂路径	98
8.9.2	损耗分裂路径	98
8.10	对称性破缺如何在 irrep 语言下导致分裂	98
8.11	拿到一个新结构后，如何自己做 Γ 点模标注	100
8.12	哪些对称性破缺最常见	100
8.13	外延不对称为什么会影响偏振和远场	101
8.14	远场整形不是“把图样画漂亮”，而是组织角谱能量	101
8.15	为什么偏振和远场必须放在同一章讨论	102

8.16 通向后续章节：偏振问题为何最终会回到器件问题	102
第九章 BIC / quasi-BIC、Γ 点辐射与远场工程	104
9.1 为什么 BIC 不是旁枝话题，而是 PCSEL 主线中的辐射语言	104
9.2 从辐射振幅到 BIC：什么叫“连续谱中的束缚态”	105
9.3 对称性保护 BIC 的物理图像：从偶极子到点群表示	106
9.4 Γ 点与 BIC 语言的对应关系	107
9.5 把法向辐射通道写成 irrep：BIC 判断才真正可操作	107
9.6 从 BIC 到 quasi-BIC：为什么真实 PCSEL 更关心后者	108
9.7 前沿补充：quasi-BFIC 与宽角高 Q 设计	110
9.8 进阶：far-field polarization vortex 与拓扑电荷语言	110
9.9 对称性破缺在 irrep 语言下如何打开 quasi-BIC	112
9.10 BIC 如何连接偏振选择与远场工程	112
9.11 为什么 BIC 仍然不能替代阈值与器件模型	113
第十章 PCSEL 的耦合波理论 (CWT)：最小模型	115
10.1 为什么在能带图之后还需要 CWT	115
10.2 最小 CWT 要解决的物理问题是什么	116
10.3 从 Maxwell 方程到场展开：四波近似从哪里来	116
10.4 投影步骤到底在做什么	117
10.5 三类矩阵项各自代表什么	120
10.5.1 先说清归一化：重叠积分和耦合系数的数值可比性	120
10.6 耦合系数为什么是几何和场重叠的压缩表达	121
10.7 四个基本波如何重新组合成实际候选模	122
10.8 把四波基底改写成 C_{4v} 的 irrep 基底	123
10.9 这套 irrep 基底怎样直接连接偏振、辐射与 BIC 判断	124
10.10 对称性破缺怎样在 CWT 矩阵里体现为 irrep 分裂	125
10.11 为什么二维八波 CWT 仍然不够：它不是错误，而是压缩得太狠	125
10.12 三维 CWT 能算什么：一页速览	125
10.13 CWT 怎样表达阈值条件	126
10.14 从线性阈值到模式竞争动力学	126
10.15 一个最典型的设计问题：几何扰动到底在帮你什么	127
10.15.1 三晶格如何直接进入 CWT：Fourier 放大因子与 180° 耦合增强	128
10.16 CWT 到底能算什么，不能算什么	128
10.17 CWT 在整本书 workflow 中的准确位置	129
10.18 2024–2026 前沿补充：CWT 的极化与晶格泛化正在变成可复用工具链	130
第十一章 三维与更高阶 CWT：完整推导与器件接口	132
11.1 为什么二维最小 CWT 迟早会不够用	133
11.2 升级思路：恢复哪三类自由度	134

11.3 从三维 Maxwell 方程重新投影：为什么会出现“平面波乘纵模”的基底	134
11.4 从二维到三维：Liang–Noda 3D CWT 的推导框架	136
11.4.1 为什么必须从矢量波动方程出发	136
11.4.2 基本波的场型假设 (separation of variables)	137
11.4.3 从分量方程到耦合方程的推导	137
11.4.4 辐射波的解析求解	138
11.4.5 高阶波的解析求解：线性组合技巧	139
11.4.6 三个子矩阵的显式矩阵形式	141
11.4.7 从四波耦合方程到完整本征问题	142
11.4.8 三维 CWT 的核心输出与二维模型的本质差异	142
11.4.9 为什么辐射波与高阶波不能共用基本波的纵向场型	143
11.4.10 圆孔与非对称孔形：干涉抵消与辐射打开	144
11.4.11 品质因子与辐射常数的提取	144
11.4.12 非 Γ 点扩展：能带结构计算	145
11.5 辐射通道为什么在三维 CWT 中必须显式处理	145
11.6 有限尺寸 CWT：从常数振幅到包络方程	146
11.7 更高阶 CWT 到底多算出了什么	148
11.8 什么时候继续升级 CWT，什么时候直接转向全波	148
11.9 三维 CWT 与器件级模型的接口为什么更自然	149
第十二章 耦合模理论、微扰观点与模式分裂	150
12.1 为什么在 CWT 之外还要学耦合模理论	150
12.2 时间耦合模理论的基本语言	151
12.3 从单模到多模：模式竞争和分裂如何写进 TCMT	152
12.4 微扰观点为什么如此重要	152
12.5 先用一个 2×2 非厄米玩具模型把左模/右模讲明白	153
12.6 为什么非厄米一阶扰动不能直接照搬厄米公式	154
12.7 双正交、伴随模与 QNM 归一化在这里各自扮演什么角色	155
12.8 开放系统中微扰不只改频率，还会改损耗	156
12.9 哪些工程公式可以继续用，哪些地方必须升级到开放系统严格形式	156
12.10 两模耦合为什么足以抓住“分裂”的核心	157
12.11 对称性破缺怎样导致模式分裂	158
12.12 TCMT、CWT 和微扰三者怎样配合使用	159
12.13 为什么本章内容会反复回到后面的器件章节	160
第三部分 数值方法	162
第十三章 平面波展开法 (PWEM) 与光子能带计算	163
13.1 为什么在能带概念之后必须把它真正算出来	164

13.2 从二维理想光子晶体到可计算本征问题	164
13.2.1 二维模型中的极化分离为什么是教学上有价值的	164
13.2.2 选取 H_z 标量问题作为教学主线	165
13.2.3 同一个二维模型里还有另一类标量问题	165
13.3 Bloch 形式与平面波展开: PWEM 的数学骨架	166
13.3.1 矩阵元为什么正好长成这个样子	166
13.4 正方晶格圆形孔洞: 把单胞几何写成傅里叶系数	167
13.4.1 几何与倒格空间定义	168
13.4.2 圆孔指标函数的傅里叶系数	168
13.5 从二维教学模型到真实 slab: 修正等效介电常数	169
13.6 如何搭建一个最小 PWEM 计算	170
13.6.1 步骤一: 定义教学型几何与参数表	170
13.6.2 步骤二: 生成倒格矢基底并做截断	171
13.6.3 步骤三: 计算傅里叶系数并组装矩阵	171
13.6.4 步骤四: 沿高对称路径逐点求解	171
13.6.5 步骤五: 做收敛检查, 而不是直接相信第一张图	171
13.7 本书附带的 Python 计算脚本与实际带图	172
13.8 如何读一张 PWEM 算出来的能带图	176
13.9 从 PWEM 结果中我们究竟能得到什么	177
13.10 PWEM 结果应该怎样交给后续章节	178
13.11 为什么真实 PCSEL 不能停在二维 PWEM	179
第十四章 RCWA 在周期开放结构与阈值分析中的应用	181
14.1 为什么在 PWEM 之后自然轮到 RCWA	181
14.2 RCWA 的基本思路: 平面内做傅里叶, 纵向做分层	182
14.3 为什么散射矩阵是 RCWA 的真正核心	184
14.4 被动谱、共振与极点: 谱线为什么会尖起来	185
14.4.1 随书实算: 从 PWEM 频窗到 RCWA 反射谱	186
14.5 RCWA 怎样进入光学阈值分析	186
14.6 为什么 RCWA 特别适合做“单胞层面的设计判断”	188
14.7 RCWA 不能替你解决什么问题	188
14.8 2024–2026 前沿补充: modified RCWA 正在覆盖 full-PCSEL 结构	189
14.9 收敛与数值卫生: 哪些检查是必须交代的	189
14.10 RCWA 在整条工作流中的准确位置	190
第十五章 FDTD 与 FEM 在 PCSEL 建模中的作用	192
15.1 为什么有限尺寸一旦重要, 全波方法就不可回避	192
15.2 同一个 Maxwell 方程, 为什么会分化出两种数值哲学	193
15.3 FDTD 的强项: 宽带、瞬态与模建立过程	194

15.4 FEM 的强项：复杂几何、本征模与局域高精度	195
15.5 几何设置为什么不能只靠“画出来”	195
15.6 边界条件：PML 不是附件，而是模型的一部分	196
15.7 激励源的选择，其实是在决定你想看什么	197
15.8 怎样从时域或频域结果里提取共振频率与品质因子	197
15.9 近场、远场与模场：怎样读出可用于后续建模的量	199
15.10 冷腔分析和带增益分析的区别	199
15.11 收敛检查：哪些量必须扫，哪些量必须复核	200
15.12 本章在整条器件主线中的位置	200
第十六章 CWT、RCWA、FDTD、FEM 的比较与互证	202
16.1 为什么单一方法几乎注定不够	202
16.2 先按“问题类型”而不是“方法名字”来比较	203
16.3 四种方法各自最擅长什么	204
16.4 比较方法时，应比较哪些量	205
16.5 数值误差从哪来：互证前的最后一项功课	205
16.6 为什么互证必须按“由粗到细”的顺序组织	206
16.7 一条适合 PCSEL 的推荐 workflow	206
16.8 2024-2026 前沿补充：代理模型与 AI 设计器如何接入互证链	206
16.9 互证不是礼仪，而是发现错误的主要手段	208
16.10 什么时候该停在某一层，不再继续加方法	208
16.11 本章在后续器件章节前的作用	209
第四部分 器件物理与设计	210
第十七章 PCSEL 的半导体外延层与纵向光学结构	211
17.1 从二维光子晶体到真实器件：建立完整纵向物理链	211
17.2 SCH 与器件可行性	212
17.3 异质结和能带对齐为什么必须在本章就讲	213
17.4 电子泄漏和空穴注入困难为什么会在 PCSEL 中显得更尖锐	215
17.5 典型 PCSEL 外延层从下到上分别承担什么角色	215
17.6 从 slab 波导方程开始理解纵向模式	216
17.7 有效折射率近似在这里为什么既必要又危险	216
17.8 重叠因子与损耗重叠	217
17.9 掺杂和金属为什么在本章就已经是光学问题	218
17.10 外延层如何改写上下辐射不对称、偏振与远场	219
17.11 层厚灵敏度与孔径工艺对比：从可算到可制造	220
17.12 本章输出给后续器件章节的量	221

第十八章 量子阱、有源区与增益模型	223
18.1 为什么本章不能写成一般半导体激光器的量子阱综述	223
18.2 量子阱为什么是 PCSEL 有源区的自然选择	224
18.3 SQW 与 MQW: 不是多放几层阱这么简单	224
18.4 有源区工程为什么必须回到能带图: 阱、势垒、限制层和阻挡层	225
18.5 2024-2026 前沿补充: 跨材料平台增益窗口不可直接平移	226
18.6 应变工程和 TE/TM 增益差异: 为什么它在 PCSEL 中尤其不能忽略	227
18.7 电子阻挡层、空穴阻挡层和限制层到底在帮谁	228
18.8 从材料增益到模增益: 上一章的重叠因子在这里真正接上线	228
18.9 透明载流子密度、微分增益和工作点灵敏度	229
18.10 简化增益模型、真实模型与它们各自的边界	229
18.11 为什么到这里仍然不能推出阈值电流	230
第十九章 电注入、载流子输运与阈值电流	233
19.1 阈值电流回答的器件问题	233
19.2 欧姆接触和肖特基接触如何进入边界条件	234
19.3 接触几何和接触电阻为什么会直接进入模式问题	234
19.4 Current Spreading Layer	235
19.5 为什么大面积器件不可能自动均匀注入	235
19.6 电流孔径、注入窗口与模式竞争为什么直接耦合	236
19.7 漂移-扩散方程组: 器件模型的骨架	237
19.8 复合、漏电与局域注入效率	238
19.9 把局域载流子分布重新接回模增益: 模式竞争在这里被重写	239
19.10 阈值电流的变量链条	239
19.11 为什么本章天然要求下一章的热反馈进入	241
第二十章 热效应、非均匀性与热-电-光耦合	243
20.1 为什么热效应在 PCSEL 中比在许多小面积激光器里更具结构性	244
20.2 热源必须写成三维分布	244
20.3 热方程求解的是温度场	245
20.4 温升如何回写到光学: 折射率、增益谱、吸收和冷腔频率全都会动	247
20.5 温升如何回写到电学	248
20.6 热斑、空间烧孔和模式竞争: 为什么高功率问题不能只看平均参数	250
20.7 把前面三章闭成热-电-光回路	250
20.8 哪些问题还可以用简化热模型, 哪些已经必须做多物理场闭环	251
20.9 不同材料平台如何让同一个 PCSEL 框架发生变形	251
第二十一章 大面积单模设计原则	255
21.1 为什么大面积单模是 PCSEL 最难也最值钱的问题	255
21.2 模鉴别要写成裕量	257

21.3 第一尺度：单胞与 Gamma 点物理决定的是候选资格	257
21.4 第二尺度：有限孔阵与边缘损耗决定的是横向可实现性	259
21.5 第三尺度：注入与温升决定的是工作点下的真实胜负	260
21.6 边缘损耗与表面提取共享同一预算	261
21.7 从 2023 年 Noda tutorial 抽出的放大设计规则	261
21.8 把设计原则翻译成仿真任务：应该在哪类工具里算什么	262
21.9 真正的设计目标是工作窗口，而不是最佳单点	263
21.10 2024–2026 设计导向补充：高功率、调制与系统链路	264
第二十二章 有限尺寸、无序与工艺误差鲁棒性	266
22.1 为什么名义值最优不等于器件最优	267
22.2 误差如何进入模型：从参数向量到性能向量	267
22.3 几何无序的第一组：孔半径无序与晶格常数漂移	268
22.4 几何无序的第二组：刻蚀深度、侧壁倾角与表面粗糙	269
22.5 装配与器件误差：金属错位、电流孔径错位和边缘截断	269
22.6 误差为什么会改变模排序、辐射、偏振、远场、注入与热	270
22.7 哪些误差可以先用微扰观点，哪些必须重算	270
22.8 如何做 tolerance analysis：不是盲扫，而是分层抽样	271
22.9 如何报告 robustness window	273
22.10 前沿补充：quasi-BFIC 视角下的鲁棒性	273
22.11 把鲁棒性设计接回工作流与失败模式	274
第二十三章 动力学、调制与噪声	276
23.1 为什么“跨过阈值”只是动态学的起点	278
23.2 最小单模速率方程：把阈值语言延伸到起振以上	278
23.3 增益压缩与多模竞争：为什么工作点不会无限平滑	279
23.4 弛豫振荡与小信号调制：从稳态到频率响应	279
23.5 热时间尺度为什么会把快动态问题改写成慢漂移问题	281
23.6 从相位噪声到线宽：Henry 因子为什么出现在这里	281
23.7 RIN、模式分配噪声与快慢噪声分层	282
23.8 外腔反馈：为什么即使本书不展开，也必须给读者一个入口	282
23.9 自发辐射耦合、Purcell 与 β_{sp} 因子：哪些微腔语言可用，哪些不能滥用	283
23.10 哪些动态结论可以直接建模，哪些必须实验闭环	284
第二十四章 线宽、相干性、光束质量与亮度	286
24.1 为什么性能指标必须和工作点一起报告	286
24.2 从一阶相干函数到线宽：定义先于经验公式	287
24.3 频率噪声谱、短时/长时线宽与测量协议	288
24.4 Petermann 因子与开放腔非正交性：什么时候必须提醒读者	290
24.5 外腔反馈与环境耦合：为什么性能章也必须给出一句硬说明	290

24.6 相干性与线宽相关，但并不完全等价	290
24.7 M^2 、亮度和功率为什么必须一起看	291
24.8 偏振纯度、远场稳定性和模式稳定性本质上连在一起	292
24.9 直接与间接可计算量	293
24.10 性能报告的正确写法：条件、区间和可复核性	294
24.11 2024–2026 前沿补充：高相干与系统化应用	294
第五部分 workflow、例题与失败模式	296
第二十五章 一套实用的 PCSEL 仿真 workflow	297
25.1 本章如何与下一章例题和失败模式章配合使用	297
25.2 workflow 的第一原则：先定义结论层级，再跑仿真	300
25.3 workflow 的第二原则：先建数据账本，再建几何	301
25.4 最小命令行复现骨架：即使使用商业软件，也应把文件流做成终端原生	302
25.5 阶段 A：先做纵向模和层栈检查，而不是急着扫平面晶格	303
25.6 阶段 B：周期单胞筛选，目标是排除错误结构而不是宣布最终器件	303
25.7 阶段 C：有限器件全波验证，检查无限周期假设到底失效到什么程度	304
25.8 阶段 D：电注入与热建模，开始把阈值增益映射成阈值电流区间	305
25.9 阶段 E：热-电-光迭代，不追求一轮算完，而追求工作窗口收敛	305
25.10 阶段 F：鲁棒性验证，检查工作窗口是否同时具有工艺窗口	305
25.11 阶段 G：实验回写链，不让模型停在“可运行”而要走到“可校准”	306
25.12 阶段 H：代理模型与 AI 加速（新增）	306
25.13 Lumerical 与 COMSOL 的两条常见实现路径	307
25.14 何时停止迭代：停机判据比“再跑一轮看看”更重要	308
第二十六章 教材式例题	310
26.1 第一层玩具模型	311
26.1.1 例题 A 的设定	311
26.1.2 第一步：把矩阵换到对称性基底	312
26.1.3 第二步：读出模式分裂和损耗排序	313
26.1.4 第三步：在简并子空间中加入一个最小对称性破缺	314
26.1.5 第四步：这一层玩具模型到底回答了什么	314
26.2 第二层例题：数值全流程全解	315
26.2.1 例题 B 的设定与已知量	315
26.2.2 解题前先写定性预期	315
26.2.3 第一步：纵向模检查	316
26.2.4 第二步：单胞候选模筛选	317
26.2.5 第三步：有限阵列检查	318
26.2.6 第四步：简化电学与热学回写	319

26.2.7 给例题 B 补一页可核对的“实算结果页”	320
26.2.8 第五步：把中间结果收束成一张教材式汇总表	322
26.3 这一章到底给出了哪些“可靠结论”	322
26.4 把例题映射回工作流与失败模式	323
第二十七章 PCSEL 建模常见失败模式	325
27.1 把本章当作工作流和例题的反向读法	326
27.2 PCSEL 中最需要警惕的失败，不是算不出来，而是错误结果被误认为可靠结论	326
27.3 第一类失败：把冷腔结果直接读成器件结论	326
27.4 第二类失败：边界条件和几何定义在不同求解器里并不一致	327
27.5 第三类失败：把有效折射率近似当成严格三维模型	327
27.6 第四类失败：把均匀增益默认当作真实电注入分布	327
27.7 第五类失败：有限尺寸建模不足，导致把单胞排序误读成器件排序	328
27.8 第六类失败：忽略热漂移，默认冷腔频率就是工作频率	329
27.9 第七类失败：过度相信单一求解器或单一图像	329
27.10 可复现故障案例：低截断怎样制造伪频移与伪分裂	329
27.11 这些失败在 Lumerical 与 COMSOL 中通常长什么样	330
27.12 推荐排错顺序：先层级，再物理，再数值	331
第六部分 附录与索引	333
第二十八章 数学附录	334
28.1 为什么数学附录不是装饰，而是主结论的一部分	334
28.2 时间因子与频域约定：先把增益和损耗的符号说清楚	335
28.3 周期介质中的傅里叶展开：为什么卷积会反复出现	336
28.4 开放 Maxwell 问题的函数空间直觉：为什么普通 Hilbert 空间不够	337
28.5 QNM / resonant states 的最小定义	338
28.5.1 QNM 理论的三个核心数学问题（进阶选读）	339
28.6 光子学中的 Green 函数方法简介	341
28.6.1 从标量波动方程的 Green 函数出发	341
28.6.2 Dyson 方程及其物理含义	342
28.6.3 矢量 Green 张量与 Maxwell 方程	342
28.6.4 周期结构中的 dyadic Green 函数与 Bloch-Floquet 展开	342
28.6.5 Green 函数极点与准正则模的对应关系	343
28.6.6 Purcell 因子的 Green 函数表达	343
28.7 伴随模、双正交与广义归一化	344
28.8 复频率、衰减率与 Q ：不要混淆振幅寿命和能量寿命	345
28.9 互易性：为什么许多耦合公式其实来自同一个定理	346
28.10 一阶微扰公式：什么时候它是定理，什么时候它只是工程近似	347

28.11 这些数学工具在 PCSEL 仿真中怎么实际使用	348
28.12 从 CWT 本征矢量提取远场偏振：方位角、椭圆率与 Stokes 参量	348
28.13 群论工具箱： C_{4v} 点群完整参考	349
28.14 关键量的首次定义位置索引	352
第二十九章 半导体激光器与器件建模附录	355
29.1 为什么器件建模最容易在“接口”上出错	356
29.2 从材料增益到阈值电流：先把符号链写全	356
29.3 第一步：材料增益怎样变成模增益	357
29.4 第二步：从模增益到净模增益，再到阈值	357
29.5 第三步：速率方程的一致性核查	358
29.6 第四步：完整器件方程组是什么，各自属于哪类模型	358
29.7 边界条件不是实现细节，而是器件物理的一部分	359
29.8 参数来源要分层管理：测量值、文献值和拟合值不是同一种可信度	359
29.9 校准顺序为什么通常是先电学，再热学，最后光学回写	360
29.10 测量 → 参数提取：先区分直接观测量与反演量	360
29.11 预测 → 回验：如何判断模型真的具备外推能力	362
29.12 仿真-实验对标协议：如何让数字与测量对话	363
29.12.1 对标前的层级对齐检查	363
29.12.2 三级对标体系：从定性到定量	363
29.12.3 关键物理量的对标技术规范	365
29.12.4 当仿真与实验不符时的诊断树	366
29.12.5 对标结果的报告伦理	367
29.13 不确定性传播：不要只给单值，要给证据范围	367
29.14 回到 PCSEL：可交付器件结论的组成	368
第三十章 仿真附录：检查清单与流程表	370
30.1 为什么验证不是额外工作，而是物理论证本身	370
30.2 通用验证逻辑：先问收敛，再问不变量，再问模型边界	371
30.3 RCWA 验证：验证对象是单胞开放散射，而不是某条窄谱线	371
30.4 FDTD 验证：时间窗、PML 和激励源决定你看到的是模还是伪影	372
30.5 FEM 验证：收敛的不只是网格，还有问题本身的弱形式实现	372
30.6 收敛性检查的目标：不是让数字不变，而是让结论不变	373
30.7 从冷腔到器件级：需要怎样的阶段门槛	373
30.8 收敛标准不应只写“已收敛”，而应写成门槛	374
30.9 网格收敛最好固定成五步流程	374
30.10 排错决策树：结果异常时按什么顺序查	375
30.11 怎样把验证结果写成可发布报告	376
30.12 数值不确定度量化与报告规范	376

30.12.1 各方法的主要误差源	376
30.12.2 最小必要收敛扫描组合	380
30.12.3 不确定度的合成与传递	381
30.12.4 交叉验证的可接受偏差范围	381
30.12.5 可重复性报告的最小信息集	382
第三十一章 术语表与符号总表	384
31.1 如何使用本章	384
31.2 核心术语表	385
31.3 符号总表	387
31.4 多义符号与上下文辨识	388
31.5 高频混淆对照	389
31.6 如何把本章用回 workflows、例题与失败模式	390

插图索引

图号	简短标题	页码
图 1.1	四种主要半导体激光器的结构对比	6
图 1.2	PCSEL 的纵向层状结构与面内反馈、法向辐射概念	7
图 1.3	光子晶体面发射激光器建模的三重视角与数据传递链	11
图 2.1	本书统一教学器件示意	15
图 2.2	本书采用的四层建模结构	19
图 2.3	全书逻辑依赖图	23
图 2.4	“六条红线”可视化速记图	23
图 3.1	复频率平面示意	30
图 4.1	正方晶格实空间原胞与第一布里渊区	42
图 4.2	二维实空间晶格与倒空间晶格示意	42
图 4.3	方格晶格第一布里渊区构造示意	43
图 4.4	方格晶格第一布里渊区高对称点路径	43
图 4.5	三角晶格第一布里渊区示意	44
图 5.1	Γ 点与非 Γ 点模式近场远场对比	54
图 5.2	正方晶格圆孔教学几何与路径示意	58
图 5.3	能带图读图示意	59
图 6.1	动量折叠与光锥耦合示意	71
图 6.2	平板导模色散与光锥边界示意	71
图 6.3	三类主要面发射机制对比	72
图 7.1	损耗机制分解示意图	80
图 7.2	阈值增益分解示意图	84
图 8.1	近场对称性到远场角谱映射示意	96
图 8.2	远场辐射方向图示例	96
图 8.3	偏振调控方法对比	99
图 9.1	远场偏振涡旋最小示意	109
图 10.1	最小四波 CWT 图像	114
图 11.1	耦合波升级路径示意	141
图 12.1	端口-腔模-端口最小模型	150

图号	简短标题	页码
图 12.2	微扰诱导模式分裂示意	156
图 13.1	PWEM 标准计算流程图	164
图 13.2	平面波矩阵结构与截断示意	165
图 13.3	二维标量能带实算结果	169
图 14.1	RCWA 阶梯离散策略	179
图 15.1	三维 FDTD 仿真配置示意	189
图 16.1	数值方法性能雷达对比	199
图 17.1	典型外延堆栈截面示意	212
图 17.2	导带/价带台阶示意	213
图 17.3	外延层纵向结构示意	214
图 17.4	量子阱重叠与损耗重叠示意	217
图 18.1	压缩应变量子阱能带结构示意	222
图 18.2	多量子阱与阻挡层能带示意	224
图 19.1	横向扩展层与注入均匀性关系示意	233
图 19.2	电流拥挤效应示意	237
图 20.1	热效应多物理场耦合路径图	244
图 20.2	热效应多物理场耦合路径图（教学版）	244
图 20.3	热路等效电路类比图	245
图 20.4	热-电-光耦合闭环量传递示意	247
图 21.1	大面积阵列相干锁定机制对比	252
图 22.1	工艺误差传播图	259
图 23.1	动态调制与眼图示意	269
图 24.1	线宽增强机制示意	280
图 25.1	工程设计工作流总图	290
图 25.2	仿真工作流示意	298
图 26.1	算例桥接图	306
图 27.1	建模排错流程图	319
图 28.1	数学对象关系图	329
图 29.1	校准-验证闭环图	347
图 30.1	建模排错决策树	361
图 31.1	符号与术语分类索引图	367

符号与记号

本书采用时间因子 $e^{-i\omega t}$ 约定。除非特别说明， ω 为角频率， $f = \omega/(2\pi)$ ，波长为 λ 。

符号	含义
$\mathbf{r} = (x, y, z)$	直角坐标； z 轴取外延层法向。
$\mathbf{k}$	波矢；在周期系统中写为 $\mathbf{k} + \mathbf{G}$ 。
$\mathbf{G}$	倒格矢（reciprocal lattice vector）。
Γ 点	布里渊区中心， $\mathbf{k} = \mathbf{0}$ 。
$\mathbf{E}, \mathbf{H}$	频域电场（electric field）与磁场（magnetic field）复振幅。
$\varepsilon(\mathbf{r}, \omega)$	介电函数；可含色散与损耗。
Q	品质因子， $Q = \omega_0/(2\gamma)$ （被动定义）。
$\gamma_{\text{rad}}, \gamma_{\text{int}}$	辐射损耗与内部损耗对应的衰减率。
κ_1, κ_2	CWT 中由主导几何傅里叶分量和纵向模重叠给出的耦合系数。
β_{slab}	slab/导波模型中的面内传播常数（单位： m^{-1} ）。
β_{sp}	自发辐射耦合因子（无量纲），定义见动态与噪声章节。
β_{Bloch}	仅在少数上下文中用于记 Bloch 波矢分量；默认更推荐直接写 $\mathbf{k}$ 分量以避免歧义。
g_{mat}	材料增益（material gain）。
g_N	微分增益（differential gain），即材料增益对载流子密度的局域导数。
N_{tr}	透明载流子密度（transparency carrier density）。
Γ_{opt}	光学约束因子（optical confinement factor）。
$\Delta G_{\star j}$	模鉴别裕量；表示目标模与第 j 个竞争模之间的净优势。
g_{mod}	模增益（modal gain），常写为 $\Gamma_{\text{opt}} g_{\text{mat}}$ （近似）。
g_{th}	阈值增益（threshold gain），属于光学阈值条件。
I_{th}	阈值电流（threshold current），属于器件级电-热-光结果。
N	载流子浓度（carrier density）。
ϕ	静电势（electrostatic potential）。
T	温度场。
$R_{\square}$	电流扩展层的片电阻（sheet resistance）。
L_s	电流扩展长度（spreading length）。

易错点

符号 Q 、 g_{th} 、 I_{th} 分别对应被动腔、光学阈值、器件阈值三个层级，不能互相替代。 β 统一优先写成 β_{slab} 或 β_{sp} ，避免跨章节误读。

证据标签规范

为避免把教学示意、工程经验和可核验文献值混写，本书统一使用下列证据标签：

标签	含义与使用规则
[文献值]	来自可核验的同行评审文献；引用时应给 DOI 或标准引文。
[本书算例]	来自本书示例脚本或示例工况；默认不可直接外推到其他平台。
[工程经验]	来自工程实践的启发式范围；必须同时给出边界条件。
[教学示意]	仅用于帮助理解机制与数量级，不作为发布级定量结论。
[预印本补充：待同行评审]	用于前沿趋势补充；若作为证据出现，正文必须同时给出“待同行评审”提示，并尽量配套同行评审兜底文献。

缩略语表

缩写	含义
PCSEL	photonic-crystal surface-emitting laser, 光子晶体面发射激光器。
PWEM	plane-wave expansion method, 平面波展开法。
RCWA	rigorous coupled-wave analysis, 严格耦合波分析。
FDTD	finite-difference time-domain, 时域有限差分法。
FEM	finite-element method, 有限元法。
CWT	coupled-wave theory, 耦合波理论。
TCMT	temporal coupled-mode theory, 时间耦合模理论。
BIC	bound state in the continuum, 连续谱中的束缚态。
QNM	quasinormal mode, 准正规模。
SCH	separate confinement heterostructure, 分离限制异质结构。
SQW / MQW	single quantum well / multiple quantum wells, 单量子阱 / 多量子阱。
PML	perfectly matched layer, 完全匹配层。
EBL / HBL	electron blocking layer / hole blocking layer, 电子阻挡层 / 空穴阻挡层。
RIN	relative intensity noise, 相对强度噪声。
SALT	steady-state <i>ab initio</i> laser theory, 稳态从头算激光理论。

第一部分

总览与电磁基础

第一章 PCSEL 是什么，为什么重要

读完本章，你能

- 从激光器应用需求出发理解 PCSEL 为何会成为一个独立方向。
- 准确区分 PCSEL、边发射激光器、VCSEL、DFB 激光器与一般光子晶体激光器。
- 建立本书的基本视角：PCSEL 同时是开放周期电磁系统、半导体有源腔与真实器件。

读前准备 光学、固体物理、基础电动力学、傅里叶光学、半导体激光器基本概念。

阅读主线 本章所处模型层级：L1→L3（总览）

本章先从”为什么需要这样一种激光器”出发，回顾从传统激光器到 PCSEL 的历史演进脉络，再说明 PCSEL 的物理图像，随后通过与相邻器件对比划清边界，最后给出贯穿全书的三重视角与关键观测量。

1.1 从传统激光器到光子晶体激光器：技术背景

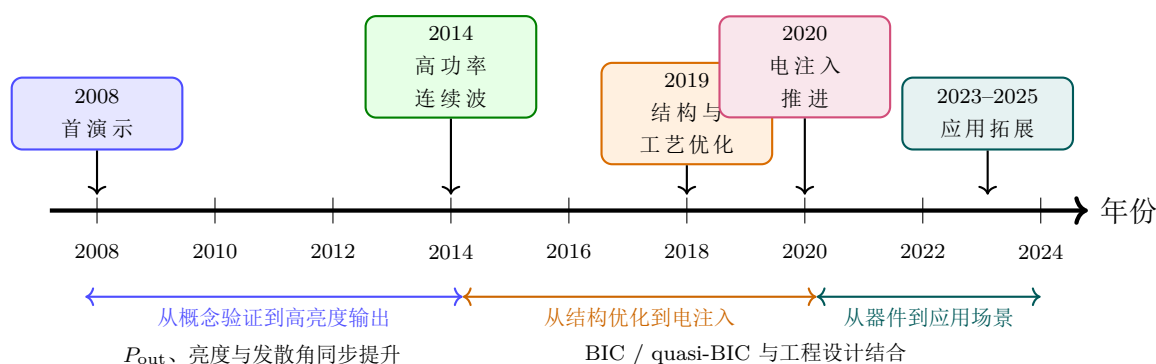


图 1.1: PCSEL 发展历程中的代表性里程碑（2008–2025）。2008 年首次实现二维光子晶体面发射激光 [43]；2014 年实现高功率、窄发散连续波输出 [20]；2019–2020 年器件结构和电注入方案进一步成熟 [75, 25]；2023–2025 年在高亮度、传感与系统集成方向持续推进 [74, 69, 77]。

要理解 PCSEL 为什么会成为独立的研究方向，需要先回答一个问题：在它出现之前，已有的半导体激光器技术遇到了什么瓶颈？本节将从三种主流半导体激光器的局限性出发，说明 PCSEL 形成独立研究方向的技术背景。图1.1展示了 PCSEL 发展的关键时间节点和技术突破。

1.1.1 边发射激光器的成就与局限

边发射激光器 (edge-emitting laser) 是最早实现商业化的半导体激光器类型。它的反馈机制建立在两个端面上：要么依靠解理面形成的法布里 - 珀罗腔，要么引入一维光栅形成 DFB (distributed feedback) 结构以获得更好的单模特性。

优势 边发射激光器具有以下优点：

- 成熟的工艺路线和可靠性记录；
- 较高的输出功率（尤其是条形结构和锥形放大器），连续波功率可达数百毫瓦至数瓦量级；
- DFB 结构可实现优异的单模特性和窄线宽，线宽可压窄至 MHz 量级。

根本局限 然而，当应用需求转向大面积、高亮度面发射时，边发射激光器暴露出三个难以克服的问题：

1. **光束质量椭圆度**：由于有源区在平行于结平面方向和垂直于结平面方向的尺寸差异巨大（常见为微米量级 vs. 亚微米量级），快慢轴发散角通常明显不对称（教材常用示意约为平行方向 $\sim 10^\circ$ 、垂直方向 $\sim 30^\circ\text{--}40^\circ$ ）[10]。虽然可以通过柱面透镜组校正，但这会增加封装复杂度和成本。
2. **面积扩展受限**：若要增大发光面积以降低功率密度和热负荷，条宽增加后会激发高阶横模，破坏光束质量和单模特性。这是一个波动光学结果：波导横向尺寸越大，可容纳模式数越多。对于“条宽超过几十微米后基模占比下降”的说法，工程上应按具体波导结构与热设计判断，不当当作固定阈值 [10]。
3. **面发射困难**：边发射的本质决定了其输出方向平行于芯片表面，这在需要二维阵列集成或与光纤端面耦合的应用中会带来额外的光学系统设计负担。

1.1.2 VCSEL 的突破与新挑战

垂直腔面发射激光器 (VCSEL) 的出现解决了边发射的部分痛点。它采用上下两组分布式布拉格反射镜 (DBR) 构成垂直谐振腔，实现了法向面发射。

优势 VCSEL 在器件形态和制造流程上带来了重要改进：

- 天然的圆形远场斑图和较低的阈值电流（小孔径器件常见 mA 量级，常被写作约 1–5 mA 的教学区间）[10, 59]；
- 无需解理即可进行晶圆级测试，大幅降低成本；
- 易于构建二维阵列，适用于并行光互连和传感应用。

物理根源导致的限制 但是，VCSEL 的结构特点也埋下了新的隐患。DBR 反射镜虽然可提供很高反射率（工程目标常超过 99.9%），但其工作原理基于多层膜干涉，这带来明显的应力与工艺窗口约束 [10]：

DBR 的生长应力与曲率效应限制了可实现的单模孔径尺寸。

具体来说，为达到足够反射带宽和低损耗，GaAs/AlGaAs 体系常需几十对 DBR（教材中常见约 30–40 对量级）[10]。层对数增加会加剧应力与曲率问题；器件孔径增大后，模式稳定性通常更难维持（“20–50 μm 量级易进入多模风险”应视为 [工程经验] 区间，而非普适临界值）。

此外，VCSEL 的微腔长度通常在波长量级（约 1λ – 3λ ），对应较大的纵模间隔（常见可达数 nm 到十 nm 量级），有利于单纵模工作 [10]。但这也意味着有源区体积较小、热与功率密度约束更强；虽然存在大孔径 VCSEL 路线，其高功率与单模光束质量往往需要额外结构和工艺协同 [59]。

1.1.3 DFB 激光器的作用与限制

DFB 激光器通过在波导中引入一维周期光栅，利用布拉格散射提供分布式反馈。它在通信领域取得了巨大成功，特别是在需要严格单模工作的场合。

核心问题 然而，传统的 DFB 激光器本质上仍是一维器件。它的反馈只发生在沿波导传播的方向上，而在横向则依赖折射率导引或增益导引。这种不对称性导致了两个后果：

1. **一维光栅 vs. 二维扩展的不兼容**：如果想要扩大发光面积，单纯增加条宽会引入横向模式不稳定；而如果在两个方向上都引入周期调制，那就已经进入了二维光子晶体的范畴。
2. **边发射的固有属性**：即使采用表面光栅耦合器来实现面发射，其效率也受到相位匹配条件的严格限制，难以在大面积上保持一致的辐射特性。

1.1.4 光子晶体概念的引入与 PCSEL 的定位

1987 年，Yablonovitch 和 John 分别提出了光子晶体的概念。他们的原始动机截然不同：Yablonovitch 希望通过构造三维周期介电结构来抑制自发辐射，从而提高 LED 效率和激光器性能；John 则关注无序介质中的安德森局域化现象在电磁波中的类比。

从三维结构到二维实现 最初的研究热情集中在三维光子晶体上，期望能实现全方位的光子带隙。然而，制备工艺的复杂性使得三维结构长期停留在微波波段和理论模拟阶段。直到微纳加工技术的进步，才使得近红外和可见光波段的二维光子晶体成为可能。

PCSEL 的独特定位 正是在这个背景下，PCSEL 形成了明确的器件定位：1999 年，Noda 等人 and Meier 等人几乎同时独立提出，可以利用二维光子晶体能带结构上的奇异点（singularity point）作为半导体激光器的腔 [50]。两者的关键差异在于所选奇异点：Meier 等人采用三角晶格 X 点的有机增益介质，由于 X 点的基本 Bloch 波只能一维耦合，未能形成二维相干共振；Noda 等人则选择了三角晶格半导体光子晶体的 Γ 点，使三个方向的基本 Bloch 波能够一维

并二维地相互耦合，并由于 Γ 点位于光锥之上，自然获得来自表面的辐射输出——“PCSEL”这一命名正是为了强调这种面发射特性 [50]。沿着这一思路，PCSEL 进一步发展出几个关键结构特征：

- 它不像三维光子晶体那样追求完整的三维带隙，而是巧妙地利用二维面内周期结构提供的分布式反馈；
- 它继承了 VCSEL 的面发射优势，但通过将反馈机制从垂直 DBR 转移到面内光子晶体，绕过了 DBR 应力问题对大面积的限制；
- 它吸收了 DFB 的单模选择能力，并将其扩展到二维，从而实现了对外向模式的主动调控；
- 同时，它通过对称性工程和辐射耦合设计，把大面积工作与单模或主模主导输出放入同一设计框架。

表1.1 给出了四种主要半导体激光器的定性比较，更详细的定量分析将在后续章节结合具体物理机制展开。

表 1.1: 四种主要半导体激光器的特性对比预览。详细机理需参阅后续对应章节。

特性	边发射	VCSEL	DFB	PCSEL
反馈维度	一维轴向	一维垂直	一维轴向	二维面内
发射方向	边发射	面发射	多为边发射	面发射
面积可扩展性	受限 (高阶模)	受限 (DBR 应力)	受限 (一维)	较好
偏振可控性	中等	较弱	较强	强 (对称性工程)
温度敏感性	中等	较高 (DBR 漂移)	较低	中等

上述历史背景说明，PCSEL 是半导体激光器在面发射、大面积扩展和模式控制需求共同推动下形成的器件路线。后续章节将从二维周期结构、开放辐射通道和有源器件约束三方面展开其工作机制。

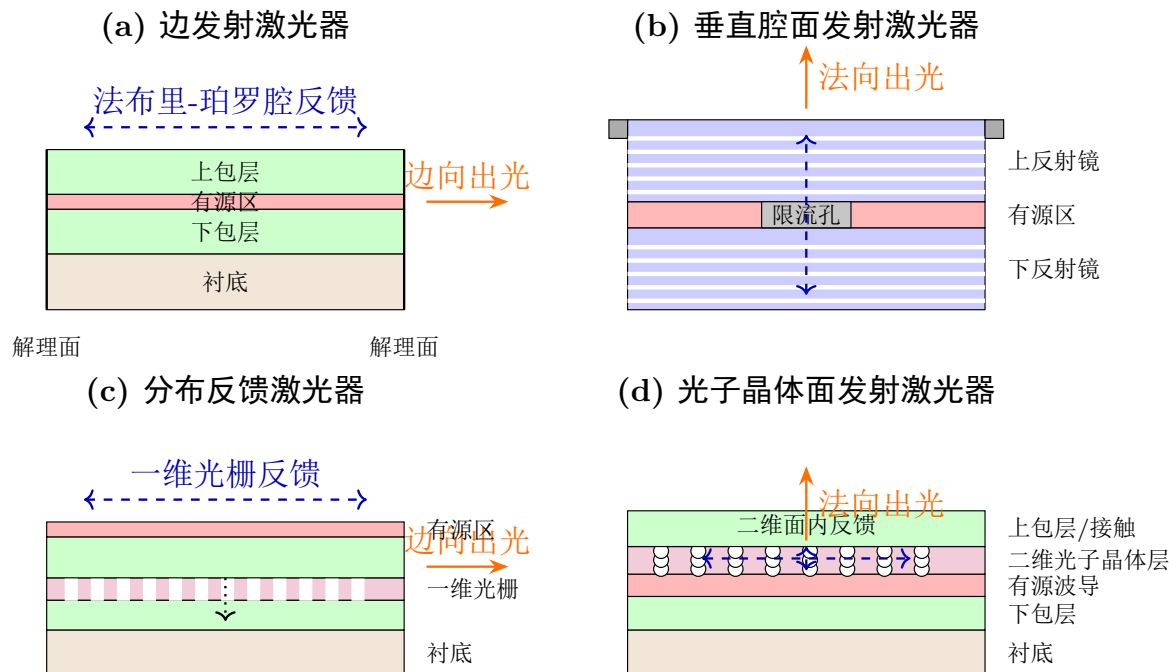
图1.2 给出了四种器件的剖面结构示意图。图中双向虚线箭头表示反馈方向，橙色实线箭头表示出光方向：边发射和 DFB 的反馈沿波导轴向，VCSEL 沿垂直方向，而 PCSEL 在面内两个方向建立反馈。这种结构差异直接决定了它们的出光特性。

1.2 为什么还需要一种新的面发射激光器

如果读者已经接触过边发射激光器 (edge-emitting laser) 或垂直腔面发射激光器 (vertical-cavity surface-emitting laser, VCSEL)，很容易产生一个疑问：现有半导体激光器种类已经很多，为什么还需要专门讨论 PCSEL？

这个问题应从光源需求来回答。许多应用同时希望得到以下几件事：

- 面发射，便于阵列集成、封装与自由空间耦合；
- 大出光面积，以承受更高功率并降低局部热负荷；



图注：四种器件的核心差异在于反馈维度（一维或二维）与出光方向（边向或法向）。光子晶体面发射激光器的关键是“二维面内反馈 + 法向面发射”的统一实现。

图 1.2: 四种主要半导体激光器的结构对比。(a) 边发射激光器: 法布里-珀罗腔沿波导轴向, 光从解理面输出; (b) 垂直腔面发射激光器: 垂直分布式布拉格反射镜谐振腔, 通过顶部氧化孔限制模式, 从表面出光; (c) 分布反馈激光器: 一维光栅提供分布式反馈, 主要为边发射 (可选表面光栅耦合器); (d) 光子晶体面发射激光器: 二维光子晶体提供面内分布式反馈, 同时实现法向面发射。

- 接近单横模或至少主模占优，以获得高亮度（brightness，单位发射面积、单位立体角内的辐射功率，单位 $\text{W} \cdot \text{cm}^{-2} \cdot \text{sr}^{-1}$ ；提高亮度要求在提升功率的同时收紧远场发散角）和良好光束质量；
- 可工程化的远场和偏振，以适配照明、探测、投影或相干系统。

这四个目标并不天然相容。出光面积做大以后，横向可容纳的模式数迅速增加；模式一多，远场会恶化，亮度会下降，偏振也更难稳定。边发射激光器通常在一维波导方向上建立反馈，VCSEL 则主要依赖垂直 DBR 腔。它们各有优势，但面对“大面积、面发射、同时保持较好模式控制”这一组目标时，都会碰到各自的边界。PCSEL 在这个背景下出现：它试图利用二维光子晶体的面内分布式反馈，一边把横向模式组织起来，一边允许能量向法向辐射输出。

直觉框

PCSEL 的问题意识是“如何让一个大面积半导体有源区，在面发射条件下仍由少数目标模稳定主导输出”，而非“如何在光子晶体里做出激光”。这一定义强调的是器件目标，而不是表面结构。

1.3 PCSEL 的基本图像：从有源导模到法向出光的完整链条

图1.3 给出了 PCSEL 最基本的层状图像。外延层沿 z 方向堆叠，光子晶体周期主要位于平面内 (x, y) 。从纵向看，它像一个有源 slab 波导；从横向看，它又像一个二维周期散射网络。这两种视角的叠加构成了 PCSEL 的特殊性。

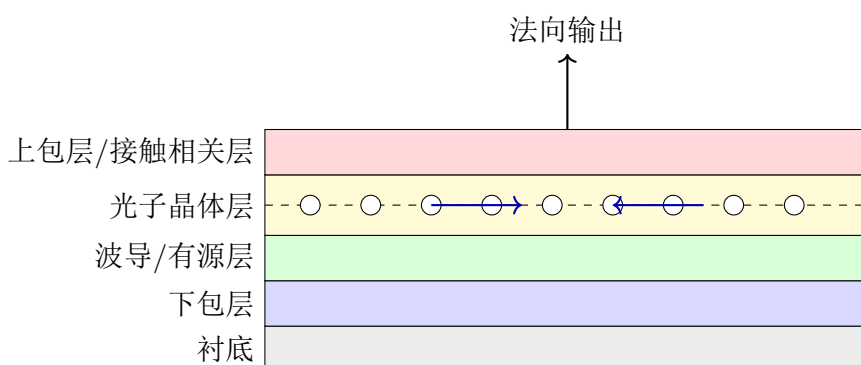


图 1.3: PCSEL 的纵向层状结构与面内反馈、法向辐射的概念图。

对初学者而言，最容易把 PCSEL 想成“一个有孔的出光面”。这种理解忽略了面内反馈与辐射通道的形成过程。下面我们分步描述 PCSEL 的出光过程，将整条物理链条连贯地呈现出来 [50]。

图1.4 把这条链条压缩成一张读图用的守门图：左侧是纵向层栈，右侧是从有源增益到真实模式选择的五个关口。它的重点是提醒读者每一步需要交付什么证据。

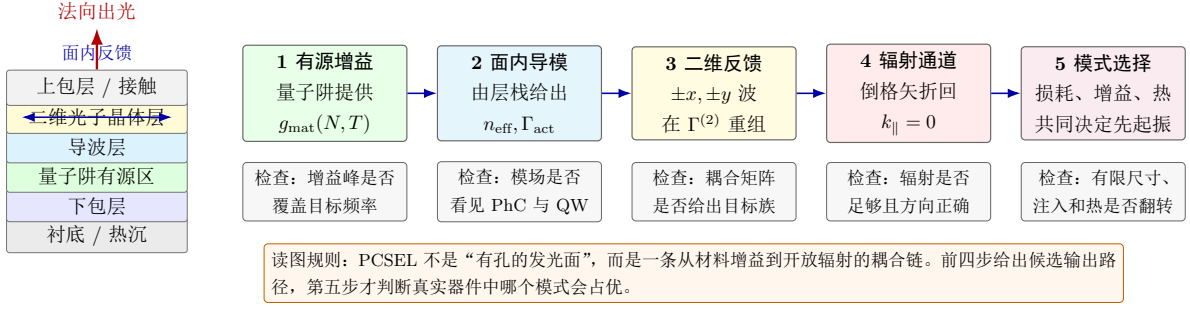


图 1.4: PCSEL 从有源导模到法向出光的五步证据链。量子阱先提供材料增益，层栈形成面内导模，二维光子晶体在 $\Gamma^{(2)}$ 点建立面内反馈，周期散射打开法向辐射通道，最后由损耗、增益、有限尺寸、注入和热反馈共同决定真实起振模式。

1.3.1 第一步：有源区产生导模

电注入使量子阱有源区产生受激辐射。有源层的折射率在整个外延层栈中通常最高，因此辐射的光场以平面导波模式被束缚在以有源层为核心的薄板波导中，沿面内方向传播。光子晶体层紧邻有源层，使得导模场通过倏逝耦合“感受到”周期折射率调制。定义光场在有源层和光子晶体层的约束因子（confinement factor）分别为 Γ_{act} 和 Γ_{pc} 。PCSEL 的结构允许对活性层-包层折射率差进行较大的调控，因此每个量子阱的 Γ_{act} 通常可达 2%–3%，高于 VCSEL 的典型值（约 1%–1.5%）[50]。 Γ_{pc} 则决定了面内光学耦合的整体强度。

1.3.2 第二步：二维布拉格散射与 $\Gamma^{(2)}$ 点共振

当光子晶体的晶格常数 a 满足二阶布拉格条件（ $a \approx \lambda/n_{\text{eff}}$ ，即布里渊区中心 $\Gamma^{(2)}$ 点条件）时，面内传播的四束基本 Bloch 波——沿 $+x$ 、 $-x$ 、 $+y$ 、 $-y$ 方向——被晶格的倒格矢 $m\mathbf{G}_1 + n\mathbf{G}_2$ 耦合在一起。对于折射率对比度较大（如空气/半导体）的正方晶格 TE 模式，这种耦合不只发生在 180° （一维反向散射）方向，还通过高阶 Bloch 波间接发生在 $\pm 90^\circ$ 方向，从而形成**二维相干共振** [50]。这也是 PCSEL 与传统二维 DFB 激光器的本质区别：后者的折射率对比度较低（半导体/半导体），高阶 Bloch 波不参与耦合，四束基本波仅能一维地两两对耦，无法建立二维相干振荡。在 $\Gamma^{(2)}$ 点，四束基本波的群速度降为零，形成驻波式的二维宽面积共振模式。

1.3.3 第三步：辐射通道与法向出光

$\Gamma^{(2)}$ 点位于光锥（light line）之上，这意味着共振模式可以通过一阶衍射与自由空间的辐射波耦合，将能量送出芯片表面。从倒空间的图像来看：每束基本 Bloch 波（波矢 $\sim \pm\mathbf{G}_1$ 或 $\pm\mathbf{G}_2$ ）被一个倒格矢散射回 $\mathbf{k}_{\parallel} = \mathbf{0}$ ，即法向传播的辐射波。辐射常数的大小取决于晶格点的形状和对称性。对于旋转对称的孔形（如圆孔），各辐射分量在远场相消干涉，辐射常数极小，即对称保护型 BIC（bound state in the continuum）效应 [50]。然而，PCSEL 追求的是高亮度输出，因此刻意破缺对称性：通过采用非对称孔形（如三角孔、椭圆孔）或“双晶格”结

构 (double-lattice)，使远场不再完全相消，在目标模上开放可控的辐射通道。由于对称保护 BIC 使辐射常数极小甚至趋零，直接称 PCSEL 为“BIC 激光器”并不恰当——PCSEL 需要主动打破 BIC 条件来获取足够的输出功率 [50]。

1.3.4 第四步：模式选择——从单晶格到双晶格再到 Hermitian/非 Hermitian 联合调控

出光过程的完整性还包括“谁先起振”这一关键问题。在 $\Gamma^{(2)}$ 点通常存在四个带边模式，它们的频率接近但辐射损耗和面内损耗各不相同。PCSEL 通过结构设计使目标基模与高阶模之间产生足够大的**阈值增益差** (threshold margin)，从而实现稳定单模工作。这一思路经历了三代演进 [50]：

1. **非对称单晶格** (2014 年) [20]：采用三角形等非对称孔形，打破 C_4 旋转对称，使不同带边模的辐射损耗产生分化，实现瓦级单模输出。
2. **双晶格** (2018 年) [75]：将一组晶格点相对另一组在 x 、 y 方向各平移约 $\lambda/(4n_{\text{eff}})$ 。由于两组孔的 180° 衍射光程差约为 λ ，发生相消干涉，有效削弱面内 180° 耦合。其结果是基模的场型在空间上展宽，高阶模从谐振区边缘的泄漏损耗相对增大，从而拉开模式间阈值差距。凭借此方案，500 μm 口径器件实现了 10 W 级脉冲单模输出。
3. **Hermitian/非 Hermitian 联合调控** (2023 年) [74]：进一步精调双晶格间距与孔径比例，同时优化面内耦合 (Hermitian 耦合，即 180° 和 90° 衍射) 与辐射耦合 (非 Hermitian 耦合，因伴随辐射损耗)。理论预测口径 3 mm 可实现 50 W to 100 W 单模工作，10 mm 口径可达 500 W to 1000 W [50]。实验上，3 mm 口径器件已实现连续波 50 W 单模输出，发散角仅 0.05° ($1/e^2$)，亮度达 $1 \text{ GW cm}^{-2} \text{ sr}^{-1}$ ，首次在连续波条件下追平了气体和固体激光器的亮度水平 [74]。

1.3.5 小结：一条连贯的出光链

将上述四步合在一起，PCSEL 的出光过程可以概括为一条完整的物理链条：

电注入 $\rightarrow$ 量子阱增益 $\rightarrow$ 面内导模 $\xrightarrow{\Gamma_{\text{pc}}}$ 二维布拉格耦合 ($\Gamma^{(2)}$ 点驻波) $\rightarrow$ 辐射通道 $\rightarrow$ 法向出光
--

因此，PCSEL 不是先形成封闭腔、再在顶部开辐射孔的结构；面内耦合和向外辐射在同一二维周期结构中同时确定。模式选择由晶格对称性、单胞形状和辐射通道共同决定。

1.4 定义 PCSEL：什么必须成立，什么不应混淆

为了避免和相近器件混淆，本书采用如下工作定义。

定义 1.1. 光子晶体面发射激光器 (photonic-crystal surface-emitting laser, PCSEL) 是指满足以下特征的一类半导体激光器：

1. 主要反馈机制来自平面内二维周期结构引起的分布式耦合，而非单纯的垂直 DBR 腔或单轴波导端面反射；
2. 主要输出方向为法向或近法向面发射；
3. 激射模式的存在与选择与布洛赫模（Bloch mode）、 Γ 点附近模、对称性和辐射耦合直接相关；
4. 器件性能不能只由被动冷腔模式决定，而必须与外延层、有源区、电注入和热效应联立理解。

这个定义中需要特别注意最后一条。很多文章在介绍 PCSEL 时会把重点放在光子晶体和能带上；但对于可工作的电注入 PCSEL，纵向层栈、电流路径、载流子输运、热源分布和掺杂损耗都是决定器件成败的主因素，而非“后处理修正项”。

严格表述（难度）

本书会反复坚持六条阅读红线（完整清单见第二章），其中最核心的三条是：被动共振不等于有源阈值，阈值增益不等于阈值电流，无限周期结果不等于有限器件工作结果。只要把这几条守住，很多概念性错误就能避免。

1.5 与相邻器件逐一比较：通过对比限定边界

如果只背定义，读者仍然容易在具体结构前犹豫“这到底算不算 PCSEL”。因此，更有办法的是把 PCSEL 与几类常见器件逐一比较。表1.2 总结了最核心的区别。

表 1.2: PCSEL 与几类相关半导体激光器的核心比较。

器件类型	主要反馈机制	主要出光方向	模选择的主控因素
边发射激光器	波导端面/DFB/DBR 轴向反馈	沿波导方向	波导横模、纵模与端面/光栅条件
VCSEL	垂直 DBR 腔	法向面发射	垂直腔谐振、横向氧化孔或电流孔径
DFB 激光器	一维光栅分布式反馈	多为边发射	一维布拉格条件与相移设计
一般光子晶体激光器	缺陷局域或局域共振	可多样	局域缺陷模或纳米腔共振
PCSEL	二维周期结构面内反馈并与辐射耦合	法向或近法向面发射	Γ 点附近模式、二维对称性、辐射耦合、外延与注入分布

1.5.1 PCSEL、VCSEL 与 RCLED 的定量对照（代表性报道）

为了避免把不同波段、不同驱动模式和不同测试窗口混成一个“万能数字”，表1.3 只列出可追溯的代表性报道，并把“器件性能”和“制造成熟度”分开看。对初学者而言，这张表的用途是建立数量级直觉：为什么三类器件在应用定位上长期并存，而不是彼此完全替代。

表 1.3: PCSEL、VCSEL 与 RCLED 的代表性定量对照（不同文献测试条件不完全一致，表中数字用于数量级认知）。

器件	输出功率（代表报道）	光束质量（ M^2 /发散）	谱宽/线宽（代表条件）	不可直接比较项	制造与成本特征
PCSEL	瓦级连续波单模 [20]；50 W 级连续波单模 [74]	$M^2 \leq 1.1$ [20]；在超大口径单模下 $M^2 \approx 2.36$ 、发散角约 0.05° [74]	文献多报告“单峰窄谱”与单模稳定窗口（线宽受工作点与观测带宽影响，详见第二十四章）	常见近红外高功率连续波条件，孔径与热边界与 VCSEL/RCLED 不同	大面积单模与高亮度潜力高；工艺链通常较复杂（二维晶格精细加工、外延与封装协同）
VCSEL	单发射单元在文中测试条件下可达 28.4 mW [76]	多结阵列发散角约 8° – 16° ；小孔径时平均 M^2 可接近 1 [76]	文中报告窄谱输出并给出随孔径变化的谱宽数据 [76]	小孔径、阵列封装与驱动方式差异显著；与大孔径 PCSEL 不能同口径并表	晶圆级制造和封装生态成熟，已在 3D 感知与车载 LiDAR 形成规模应用 [76]
RCLED	0.5 mW（88 μm 孔径，1300 nm 波段） [58]	窄角锥发射、纤耦合友好 [58]	发射 FWHM 约 5.2 nm [58]	发光机理与激光不同，谱宽与相干性定义口径不可直接对照	常作为短中距系统中的低成本光源替代方案 [58]

从这张表可以看出一条很实用的工程判断：如果目标是“成熟供应链 + 中等功率 + 成本优先”，VCSEL/RCLED 往往更直接；如果目标是“大面积单模 + 极窄发散 + 高亮度”，PCSEL 的结构优势会更明显，但代价是设计与工艺协同门槛更高。新增“不可直接比较项”列的目的，就是把跨器件并表时最容易被忽略的边界条件显式化。

报告规范

跨平台外推三元注释

- **外推来源：**异构器件对照来自不同平台文献（PCSEL/VCSEL/RCLED）[74, 76, 58]
- **边界条件：**并表前需对齐波段、孔径/面积、驱动模式、温度与测量协议
- **不可迁移项：**不可把异构条件下的单点数字直接推导成“器件路线优劣定理”

这里尤其要强调两点。

在强调之前先补一个常被忽略的器件细节：VCSEL 中的氧化孔（oxide aperture）通过选择性氧化高 Al 组分 AlGaAs 形成绝缘环，其开口直径通常为 $3\mu\text{m}$ to $20\mu\text{m}$ ，把有效增益区压缩在小孔径内，从而抑制高阶横模，本质上是空间选择性增益。PCSEL 没有这样的物理孔径，它的横向模式控制依赖二维布拉格散射在大面积孔阵上的相干耦合，模式由散射网络的对称性与损耗分布决定，而不是几何孔径决定。

第一，PCSEL 不是 VCSEL 的“打孔版”。VCSEL 的核心是垂直腔，PCSEL 的核心是平面内耦合。两者都可能面发射，但形成面发射的物理路径不同。VCSEL 中，模式主要先在垂

直方向被 DBR 限定；PCSEL 中，模式先要在平面内通过周期散射建立特定的反馈结构，再通过允许的辐射通道向外输出。

第二，PCSEL 也不是一般“photonic crystal laser”的同义词。很多光子晶体激光器依赖的是小体积极强局域模式，而 PCSEL 的一个重要目标恰恰是做大有效出光面积，同时尽可能维持良好的模式控制。一个靠纳米缺陷局域实现激射的光子晶体纳米腔，原则上不应自动归入 PCSEL。

如果一个结构具有二维周期孔阵，但主要反馈实际上来自上下 DBR 形成的垂直腔，而孔阵只是辅助出光或横向整形，那么把它称为 PCSEL 就需要非常谨慎。是否属于 PCSEL，不看表面形状，而看主要反馈是谁提供的。

1.6 为什么大面积单模困难，而 PCSEL 可能有优势

对于任何激光器，只要横向尺寸变大，允许存在的横向模式数目就会增加。若没有额外机制筛选模式，不同模式的阈值会变得接近，结果就是：

- 起振模式不稳定，容易随注入变化切换；
- 副瓣增强，远场变差；
- 虽然总功率升高，但亮度（brightness）和光束质量（beam quality）不一定改善。

这里的亮度不是总功率的同义词：器件即便输出总功率更高，只要发光面积显著变大或远场发散角明显变宽，亮度仍可能下降。对 PCSEL 来说，困难在于同时扩大有效出光面积、压住远场发散角，并尽量维持单模与单偏振工作。

定量上常用近似写法是

$$B \approx \frac{P}{A_{\text{eff}} \Omega}, \quad \Omega \approx \theta_x \theta_y, \quad (1.1)$$

其中 P 为输出功率， A_{eff} 为有效出光面积， Ω 为远场立体角（小角度时可用发散半角 θ_x, θ_y 近似）。这一定义提醒我们：亮度提升必须同时压缩发散角或降低等效束腰面积，不能只靠总功率上升。

例子

[本书算例] 用 3 mm、50 W 和 0.05° 估算亮度。把出光口径近似成直径 3 mm 的圆，则 $A_{\text{eff}} \approx \pi(0.15 \text{ cm})^2 = 0.0707 \text{ cm}^2$ ；再取 $\theta_x = \theta_y = 0.05^\circ = 8.73 \times 10^{-4} \text{ rad}$ ，得到

$$B \approx \frac{50}{0.0707(8.73 \times 10^{-4})^2} \approx 9.3 \times 10^8 \text{ W cm}^{-2} \text{ sr}^{-1} \approx 0.93 \text{ GW cm}^{-2} \text{ sr}^{-1}.$$

若功率和口径不变而发散角增至 1° ，亮度会降至约 $2.3 \times 10^6 \text{ W cm}^{-2} \text{ sr}^{-1}$ ，相差约 400 倍。这个比较说明，窄发散对高亮度的贡献不是修饰项，而是平方进入分母的核心量。

PCSEL 的潜在优势在于，它并非先做大面积，再事后用损耗把高阶模压下去；相反，它可以在结构层面把模式选择机制提前嵌入平面内反馈网络中。二维光子晶体带来的是“可工

程化的耦合通道”：通过晶格常数、孔形、扰动、双晶格、对称性破缺等设计，研究者可以调节模式分裂、辐射方向性和偏振选择规则。

这里需要进一步限定“可能有优势”的含义。PCSEL 的大面积单模潜力来自**模式组织机制**，而不是天然保证。更准确地说，PCSEL 的大面积单模能力不是 band diagram 单独的性质，而是

$$\text{band} + \text{radiation} + \text{finite size} + \text{injection} + \text{thermal}$$

这五类因素共同决定的联合性质。只要其中任意一环失控，例如有限阵列边缘引入新泄漏通道、电流拥挤把增益峰拖离目标模，或者温升把模式身份重新排序，那么“名义上看起来很优的候选模”就未必能在器件里兑现成稳定单模工作。

1.7 PCSEL 必须同时用三种视角来理解

从本章开始，本书会一直交替使用三种视角。这三种视角是一条连续推理链上的三个层面。

1.7.1 第一种视角：开放周期电磁系统

从这一视角看，PCSEL 是一个与自由空间耦合的二维开放周期系统。关键词是 Maxwell 方程、Bloch 理论、 Γ 点、辐射损耗、复频率、远场和偏振选择。这个层面回答的问题包括：有哪些候选模？哪些模能向法向辐射？辐射损耗大小规律是什么？

1.7.2 第二种视角：半导体有源光学腔

从这一视角看，PCSEL 不只是“有模”，而是“有需要被增益补偿的损耗”。关键词是材料增益、模增益、光学约束因子、内部损耗、阈值增益和模式竞争。这个层面回答的问题包括：有源区与目标模重叠如何？哪一个模式更容易先达到光学阈值？

1.7.3 第三种视角：真实电注入器件

从这一视角看，PCSEL 不再只是光场问题，而是一个由外延层、电流扩展、载流子输运、Joule 热和热致折射率变化共同决定的器件。关键词是漂移-扩散 (drift-diffusion)、Poisson 方程、热方程、注入效率、电流拥挤、温升和热-电-光耦合。这个层面回答的问题包括：阈值电流是多少？温升会不会改变模式排序？器件在多大电流窗口内能稳定工作？

这三种视角之间的关系可以概括为：第一层决定候选模，第二层决定光学阈值排序，第三层决定器件能否达到并维持该排序。更具体的接口链条可以用一条最小的量级传递线来记：

$$Q, \tilde{\omega} \Rightarrow \alpha_{\text{rad}} (= \omega/2Qv_g) \Rightarrow g_{\text{th}} = \alpha_{\text{rad}} + \alpha_{\text{int}} \Rightarrow I_{\text{th}} \text{ (由 } g_{\text{th}} \text{ 进入载流子/热方程)}. \quad (1.2)$$

它提醒读者： Q 与辐射损耗是 L1 的输出，阈值增益是 L2 的入口，阈值电流与工作窗口则必须进入 L3/L4 的输运与热回写。

为了直观展示三重视角之间的数据传递关系，图1.5显示了从 L1 到 L3 的完整工作流。注意每个层级的输入输出量以及热回写反馈路径。

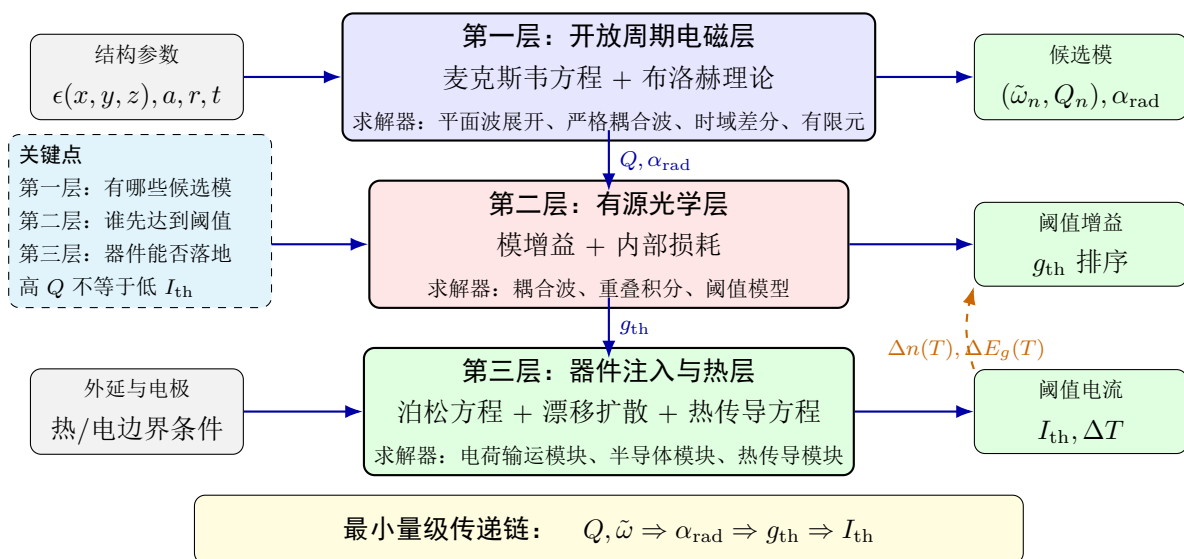


图 1.5: 光子晶体面发射激光器建模的三重视角及其数据传递链。L1（被动电磁层）求解麦克斯韦方程得到候选模的品质因子 Q 和辐射损耗 α_{rad} ；L2（有源光学层）加入内部损耗 α_{int} 和有源区增益特性，计算阈值增益 g_{th} ；L3（器件层）通过电注入和热传导方程求得阈值电流 I_{th} 和工作温度。橙色虚线箭头表示热-电-光耦合反馈：温升会通过折射率温度系数 dn/dT 和带隙温度系数 dE_g/dT 回写到光学性质，可能改变模式排序。底部的黄色框总结了最小量级传递链。右侧青色框强调了三个关键区分：被动共振不等于有源阈值，阈值增益不等于阈值电流，无限周期结果不等于有限器件工作结果。

1.8 关键观测量：本书后面所有章节都在回答哪些量

本书后文经常出现的观测量共同刻画了 PCSEL 的工作状态。

- 品质因子 Q 与光子寿命：属于被动腔信息，用来表征损耗时间尺度；
- 阈值增益 g_{th} ：属于有源光学层，回答“补偿损耗至少需要多少模增益”；
- 阈值电流 I_{th} ：属于器件层，回答“真实器件达到阈值所需的注入电流”；
- 远场与偏振：回答输出是否在目标方向、目标偏振上集中；
- 线宽、相干性、亮度与 M^2 ：表征器件作为系统光源的可用性。

把这些量按层级区分非常重要：比如 Q 很高不自动意味着 I_{th} 很低（红线 #2，见第2.10节）。后文每一章都会回到这些量，但会明确说明自己讨论的是哪一个层级。

表 1.4: 本书最常用观测量的“物理量 → 模型层级 → 求解器 → 实验量”总映射。表中“求解器”表示典型接口，而非唯一实现。

物理量	主要模型层级	常见求解器或方法接口	典型实验对应量
带图、Gamma 点候选模	L1 被动电磁层	PWEM、CWT、RCWA	反射/透射谱中的候选特征、偏振选择趋势
被动 Q 、辐射损耗、上下出光比例	L1 被动电磁层	RCWA、FDTD、FEM、周期频域本征/散射模型	线宽收缩趋势、反射谷宽度、远场与上下出光比
阈值增益 g_{th}	L2 有源光学层	CWT、RCWA 阈值代理、重叠积分模型	起振前后增益需求、阈值附近谱特征变化
阈值电流 I_{th}	L3/L4 器件层	Poisson + 漂移扩散 + 热回写	L-I 曲线拐点、J-V-L、热滚降起点
远场、偏振、 M^2	L1/L4 之间的输出层	RCWA、FDTD、FEM、近远场变换	CCD 远场、偏振仪、束质测量
线宽、相干性、频率噪声	动态/噪声层 + 实验闭环	速率方程、降阶噪声模型、热灵敏度模型	频谱仪、自外差/拍频、频率噪声谱测量
载流子分布、温度场	L3/L4 器件层	半导体电学求解器、热传导求解器（如 CHARGE、Semiconductor、HEAT、Heat Transfer）	热阻、热像、I-V、谱漂移、阈值随温度变化

1.9 本书的主线：从 Maxwell 到工作器件

如果只用一句话概括全书的结构，可以写成下面这条链：

$$\begin{aligned}
 &\text{开放 Maxwell 方程} \rightarrow \text{Bloch 模与}\Gamma\text{ 点} \rightarrow \text{耦合与辐射} \\
 &\quad \rightarrow \text{阈值增益} \rightarrow \text{电注入与热回写} \rightarrow \text{器件工作窗口.}
 \end{aligned}
 \tag{1.3}$$

表 1.5: 式 (1.3) 主线中每一步的最小物理意义对照。

步骤名称	这一跳在物理上做了什么
开放 Maxwell 方程 → Bloch 模与 Γ 点	在周期开放结构里先找“允许存在的候选场型”和它们在倒空间的位置标签。
Bloch 模与 Γ 点 → 耦合与辐射	判断候选模如何通过周期散射彼此耦合，以及哪些通道会把能量送进自由空间。
耦合与辐射 → 阈值增益	把辐射/内部损耗与有源区重叠放到同一阈值条件里，得到“补偿总损耗所需的最小增益”。
阈值增益 → 电注入与热回写	把光学阈值映射到可实现的载流子分布与温升回写，进入器件级约束。
电注入与热回写 → 器件工作窗口	由阈值电流、模式竞争、热稳定性与远场/偏振共同决定可工作的参数区间。

这个顺序对应模型依赖关系。模式分析依赖电磁边界条件；阈值分析依赖损耗分解；阈值电流依赖载流子输运和注入效率；最终远场和稳定性还依赖温升对折射率、增益和模式排序的回写。

1.10 2024–2026 进展速览：从“能激射”走向“系统可用”

面向 2024–2026 的文献趋势，PCSEL 研究重点已从“证明可高亮度单模”继续推进到“系统级可用”：一类工作聚焦高功率调制与相干自由空间通信，直接把调制、远场与链路指标并列报告 [26]；另一类工作通过谐振腔嵌入式设计强化模式选择与噪声控制 [9]；还有一类工作在 BIC 语境下探索更宽角谱的高 Q 机制（如 quasi-BFIC），提升有限尺寸与误差条件下的角度容差 [54]。因此，本书后续结构既讨论模式存在的电磁条件，也讨论这些模式在真实器件和系统约束下是否仍可工作、可制造、可复核。

表 1.6: 2024–2026 PCSEL 前沿对照表：进展–应用场景–设计启示。

进展主题	代表年份/文献	对应应用场景	设计启示
高亮度大面积单模持续推进	2023, [74]	高功率远场照明、材料加工、远距主动照明	需把“低损耗”升级为“模裕量 + 热窗口 + 工艺窗口”联合目标
横向耦合阵列规模化	2024, [16]	大面积相干输出与可扩展阵列	阵列耦合设计应与边缘终止和误差统计一起优化
系统导向的高相干/调制路线	2025, [26]	相干自由空间链路 with 高速调制	性能报告需联动调制协议、链路稳定性与环境边界
谐振腔嵌入式模式约束	2024, [9]	噪声约束和模式稳定增强	线宽、功率与远场应采用多指标协同优化而非单点极值
quasi-BFIC 与有限尺寸边界研究	2025, [54, 52]	宽角容差与鲁棒器件设计	必须显式报告有限尺寸验证，禁止把无限周期结论直接外推到器件

回看本章

PCSEL 值得单独讨论，并非因为名称中包含“光子晶体”，而是因为二维周期电磁学、半导体有源腔理论和真实器件电热问题在同一器件中同时出现。理解 PCSEL，需要从一开始区分并连接这三个层面。

从下一章起，本书先暂停物理叙述，用一章时间统一坐标系、时间因子、模型层级和常用符号。这一步属于符号和层级约定，却是后续推导保持一致的前提：先固定这些约定，后文就不必反复解释复频率虚部、增益和损耗的符号含义。

练习题

1. [基础] 不查资料，用自己的语言说明：为什么“大面积面发射”与“单模高亮度”之间天然存在张力？
2. [基础] 试着分别用一句话概括边发射激光器、VCSEL 和 PCSEL 的主要反馈机制。
3. [进阶] 给出一个你认为“不应称为 PCSEL”的结构例子，并说明理由。
4. [综合]（思路提示：先定层级，再列关键式）解释为什么“高 Q 候选模”不能直接推出“低阈值电流器件”。

延伸阅读

- [28] 中关于周期介质与光子晶体基本概念的章节。
- [10] 中关于半导体激光器工作机制与器件指标的章节。

第二章 全书路线图、符号、坐标与基本约定

读完本章，你应能

- 建立本书统一的坐标系、时间因子与频率约定。
- 统一“模、Bloch 模、 Γ 点、带边、被动共振、阈值增益、阈值电流”等术语含义。
- 明确本书中不同模型层级的关系，避免跨层误读结论。

读前准备 复数表示、向量分析、傅里叶级数与基本晶体学记号。

阅读主线 本章所处模型层级：L1→L4（层级接口总览）

本章不推导新的器件结果，而是固定全书使用的共同语言：先固定 benchmark 教学器件与高频符号工具页，再统一坐标和频域约定，随后说明模型层级与术语边界，最后给出阅读全书时必须守住的六条“红线”。晶格与倒格矢记号在 第4章 首用处建立，归一化约定在 第10章 首用处说明。

2.1 为什么在第二章先停下来统一语言

初学者读 PCSEL 文献时，最先遇到的困难往往是同一个词在不同上下文里意思变了，而非公式本身。比如“阈值”这个词，在某些地方指的是阈值增益，在另一些地方指的是阈值电流；“模式”有时指无限周期的 Bloch 模，有时又指有限器件里的实际振荡模；“TE 模”在二维教材里是严格分类，在 slab 结构里却经常只是近似称呼。若不先限定这些边界，后续公式和结论容易被跨层误读。

因此，本章的作用是给后续所有推导建立语义基础：为全书固定坐标、符号和模型层级。

2.2 本书默认的 benchmark 教学器件

为了避免读者在前半本不断在脑中切换不同器件，本书从这里开始固定一个默认教学参照：方格晶格 air-hole slab PCSEL。除非某章明确声明换了结构，否则本书的几何、对称性、PWEM 候选模、CWT 四波图像、单胞筛选、有限阵列检查与 worked example 都默认围绕这一 benchmark 展开。

它的最小身份是：

- 平面晶格：方格晶格；
- 单胞类型：高折射率背景中的圆形空气孔；
- 纵向结构：有源 slab 波导，上下包层不必严格对称；
- 材料平台：以 GaAs-like 近红外半导体平台作教学参照；
- 目标波长：近红外 $0.8\mu\text{m}$ to $1.0\mu\text{m}$ 量级；
- 目标问题：从 Γ 点候选模出发，逐层推进到有限尺寸、电注入与热回写。

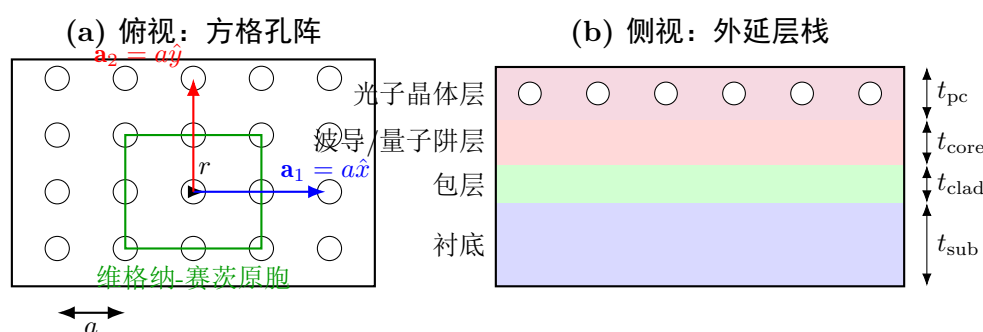


图 2.1: 本书统一教学器件示意。(a) 俯视图给出方格孔阵和几何参数 a, r ；(b) 侧视图给出层栈与常用纵向参数 ($t_{\text{pc}}, t_{\text{core}}, t_{\text{sub}}$ 等)。后续章节若未特别声明，默认都以此结构作为教学对象。

两类贯穿案例及证据边界 为了让后半本的器件章节持续追踪同一组对象，本书把案例分成两个层次。B0 是可计算的教学基准，用于演示从外延到热回写的接口；R1 是有文献依据的性能锚点，用于检验放大后的设计逻辑。两者不能混写：B0 的参数带有 [本书算例] 标记，不代表已发表器件的完整外延配方；R1 只引用论文明确报告的尺寸和性能，不用缺失的层厚、电流或温升数据补齐故事。

这并非说全书结论只对这一类器件成立，而是说：**若某章没有特别声明，我们优先把抽象公式和分类工具投影回这一个统一对象。**这样做的好处是，读者从 第四章、第五章和第十三章读到第二十五章、第二十六章时，脑中追踪的是同一个器件在不同层级的投影，而不是每章换一个新结构。

2.3 前置工具页：全书最高频符号先一次看全

为了避免读者在后文反复翻页找定义，本节把全书最高频、且最容易跨层混淆的量先压成一张工具表。更完整版本见书末术语表，但这里先给出读书时最常需要随手回看的核心对象。

2.4 空间坐标与结构坐标：先分清什么是纵向，什么是面内

本书统一采用直角坐标 $\mathbf{r} = (x, y, z)$ 。其中：

表 2.1: 贯穿全书的双层案例。B0 提供透明的计算输入，R1 提供可追溯的实验性能边界。

代号	身份与尺度	本书采用的数据	在器件章节中的用途
B0	[本书算例] GaAs-like 方格圆孔 PC-SEL, 目标波长 940 nm	上、下 SCH 各约 200 nm, 3-5 个 InGaAs 量子阱; 热学入门估算先用直径 200 μm 的热源。具体组分、掺杂和接触参数必须在真实项目中重新标定	第十七章至第二十章 依次交接 n_{eff} 、重叠因子、局域增益、 $J(x, y)$ 与 $T(x, y, z)$
R1	[文献值] Noda-Yoshida 双晶格放大路线	500 μm 级器件达到约 10 W 脉冲输出; 1 mm 级器件实现 10 W 连续输出与低于 1.23 kHz 的内禀线宽; 3 mm 级器件实现 50 W 连续输出和约 0.05° 发散角 [75, 74, 50]	第二十一章、第二十四章 用这些公开指标检验模鉴别、热补偿和性能报告是否保持在证据边界内

表 2.2: 全书最高频核心符号的前置工具页。表中“典型量级”用于 [教学示意] 与 [本书算例] 对齐，跨平台定量比较需回到各章文献值与边界条件。

符号	所属层级	典型单位/量级	最小含义
$\tilde{\omega} = \omega_r - i\gamma$	L1	rad s^{-1} ($\omega_r \sim 10^{15}$)	开放系统复频率; 实部给出中心频率, 虚部给出衰减率
Q, τ_p	L1/L2	Q 无量纲; τ_p 常为 ps-ns	品质因子与光子寿命; 描述被动或线性工作点下的损耗时间尺度
$g_{\text{mat}}, g_{\text{mod}}, g_{\text{th}}$	L2	cm^{-1} (常见 10^1 - 10^3)	材料增益、模增益、阈值增益; 回答“增益补偿损耗需要到哪一步”
I_{th}	L3/L4	mA-A (随面积变化)	阈值电流; 真实器件达到阈值工作点所需的注入电流
$\Gamma_{\text{opt}}, \Gamma_{\text{loss}}$	L1/L2	无量纲 (常见 10^{-2} - 10^{-1})	目标模与有源区、损耗区的重叠因子
$\alpha_{\text{rad}}, \alpha_{\text{int}}$	L1/L2	cm^{-1} 或 m^{-1}	辐射损耗与内部损耗的等效损耗系数
N, S, T, J	L3/L4	$N: \text{cm}^{-3}$; $T: \text{K}$; $J: \text{A cm}^{-2}$	载流子、光子数、温度、电流密度; 器件动态与多物理场的最小状态量
$E(\mathbf{r})$ 、角谱 $\tilde{E}$	L1	归一化或 V m^{-1}	近场与角谱/远场表示, 用于连接模场、偏振与远场

- z 轴沿外延层法向，也就是器件纵向；
- (x, y) 平面是光子晶体的周期平面，也叫面内平面；
- 当需要把平面内向量单独拿出来时，记作 $\rho = (x, y)$ 。

这个约定看似简单，但会反复用到。因为对 PCSEL 而言，纵向与横向承担的是不同物理角色：纵向层栈决定导模约束、外延不对称和吸收层分布；横向周期结构决定 Bloch 模分类、布拉格耦合与远场工程。后文如果不特别说明，看到”纵向”就默认与外延层栈相关，看到”面内”就默认与二维周期和横向模式选择相关。

2.5 时间因子与复频率：为什么符号约定会影响物理解释

本书统一采用时间因子

$$e^{-i\omega t}. \quad (2.1)$$

这意味着如果一个本征频率写成

$$\tilde{\omega} = \omega_0 - i\gamma,$$

那么对应场随时间的包络衰减为

$$e^{-i\tilde{\omega}t} = e^{-i\omega_0 t} e^{-\gamma t}.$$

因此在本书的约定下， $\gamma > 0$ 代表衰减。这个约定必须记住，因为不同论文可能采用 $e^{+i\omega t}$ ，那时复频率虚部的符号会相反。

2.5.1 角频率、普通频率与波长

除非特别说明，本书默认：

$$\omega = 2\pi f, \quad k_0 = \frac{\omega}{c} = \frac{2\pi}{\lambda}.$$

其中 ω 是角频率， f 是普通频率， λ 是真空波长。若材料色散明显，折射率、介电常数与增益都应理解为频率相关量，但为了避免公式被符号淹没，后文经常不把这种依赖逐项写出。

严格表述（难度）

一旦进入开放系统，本征频率通常是复数。复频率虚部是模式辐射或损耗的时间尺度编码，而非“数值噪声的副产物”。若把开放系统错当成只有实频率的封闭系统，就会直接丢失辐射损耗信息。

2.6 本书的模型层级：为什么同一个问题会有不同答案

贯穿全书的一个核心思想是：PCSEL 研究是一串层级递进的模型，而非单一模型。我们采用四层划分。

表 2.3: PCSEL 四层建模体系的接口规范。以“计算阈值电流”为例说明数据如何在各层之间传递。每一层的输出成为下一层的输入，形成完整的证据链。

层级	核心问题	输入参数	输出结果	典型求解器
L1: 被动电磁	有哪些候选模？哪些能辐射？	- 晶格常数 a - 填充因子 f_{fill} - 刻蚀深度 h - 折 射 率 $n(x, y, z)$	- 复频率 $\tilde{\omega}_n$ - 品质因子 Q_n - 辐射损耗 α_{rad} - 远 场 图 案 $E(\theta, \phi)$	平面波展开法（能带）、严格耦合波法（辐射）、时域有限差分法（瞬态）、有限元法（本征模）
L2: 有源光学	哪个模阈值最低？	- L1 输出的 Q_n - 内部损耗 α_{int} - 增益谱 $g(\lambda, N)$ - 光学限制因子 Γ	- 阈值增益 g_{th} - 模增益排序 - 模式竞争结果 - 偏振选择比	耦合波法、重叠积分模型、速率方程稳态解
L3: 电注入器件	需要多大电流？	- L2 输出的 g_{th} - 外延掺杂分布 - 电极几何形状 - 迁 移 率 模 型 $\mu(N, T)$	- 阈值电流 I_{th} - 载 流 子 分 布 $N(x, y, z)$ - 电 流 密 度 $J(x, y)$ - 串联电阻 R_s	泊松方程 + 漂移扩散方程（电荷输运与半导体模块）
L4: 热 - 电 - 光耦合	能否稳定工作？	- L3 输出的 P_{heat} - 热导率 $\kappa(T)$ - 热边界条件 - 折射率温度系数、带隙温度系数	- 温 度 场 $T(x, y, z)$ - 热阻 R_{th} - 温升 ΔT - 模式排序变化	热传导方程（热传导模块）+ 光学回写迭代

注：实际工作流程中，L1→L2→L3→L4 是单向传递，但 L4 产生的温升会通过折射率温度系数 ($dn/dT \approx 2 \times 10^{-4} \text{ K}^{-1}$ ，以砷化镓为例) 和带隙温度系数 ($dE_g/dT \approx -0.4 \text{ meV/K}$) 回写到 L1 和 L2，形成闭环耦合。当 $\Delta T > 30 \text{ K}$ 时，这种反馈不可忽略。

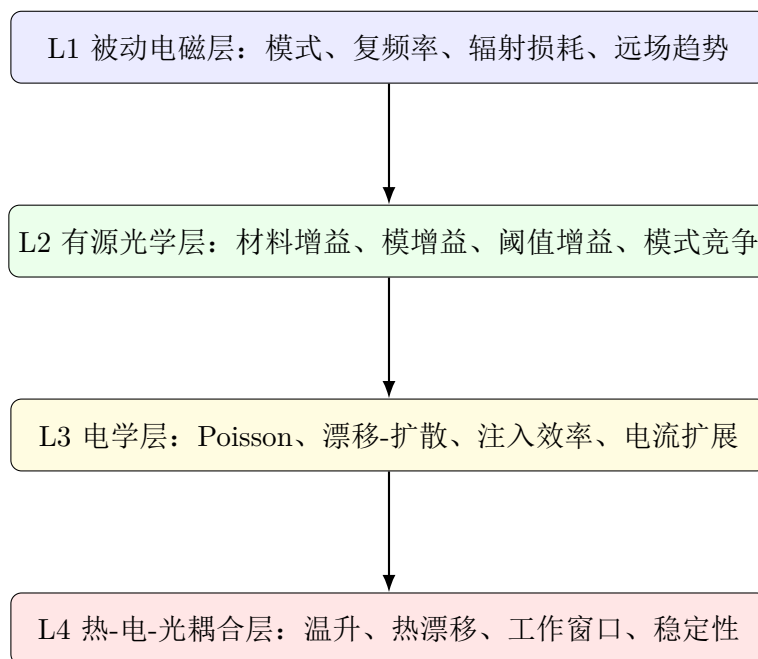


图 2.2: 本书采用的四层建模结构。越往下越接近真实器件工作状态。

表2.3 进一步给出四层模型的接口规范：每一层的输出成为下一层的输入。实际建模时应当按这张表核对数据传递。

这四层的区别，是“谁回答的问题不同”，而非“谁更高级”。

- L1 被动电磁层：回答候选模是什么、频率在哪里、辐射损耗如何；
- L2 有源光学层：回答要补多少增益才能达到阈值、哪些模式更可能竞争；
- L3 电学层：回答注入是否均匀、载流子是否能有效进入有源区；
- L4 热-电-光层：回答真实器件在工作时会不会因为温升和非均匀性而改变模式排序。

直觉框

判断一个具体分析属于哪一层的最直接方法，是看它回答的问题：

若要回答的问题是……	最低需要到哪层
候选模频率、辐射损耗、远场对称性趋势	L1
哪个候选模阈值增益最低	L2
器件阈值电流是多少	L3
高温下哪个模先激射、线宽如何变化	L4

若结论跨越了当前层级（例如用被动 Q 直接报告阈值电流），就是“层级偷换”——这是全书反复提醒的最常见错误。

2.7 几个最容易混淆的术语：先立边界，细节由首用章展开

下面四对术语会反复出现。它们的完整辨析放在各自的首用章节（“模 vs Bloch 模”见第四章，“ Γ 点 vs 带边”见第五章，“被动共振 vs 有源阈值”与“阈值增益 vs 阈值电流”见第七章），高频混淆对照表统一收在第三十一章。这里只先立四条最低边界：

- “模”泛指满足给定方程和边界条件的场分布；“Bloch 模”专指无限周期结构中满足 Bloch 条件的本征场，天然属于 L1 层；
- Γ 点是布里渊区中心（ $\mathbf{k} = \mathbf{0}$ ）；带边是群速度趋近于零的位置——两者可能重合，但不是同义词；
- 被动共振提供候选模，有源阈值决定谁先赢，这个顺序不能反过来；
- 阈值增益 g_{th} 是 L2 层的光学条件，阈值电流 I_{th} 是 L4 层的器件条件，中间隔着电学与热学两层桥。

易错点

“某模式在 RCWA/FEM 中 Q 最高，因此器件阈值电流最低”这句话同时混淆了被动与有源、阈值增益与阈值电流、无限周期与有限器件三个层级。

2.8 如何阅读后续章节：每一章都在回答链条上的哪一环

为了让整本书读起来像一条连续链，而不是二十多篇松散讲义，这里先给出主线。

$$\begin{aligned} \{\varepsilon(\mathbf{r}), \mu(\mathbf{r})\} &\xrightarrow{\text{Maxwell} + \text{Bloch}} \{\tilde{\omega}_m, \mathbf{F}_m\} \\ &\xrightarrow{\text{gain/loss overlap}} g_{\text{th},m} \xrightarrow{\text{DD} + \text{Poisson} + \text{Heat}} I_{\text{th},m}. \end{aligned} \quad (2.2)$$

这条链里每个箭头对应后续一批章节：

- 第三章至第五章 建立 L1 层；
- 第六章、第七章和第十章 解释模耦合与阈值增益；
- 第十七章至第十九章 把外延层和电学结构引入；
- 第二十章、第二十五章 则把这些模块连成真实器件工作流。

理解式 (2.2) 的意义在于：后面任何局部公式都不再是孤立的。它们总是在为某一个箭头服务。

2.9 全书模型依赖图：每章向下一章交付什么

仅知道“后面有哪些章节”还不够，还需要知道每一段向下一段交付什么中间量。表 2.4 给出本书最关键的章节交接关系。

表 2.4: 全书主线中的章节交接表。每一行表示“上一段向下一段交付什么”。

上游章节	向下游交付的核心中间量	下游章节如何使用
第三章至第五章	Bloch 模、 Γ 点候选模、对称性位置、候选频段	第六章、八至十和第十三章 用来组织辐射、偏振和降阶模型基底
第六章、第八章和第九章	辐射通道、偏振选择、BIC / quasi-BIC 约束	第十章、第十四章 用来判断哪些模值得进入阈值代理与开放散射分析
第十章至第十二章	耦合矩阵、微扰灵敏度、候选排序与阈值代理语言	第十三章至第十六章 用来缩小全波搜索空间和解释结果变化
第十三章至第十六章	候选模目录、被动损耗、有限尺寸远场、方法互证结果	第十七章、第二十五章和第二十六章 用来进入器件层与完整建模链
第十七章至第二十章	纵向重叠、模增益映射、注入分布、温度场与工作窗口	第二十一章至第二十四章 用来定义真实器件层面的稳定性与性能
第十九章、第二十章	载流子分布 $N(\mathbf{r})$ 、局部温度 $T(\mathbf{r})$ 与电热回写系数	第十九章、第二十章 中把 $T(\mathbf{r})$ 反馈进迁移率/增益/折射率模型，完成 L3→L4 闭环
第二十五章至第二十七章	正向流程、执行样例、反向诊断	第二十九章、第三十章 用来固化为可复核的研究模板

2.10 全书需要反复区分的六类边界

到这里，可以把本章压缩成六条阅读边界。后文重复触及这些误区时，统一用“红线 #N”半句回指本节，不再整段重述：

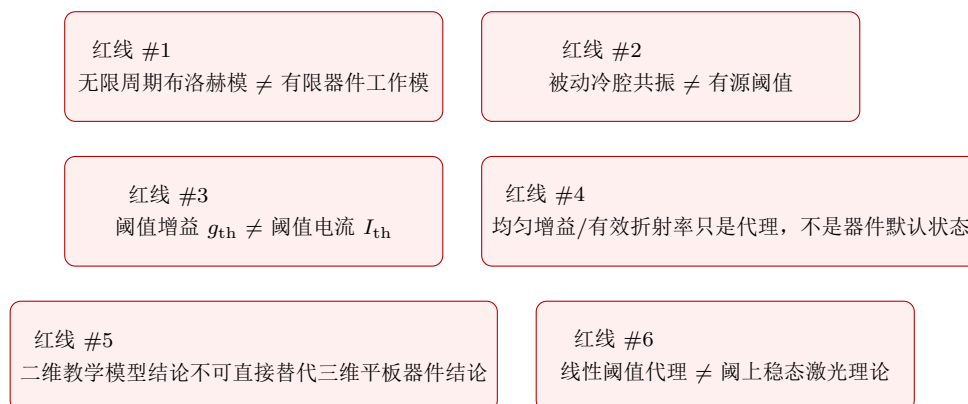


图 2.3: “六条红线”可视化速记图。后文首次触及对应误区时，统一用“红线 #N”作边注提示。

1. 无限周期的 Bloch 模不是有限器件的最终工作模 (L1 $\rightarrow$ 器件层);
2. 被动冷腔共振不是有源阈值 (L1 $\rightarrow$ L2);
3. 阈值增益不是阈值电流 (L2 $\rightarrow$ L3/L4);
4. 均匀增益或有效折射率只是代理模型，不是真实器件默认状态 (L2 $\rightarrow$ L3/L4);
5. 二维教学模型不是三维 slab 结构结论，除非已明确说明哪些纵向自由度被压缩进了等效参数 (2D $\rightarrow$ 3D 层级替换);

6. 线性阈值代理不是阈上稳态激光理论，除非已显式引入有源自洽场或饱和非线性框架 ($L2 \rightarrow$ 阈上动力学)。

如果读者在后续章节中不断回看这六条，整本书会清楚得多。

回看本章

本章的任务是把全书的语义基础固定下来。自此以后，看到坐标、频域符号、 Γ 点、带边、阈值增益、阈值电流等术语，都应自动联想到它们所属的模型层级与物理含义。

练习题

1. [基础] 在本书采用的时间因子约定下，解释为什么 $\tilde{\omega} = \omega_0 - i\gamma$ 表示衰减。
2. [进阶] 试着用自己的话区分“Bloch 模”“被动共振模”“激射模”。
3. [综合] (思路提示：先定层级，再列关键式) 给出一个例子，说明为什么 g_{th} 很低时， I_{th} 仍可能很高。

延伸阅读

- [57] 中关于晶格、倒格矢与对称性记号的章节。
- 建议将本书的符号表与本章一起对照阅读，随时检查各符号所属的模型层级。

第三章 周期介质中的 Maxwell 方程

想象一只“关不上门”的腔：顶部完全封闭，光出不来；顶部一旦打开，能量又必然泄漏。PCSEL 必须在这两者之间工作，所以它的模式从一开始就不是普通封闭腔里的实频本征态，而是带辐射寿命的开放共振。本章要建立的 Maxwell 语言，正是为了把这扇门怎样打开、打开后频率和损耗怎样一起出现说清楚。

读完本章，你能

- 从时域 Maxwell 方程出发，建立本书后续使用的频域本征问题。
- 理解何时可以把 Maxwell 算符看成厄米问题，何时必须视为非厄米开放问题。
- 建立开放 Maxwell 问题的最小函数空间直觉，知道为什么共振模、QNM 与封闭腔本征模不是同一类对象。
- 弄清楚复频率、品质因子、光子寿命和辐射损耗之间的关系。

读前准备 大学电动力学中的 Maxwell 方程、边界条件与向量恒等式。

阅读主线 本章所处模型层级：L1

本章先从 Maxwell 方程写起，再转入频域形式，随后讨论在何种条件下它成为本征值问题。接着我们解释为什么 PCSEL 从根本上是开放系统，并补上开放问题关键的数学直觉：共振模为什么不再属于普通平方可积函数空间、QNM / resonant states 应如何作最小理解。最后再把这一章和后面的 Bloch 理论、阈值分析接起来。

3.1 从时域方程出发：我们到底在求解什么

本章先明确所求解的电磁问题：PCSEL 中的光场由什么方程决定？这些方程中的哪些部分来自材料本身，哪些部分来自边界，哪些部分来自我们选择的近似？

在宏观电磁学层面，时域 Maxwell 方程写为

$$\nabla \cdot \mathbf{D} = \rho_f, \quad (3.1)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (3.2)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (3.3)$$

$$\nabla \times \mathbf{H} = \mathbf{J}_f + \frac{\partial \mathbf{D}}{\partial t}, \quad (3.4)$$

其中 ρ_f 与 $\mathbf{J}_f$ 是自由电荷和自由电流。在本书讨论的腔模、本征模和被动共振问题中，我们通常取无外加源情形，即 $\rho_f = 0$ 、 $\mathbf{J}_f = 0$ 。这并非说器件里没有载流子，而是说在电磁子问题中，载流子的作用被压缩进本构关系，例如介电函数、增益和损耗之中。

为了把 Maxwell 方程变成可计算问题，还需要本构关系。最基本的写法是

$$\mathbf{D} = \varepsilon_0 \varepsilon_r(\mathbf{r}, \omega) \mathbf{E}, \quad \mathbf{B} = \mu_0 \mu_r(\mathbf{r}, \omega) \mathbf{H}.$$

在大多数半导体光学问题中，磁响应不显著，后文通常取 $\mu_r \approx 1$ 。这一近似在近红外到可见光频段（约 400 nm–2 μm ）的非磁性半导体（GaAs、InP、GaN 等）中成立，因为轨道磁响应极弱，对应磁化率通常满足 $\chi_m = \mu_r - 1 \sim 10^{-5}$ 。对 THz 频段或人工磁响应超材料，此近似不成立；本书讨论的材料体系均满足 $\mu_r \approx 1$ 。主要复杂性集中在 ε_r ：它既包含空间分布，又可能包含频率色散、材料损耗、载流子注入和温度依赖。也就是说，PCSEL 的全部复杂性，最终都会不同方式写进 $\varepsilon_r(\mathbf{r}, \omega)$ 之中。

3.2 转到频域：为什么教材几乎总在频域讨论模式

若系统在单频附近工作，或者我们要研究本征模与共振，最自然的做法是采用第二章中约定的时间因子 $e^{-i\omega t}$ ，把场写成

$$\mathbf{E}(\mathbf{r}, t) = \Re [\mathbf{E}(\mathbf{r}) e^{-i\omega t}], \quad \mathbf{H}(\mathbf{r}, t) = \Re [\mathbf{H}(\mathbf{r}) e^{-i\omega t}].$$

代回 Maxwell 方程时，关键步骤如下。以 Faraday 定律为例：

$$\nabla \times \mathbf{E}(\mathbf{r}, t) = -\frac{\partial \mathbf{B}}{\partial t} = -\frac{\partial}{\partial t} [\mu_0 \mu_r \mathbf{H}(\mathbf{r}) e^{-i\omega t}] = i\omega \mu_0 \mu_r \mathbf{H}(\mathbf{r}) e^{-i\omega t}.$$

消去公共因子 $e^{-i\omega t}$ 后即得频域 Faraday 方程。注意正号的来源：时间导数给出 $-(-i\omega) = +i\omega$ ，两个负号相消。若改用 $e^{+i\omega t}$ 约定，则 $\partial_t \rightarrow +i\omega$ ，所有含 $i\omega$ 的符号均翻转——不同文献中频域 Maxwell 方程正负号不同，即源于时间因子约定不同。

对 Ampère 定律也应把符号完整走一遍，而不只是“照抄结果”。从

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \frac{\partial \mathbf{D}}{\partial t} = \frac{\partial}{\partial t} [\varepsilon_0 \varepsilon_r \mathbf{E}(\mathbf{r}) e^{-i\omega t}] = -i\omega \varepsilon_0 \varepsilon_r \mathbf{E}(\mathbf{r}) e^{-i\omega t}$$

可见，在 $e^{-i\omega t}$ 约定下，Ampère 定律右端会直接给出负号。消去公共因子 $e^{-i\omega t}$ 后，频域方程写成

$$\nabla \times \mathbf{E} = i\omega \mu_0 \mu_r \mathbf{H}, \quad (3.5)$$

$$\nabla \times \mathbf{H} = -i\omega \varepsilon_0 \varepsilon_r \mathbf{E}. \quad (3.6)$$

对式 (3.5) 再取一次旋度：

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla \times (\mathrm{i}\omega\mu_0\mu_r\mathbf{H}) = \mathrm{i}\omega\mu_0\mu_r\nabla \times \mathbf{H}.$$

若 μ_r 空间变化，则需写成 $\nabla \times [\mu_r^{-1}(\nabla \times \mathbf{E})] = \mathrm{i}\omega\mu_0(\nabla \times \mathbf{H})$ 。再用式 (3.6) 消去 $\nabla \times \mathbf{H}$ ，得到电场主方程

$$\nabla \times [\mu_r^{-1}(\mathbf{r})\nabla \times \mathbf{E}] - k_0^2\epsilon_r(\mathbf{r})\mathbf{E} = 0, \quad k_0 = \frac{\omega}{c}. \quad (3.7)$$

这是后文常用的频域 Maxwell 方程。形式上它已经非常像一个本征值问题，但若要把它当作本征值问题，还需要说明两个问题：边界条件是什么？介质是否损耗、是否开放？

坡印廷定理与能量流向 回答这两个问题的最直接工具是能量流。利用矢量恒等式 $\nabla \cdot (\mathbf{E} \times \mathbf{H}^*) = \mathbf{H}^* \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{H}^*)$ ，并将式 (3.5) 和式 (3.6) 代入，经过直接计算可得

$$\nabla \cdot \left(\frac{1}{2} \mathbf{E} \times \mathbf{H}^* \right) = \frac{\mathrm{i}\omega}{2} (\mu_0\mu_r|\mathbf{H}|^2 - \epsilon_0\epsilon_r^*|\mathbf{E}|^2). \quad (3.8)$$

对全空间积分并取实部，得到时间平均功率流

$$P_{\text{out}} = \oint_{\partial V} \frac{1}{2} \text{Re}(\mathbf{E} \times \mathbf{H}^*) \cdot d\mathbf{s} = - \int_V \frac{\omega}{2} \text{Im}(\epsilon_r)\epsilon_0|\mathbf{E}|^2 dV. \quad (3.9)$$

若边界为理想导体 ($\mathbf{n} \times \mathbf{E} = 0$) 且介质无损 ($\text{Im} \epsilon_r = 0$)，则两侧均为零，能量既不能流出也不能耗散，只能永远囚禁在腔内；而 PCSEL 表面没有金属屏蔽，光可以从衍射栅耦合出去，坡印亭净通量不为零。因此，判断一个系统是“封闭”还是“开放”，不应只看几何形状，而应看坡印亭矢量的净通量是否为零。下一节将看到，这一能量差别决定了本征频率是实数还是复数。

3.3 算符观点：为什么有时本征频率是实数，有时必须是复数

为了建立后文需要的概念，我们把式 (3.7) 记成算符形式

$$\hat{\mathcal{L}}(\omega)\mathbf{E} = 0, \quad \hat{\mathcal{L}}(\omega) = \nabla \times \mu_r^{-1}\nabla \times - k_0^2\epsilon_r(\mathbf{r}, \omega). \quad (3.10)$$

物理含义：方程式 (3.10) 定义了开放光学系统的本征问题框架。算符 $\hat{\mathcal{L}}(\omega)$ 的第一项 $\nabla \times \mu_r^{-1}\nabla \times$ 描述磁场的旋度能量，第二项 $k_0^2\epsilon_r$ 描述电场在介质中的存储能力。求解 $\det[\hat{\mathcal{L}}(\omega)] = 0$ 即得到系统的本征频率和本征模式。

如果材料无损、无色散，边界又是理想封闭的，那么这个算符可以在适当函数空间上写成自伴形式。此时本征频率是实数，不同模式之间具有良好的正交关系。这是很多电磁教材首先讲授的情形，因为它数学上干净，也有助于建立能量守恒图像。最小类比是一自由度 RLC 共振器：理想无损时本征频率为实数，串入电阻后频率变成“振荡频率 + 衰减率”的复数——厄米对应“无电阻”，非厄米对应“有电阻”。

但 PCSEL 不是这样的系统。它要向自由空间辐射输出；只要允许辐射，系统就不再是封闭腔，模式就不再是严格束缚态，而会变成准束缚态 (quasi-bound state) 或共振态。在后面

的第 9 章中（完整内容见第 8b 章，位于 Ch08 与 Ch09 之间）我们会进一步看到，某些 Γ 点模甚至会以 BIC / quasi-BIC 的形式组织辐射通道。此时本征频率通常写成

$$\tilde{\omega} = \omega_0 - i\gamma,$$

其中 γ 不再是“数值误差”，而是模式能量逸出或耗散的速率编码。因此，PCSEL 的被动模分析从一开始就必须带着“非厄米开放问题”的意识进行。图 3.1 展示了这些模式在复频率平面上的位置和物理含义。

封闭谐振腔与开放 PCSEL 系统的对比		
性质	封闭腔 (金属壁)	开放腔 (PCSEL)
本征频率	实数 ω_n	复数 $\omega_n - i\gamma_n$
模式正交性	标准内积正交	需要双正交基
边界条件	$\mathbf{n} \times \mathbf{E} = 0$	Sommerfeld 辐射条件
Q 因子来源	导体损耗/介质损耗	辐射损耗 + 介质损耗
完备性	离散谱完备	离散谱 + 连续谱

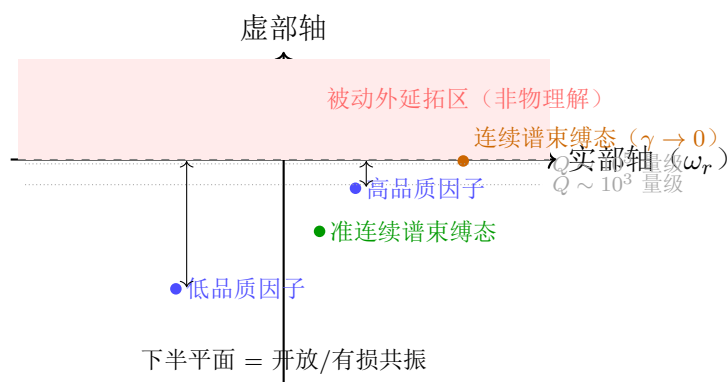


图 3.1: 复频率平面示意。被动开放系统的共振模位于下半平面 ($\gamma > 0$)，距实轴越近表示 Q 越高。连续谱束缚态理想地处于实轴上 ($\gamma = 0$)，准连续谱束缚态则刚刚离开实轴。

对 PCSEL 常见近红外工作点 ($\omega_0 \sim 10^{15}$ rad/s)，由

$$Q = \frac{\omega_0}{2\gamma} \Rightarrow \frac{\gamma}{\omega_0} = \frac{1}{2Q}$$

可得： $Q = 10^3$ 时 $\gamma/\omega_0 \approx 5 \times 10^{-4}$ ， $Q = 10^5$ 时 $\gamma/\omega_0 \approx 5 \times 10^{-6}$ 。因此“离实轴多远算高品质因子”可以直接用这个比例估算，而不必只靠定性描述。

若要把“无损、无色散、封闭边界下 Maxwell 问题可写成自伴形式”说得更正式，较自然的写法不是直接把式 (3.7) 当普通矩阵本征值问题，而是先把它限制在

$$\mathcal{X} = \{ \mathbf{E} \in H(\text{curl}; V) \mid \nabla \cdot (\varepsilon \mathbf{E}) = 0, \text{ 满足封闭边界条件} \}$$

这一约束子空间上，再把它改写成广义本征值问题或等价的加权自伴算子问题。下面的命题与证明只给出**教学化 proof sketch**：它足以解释物理上为什么得到实频率，但不是函数分析层面的完整证明。

命题 3.1. 在无损、无色散、封闭边界条件下，*Maxwell* 本征问题可以写成自伴形式，因此本征频率为实数。

教学化证明思路. 对两个试探场 $\mathbf{E}_1, \mathbf{E}_2$ ，构造加权内积

$$\langle \mathbf{E}_1, \mathbf{E}_2 \rangle_\varepsilon = \int_V \varepsilon_r(\mathbf{r}) \mathbf{E}_1^* \cdot \mathbf{E}_2 dV.$$

在上述约束子空间内，下面只抓住对“为什么封闭无损时频率为实数”关键的一步：旋度项在该加权内积下可通过分部积分变成对称形式。利用向量恒等式 $\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B})$ 对旋度做分部积分，可以把作用在 $\mathbf{E}_2$ 上的旋度“转移”到 $\mathbf{E}_1$ 上：

$$\int_V \mathbf{E}_1^* \cdot (\nabla \times \nabla \times \mathbf{E}_2) dV = \int_V (\nabla \times \mathbf{E}_1)^* \cdot (\nabla \times \mathbf{E}_2) dV - \oint_{\partial V} (\mathbf{E}_1^* \times \nabla \times \mathbf{E}_2) \cdot \hat{\mathbf{n}} dS.$$

若边界为理想导体 (PEC)，则在边界上 $\hat{\mathbf{n}} \times \mathbf{E} = \mathbf{0}$ ，表面积分项为零；若为周期边界，相对两侧的法向相反且场值满足 Bloch 条件，表面项成对抵消。更具体地，在两侧边界 $\mathbf{r}$ 与 $\mathbf{r} + \mathbf{R}$ 上有

$$\mathbf{E}(\mathbf{r} + \mathbf{R}) = \mathbf{E}(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{R}},$$

表面积分的相位因子与法向翻转带来的负号相抵，因而相消。这一步只在 $\mathbf{k}$ 为实向量时成立；若 $\mathbf{k}$ 取复值（表示面内衰减），表面积分不再严格抵消，算符也不再严格自伴。对调 $\mathbf{E}_1$ 和 $\mathbf{E}_2$ 后结果不变，于是

$$\langle \mathbf{E}_1, \hat{\mathcal{L}} \mathbf{E}_2 \rangle_\varepsilon = \langle \hat{\mathcal{L}} \mathbf{E}_1, \mathbf{E}_2 \rangle_\varepsilon,$$

这也是自伴性的定义。自伴算符的本征值必为实数。

但对开放辐射边界， ∂V 延伸到无穷远，出射波携带能量使表面积分不为零且不对称，算符不再自伴。本征频率必须取复数 $\tilde{\omega} = \omega_0 - i\gamma$ ，其中 $\gamma > 0$ 直接编码能量向外辐射的速率。□

3.4 开放 Maxwell 问题的函数空间直觉：为什么它不再是普通闭腔本征值问题

这里需要把一个常被轻描淡写、但对 PCSEL 极关键的数学事实讲清楚。封闭腔本征模通常属于普通的平方可积函数空间，因为场在腔外被边界截断或指数衰减，能量积分是有限的。开放系统中的共振模则不同。若某个模满足出射波条件，其远场在三维直觉上可写成

$$\mathbf{E}(\mathbf{r}) \sim \frac{e^{i\tilde{k}r}}{r} \mathbf{f}(\hat{\mathbf{r}}), \quad \tilde{k} = \frac{\tilde{\omega}}{c}, \quad (3.11)$$

其中 $\tilde{\omega} = \omega_r - i\gamma$ ，且 $\gamma > 0$ 。把指数项拆成振荡部分与增长部分后可得

$$e^{i\tilde{k}r} = e^{i\omega_r r/c} e^{+\gamma r/c}. \quad (3.12)$$

这说明一个反直觉但标准的事实：若把开放共振模按纯出射边界条件延拓到无穷远，它在空间上通常不是衰减而是增长的。这并不表示它“物理上发散”，而是说它对应的是有限时间窗口内的辐射拖尾（因果响应），并不要求在 $t \rightarrow +\infty$ 的真实场中保持该增长；在严格的散射/Green 张量解析延拓意义下，它是共振极点态，而不是普通 Hilbert 空间里的束缚态。

这一步必须讲明白，因为它直接解释了为什么开放系统中的归一化、正交性和微扰理论不能直接照搬封闭厄米问题：问题的根源不只是“损耗让本征值变成复数”，而是函数空间本身已经换了。

3.5 QNM / resonant states 的最小定义

在本书语境中，一个准正规模（quasinormal mode, QNM）或共振态（resonant state），最小可以定义为：在无外加源条件下满足 Maxwell 方程，并同时满足纯出射边界条件的非平凡解，

$$\hat{\mathcal{L}}(\tilde{\omega}_n)\mathbf{E}_n^{\text{R}} = 0, \quad \text{仅允许 outgoing-wave solution.} \quad (3.13)$$

一个常用的等价描述是 Sommerfeld 辐射条件。以一维标量场为例：

$$\lim_{r \rightarrow \infty} r \left(\frac{\partial E}{\partial r} - ik_0 E \right) = 0, \quad (3.14)$$

表示无穷远处只有向外传播分量，没有向内入射分量。这与束缚态要求的 $E \rightarrow 0$ ($r \rightarrow \infty$) 完全不同；正因为 $\tilde{\omega}$ 的虚部存在，QNM 在空间上可指数增长，导致其 L^2 范数发散。数值计算中采用 PML，正是将这一出射条件以可计算的方式实现。这里上标 R 表示右模（right mode），是为了和后面会出现的伴随模区分。对应的复频率 $\tilde{\omega}_n$ 可等价地理解为 Green 张量或散射矩阵的极点位置；也就是说，开放系统的模不是先于散射存在的独立对象，而正是散射响应在复频率平面上的结构化奇点 [13, 60, 3]。

这个定义已经足够支撑本书后面的物理主线。它告诉我们：

- QNM 的本征频率天然为复数；
- QNM 的场型天然带有辐射信息，而不是只描述腔内驻波；
- 只要讨论辐射损耗、准 BIC、远场工程或开放系统微扰，QNM 语言就比封闭腔本征模语言更贴近问题本身。

围绕这一极点态身份重新建立的归一化、伴随模与双正交框架，见第二十八章中“伴随模、双正交与广义归一化”一节，尤其可对照式 (28.48) 及其前后说明。

3.6 为什么 PCSEL 作为开放系统

到这里，我们可以回答一个看似简单、实际上极其关键的问题：为什么本书反复强调 PCSEL 是开放系统？

原因很直接。PCSEL 的目标之一就是法向出光。既然要出光，就必须允许电磁能量离开器件进入自由空间。能量能够离开，模式就不可能是理想封闭腔中的永不衰减本征态。于是：

- 模式频率一般是复数；
- 模式品质因子 Q 与衰减率 γ 相关；
- 模式场的远场形状与辐射边界实现密切相关；
- 仿真中边界条件的处理不再是数值细节，而是物理模型的一部分。

因此，即使只进行能带分析，也需要保留开放性的影响范围。一个在封闭近似下看起来很好的模式，到了开放系统中可能因为辐射耦合太强而损耗很大；反过来，一个能有效法向出射的模式也一定不是完全束缚的。PCSEL 的设计任务始终围绕“控制目标模式的辐射通道和泄漏速率”，而不追求“完全抑制泄漏”。

3.7 复频率、品质因子与光子寿命

设某个被动共振模的复频率为

$$\tilde{\omega} = \omega_0 - i\gamma.$$

若把时间依赖写成 $\Re[\mathbf{E}(\mathbf{r})e^{-i\tilde{\omega}t}]$ ，则模场振幅满足

$$|\mathbf{E}(t)| \propto e^{-\gamma t}.$$

于是，振幅本身的衰减时间尺度是 $\tau_{\text{amp}} = 1/\gamma$ 。但器件分析里更常用的是储能或光子数寿命。由于电磁能量与场强平方成正比，

$$U(t) \propto |\mathbf{E}(t)|^2 \propto e^{-2\gamma t},$$

因此损耗功率满足

$$P_{\text{loss}} = -\frac{dU}{dt} = 2\gamma U.$$

这就解释了为什么品质因子的分母是 2γ 而不是 γ ： Q 比较的是单位振荡周期内储能衰减得有多慢，所以它天然对应能量寿命，而不是振幅寿命。于是可定义光子寿命

$$\tau_p = \frac{1}{2\gamma}. \quad (3.15)$$

进一步，品质因子定义为

$$Q = \frac{\omega_0}{2\gamma} = \omega_0 \tau_p. \quad (3.16)$$

这两个式子在后文会不断出现，因为它们把“复频率”这种看似抽象的本征值，转换成了器件工程师更熟悉的时间尺度和损耗尺度。

但 Q 只是被动冷腔信息，式 (3.16) 是后续阈值分析的输入之一，而不是终点（红线 #2，见第2.10节）。

例子

在数值仿真中，如果用 FDTD 计算一个 PCSEL 冷腔模式的衰减时间常数，再由式 (3.15) 和式 (3.16) 提取 Q ，这一步得到的是冷腔品质因子。它可以用于与 RCWA、FEM 做交叉验证，也可以为阈值增益估计提供被动损耗信息。

3.8 开放边界的实现：数学边界条件与数值边界条件不是一回事

在数学上，开放系统的理想边界即式 (3.14) 的出射波条件 (outgoing-wave condition)：远处只允许向外跑的波，不允许从无穷远无源返回的入射波。对矢量 Maxwell 问题，严格形式会更复杂，但物理含义完全相同。边界条件本身，就是开放系统定义的一部分。

数值计算无法把空间做到无穷大，于是需要用辐射边界条件、端口条件或完全匹配层 (perfectly matched layer, PML) 来近似它。

(a) 辐射边界 (b) 完全匹配层 PML (c) 周期性边界



(d) 边界条件选择指南

边界类型	适用场景	注意事项
辐射边界	解析模型、远场	仅适用于外向辐射
PML	FDTD、FEM 全波仿真	需要足够厚度并验证准确性
周期性边界	FEM、RCWA 单胞问题	验证周期性器件

图 3.2: 开放边界条件的数值实现方式对比。(a) 辐射边界条件：在计算域边界施加出射波条件，要求波只能向外传播；(b) 完全匹配层 (PML)：通过在边界处引入复坐标延拓使外向波指数衰减，理想情况下无反射；(c) 周期性边界条件：利用 Bloch 定理将无限周期结构约化到单个原胞；(d) 边界条件选择指南：根据问题类型（本征模、散射、周期结构）选择合适的边界条件组合。注意 PML 并非物理实体，而是将无限空间折叠进有限计算域的数学技巧。

图3.2对比了几种常用的开放边界实现方式及其适用场景。

PML 的核心思想值得在这里做一个最小介绍。它在计算域边界处引入一层特殊介质，通过对坐标做复数延拓

$$x \rightarrow x + \frac{i\sigma(x)}{\omega},$$

使出射波在该层中指数衰减，但由于变换是连续的、在物理-PML 界面上梯度为零，波在进入该层时不产生反射。需要注意的是， σ/ω 显式依赖频率；对 QNM 这类复频率 $\tilde{\omega}$ 问题，取用哪一个 ω 会影响虚部精度，因此常需做收敛性检查。

PML 参数设置的实践建议

PML 的性能强烈依赖于参数选择。以下是经过验证的设置准则：

1. **PML 厚度**：通常取半个到一个工作波长 ($d_{\text{PML}} \approx \lambda_0/2 - \lambda_0$)。对高 Q 腔 ($Q > 10^4$)，建议增加到 $1.5-2\lambda_0$
2. **吸收剖面**：推荐使用二次或三次渐变剖面 $\sigma(x) = \sigma_{\text{max}}(x/d_{\text{PML}})^m$ ($m = 2-3$)，而非均匀吸收
3. **最大吸收率**： σ_{max} 应满足 $\sigma_{\text{max}}d_{\text{PML}}/c \approx 3-6$ ，对应单程衰减 $e^{-3} \approx 5\%$ 到 $e^{-6} \approx 0.25\%$
4. **收敛性检验**：对高 Q 模式，需系统扫描 PML 厚度和 σ_{max} ，直到 $\text{Re}(\omega)$ 和 $\text{Im}(\omega)$ 的变化小于目标精度（如 $< 0.1\%$ ）
5. **复频率修正**：对 QNM 求解，若初始估计频率为 ω_{est} ，建议在 PML 中使用 $\omega = \text{Re}(\omega_{\text{est}})$ 而非固定实频率，以减少频率失配误差

经验法则：若发现 Q 值随 PML 厚度振荡或不单调变化，通常表明 PML 参数未优化或存在驻波干涉，需增加厚度或调整吸收剖面。

若想看到“为什么会衰减”的最小代数，可取一维右行波 $E(x) = e^{ikx}$ 。代入复坐标延拓后：

$$E \rightarrow \exp\left[ik\left(x + \frac{i\sigma x}{\omega}\right)\right] = e^{ikx} e^{-k\sigma x/\omega}.$$

后面的指数因子 $e^{-k\sigma x/\omega}$ 就是 PML 区域中的吸收包络；它直接展示了“复坐标 $\rightarrow$ 指数衰减”的物理效果。

一旦开放边界实现不当，就会发生两类严重误判：

- 人工反射把本应泄露的模式困住，导致虚假高 Q ；
- 人工吸收过强，把本应存在的模式洗掉，导致错误低估共振质量。

常见错误：开放边界 vs 封闭边界

错误理解：将 PCSEL 当成封闭谐振腔，使用理想导体 (PEC) 边界条件。封闭边界会人为产生反射驻波，导致 Q 值被严重高估（可能差 3-5 个数量级）、模式频率偏移、远场方向图完全错误。

判别方法：如果仿真得到 $Q > 10^6$ 但没有使用 BIC 机制，很可能是边界条件错误。反过来，“复频率的虚部很小，说明数值做得很准”这种说法也没有物理意义：虚部大小首先反映的是模式本身的损耗与开放性，只有在收敛和边界条件正确的前提下，才谈得

上数值精度。

3.9 这一章和后文的关系：为什么先讲 Maxwell，再讲 Bloch

到这里，读者应该能看出：本章建立的是问题类型，而不是最终解法。我们已经知道 PCSEL 的被动光学子问题来自开放系统中的 Maxwell 方程，但还没有利用二维周期性。周期性一旦被引入，问题就可以借助 Bloch 定理进行分类和简化，这也是下一章的主题。

从逻辑上说，本章回答的是“PCSEL 的电磁问题是什么”；下一章回答的是“在周期性存在时，这个问题如何被组织和分类”；再下一章才会进一步问“在这些被组织好的模式中，哪些最有可能成为 PCSEL 关心的候选模”。

回看本章

本章把 PCSEL 的电磁基础明确成了一个开放、通常非厄米的 Maxwell 本征问题。频域 Maxwell 方程给出模式条件，复频率给出损耗时间尺度，开放边界决定模式是否能正确反映真实面发射结构，而 QNM / resonant states 则给出了开放共振模最小且严格的数学身份。后续所有能带、 Γ 点、耦合波和阈值分析，都建立在这一本章奠定的框架之上；其中更严格的归一化、伴随模与非厄米微扰公式将在 第十二章、第二十八章 中继续展开。

练习题

1. [基础] 从频域 Maxwell 方程出发，推导磁场形式的主方程，并说明它与电场形式在数值上可能各有什么优缺点。
2. [基础] 解释为什么“面发射”这一目标本身就迫使 PCSEL 成为开放系统。
3. [进阶] 在本书的时间因子约定下，证明式 (3.16) 成立。
4. [进阶] 利用式 (3.11) 和式 (3.12)，解释为什么开放共振模通常不属于普通平方可积函数空间。
5. [综合]（思路提示：先定层级，再列关键式）试举出一个边界条件实现错误会如何误导 PCSEL 设计判断的例子。

延伸阅读

- [64] 中关于开放边界与时域电磁计算的章节。
- 关于开放共振模、QNM 归一化与散射极点视角，可结合 [60, 34, 3] 阅读。
- 配合 第十四章、第十五章 回看本章，可更清楚地理解数值边界为什么不能被当作纯技术细节。

第四章 Bloch 定理与二维光子晶体基础

读完本章，你应能

- 理解平移对称性为什么会导出 Bloch 定理，而不只是记住 Bloch 形式。
- 掌握二维晶格、倒晶格、布里渊区和高对称点的基本几何意义。
- 认识 slab 结构中 TE/TM 分类的适用条件与局限。

读前准备 第二章、第三章 中的频域 Maxwell 方程、坐标约定和倒格矢记号。

阅读主线 本章所处模型层级：L1

本章先从“周期性如何约束解”讲起，再给出 Bloch 定理的证明思路，随后介绍二维光子晶体几何和布里渊区，最后把这些理想周期概念放回具有有限厚度的 PCSEL slab 结构中，解释 TE/TM 为什么经常只是近似语言。

4.1 为什么周期性如此重要

上一章已经说明，PCSEL 的被动光学问题来自开放系统中的 Maxwell 方程。现在我们加入新的结构信息：介电函数在平面内是周期的。设存在一组晶格平移 $\mathbf{R}$ ，使得

$$\varepsilon(\rho + \mathbf{R}, z) = \varepsilon(\rho, z). \quad (4.1)$$

这条周期性条件直接改变了解空间的结构。它意味着：虽然结构在平面上无限延展，但解并不需要在整个无限空间里“毫无组织地”寻找。周期性会把解空间分解成一族由波矢 $\mathbf{k}$ 标记的子空间。Bloch 定理正是这种分解的数学表达。

对初学者来说，Bloch 定理最容易被误记成一句形式化定义

$$\mathbf{E}_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k} \cdot \rho} \mathbf{u}_{\mathbf{k}}(\mathbf{r}),$$

然后就往下继续算。这样当然可以做题，但难以说明这一形式为何与倒格矢、高对称点和能带图相联系。下面从平移算符说明其来源。

4.2 Bloch 定理的来源：平移算符与共同本征态

考虑晶格平移算符 $\hat{T}_{\mathbf{R}}$ ，定义为

$$\hat{T}_{\mathbf{R}}\mathbf{E}(\rho, z) = \mathbf{E}(\rho + \mathbf{R}, z).$$

若介质满足式 (4.1)，则 Maxwell 算符在平面内与所有 $\hat{T}_{\mathbf{R}}$ 对易。这个对易关系最好显式写一次。设频域算子记为

$$\hat{\mathcal{L}} = \nabla \times \mu_r^{-1}(\rho, z) \nabla \times - \frac{\omega^2}{c^2} \varepsilon_r(\rho, z),$$

则对任意场 $\mathbf{E}$ 有

$$\hat{T}_{\mathbf{R}}\hat{\mathcal{L}}\mathbf{E}(\rho, z) = \hat{\mathcal{L}}(\rho + \mathbf{R}, z)\mathbf{E}(\rho + \mathbf{R}, z).$$

一旦结构满足

$$\varepsilon_r(\rho + \mathbf{R}, z) = \varepsilon_r(\rho, z), \quad \mu_r(\rho + \mathbf{R}, z) = \mu_r(\rho, z),$$

就有 $\hat{\mathcal{L}}(\rho + \mathbf{R}, z) = \hat{\mathcal{L}}(\rho, z)$ ，从而

$$\hat{T}_{\mathbf{R}}\hat{\mathcal{L}}\mathbf{E} = \hat{\mathcal{L}}\hat{T}_{\mathbf{R}}\mathbf{E}, \quad \implies \quad [\hat{T}_{\mathbf{R}}, \hat{\mathcal{L}}] = 0.$$

对易意味着可以寻找同时是 Maxwell 算符和平移算符本征态的场。于是存在一组复数 $\lambda(\mathbf{R})$ 使得

$$\hat{T}_{\mathbf{R}}\mathbf{E}(\rho, z) = \lambda(\mathbf{R})\mathbf{E}(\rho, z).$$

现在利用群性质：

$$\hat{T}_{\mathbf{R}_1}\hat{T}_{\mathbf{R}_2} = \hat{T}_{\mathbf{R}_1 + \mathbf{R}_2}.$$

对同一本征态作用后得到

$$\lambda(\mathbf{R}_1)\lambda(\mathbf{R}_2) = \lambda(\mathbf{R}_1 + \mathbf{R}_2).$$

满足这种可乘性质的连续表示只能写成

$$\lambda(\mathbf{R}) = e^{i\mathbf{k} \cdot \mathbf{R}},$$

其中 $\mathbf{k}$ 是某个二维向量。一个简要论证如下：若 $\lambda(\mathbf{R})$ 连续且满足 $\lambda(\mathbf{R}_1)\lambda(\mathbf{R}_2) = \lambda(\mathbf{R}_1 + \mathbf{R}_2)$ ，则可写 $\ln \lambda(\mathbf{R}) = \mathbf{c} \cdot \mathbf{R}$ ，其中 $\mathbf{c}$ 为常向量。无损周期系统要求 $|\lambda| = 1$ ，因此 $\mathbf{c}$ 必为纯虚数，记作 $i\mathbf{k}$ ，于是得到指数形式。若允许有损或倏逝 Bloch 解，则 $\mathbf{k}$ 可取复数（“复 Bloch 波矢”），对应沿晶格方向的衰减或放大。

于是便有

$$\mathbf{E}_{\mathbf{k}}(\rho + \mathbf{R}, z) = e^{i\mathbf{k} \cdot \mathbf{R}}\mathbf{E}_{\mathbf{k}}(\rho, z), \quad (4.2)$$

再把快速相位因子提出，就得到 Bloch 形式

$$\mathbf{E}_{\mathbf{k}}(\rho, z) = e^{i\mathbf{k} \cdot \rho} \mathbf{u}_{\mathbf{k}}(\rho, z), \quad \mathbf{u}_{\mathbf{k}}(\rho + \mathbf{R}, z) = \mathbf{u}_{\mathbf{k}}(\rho, z). \quad (4.3)$$

图4.1通过一维示例直观展示了 Bloch 定理的物理图像。

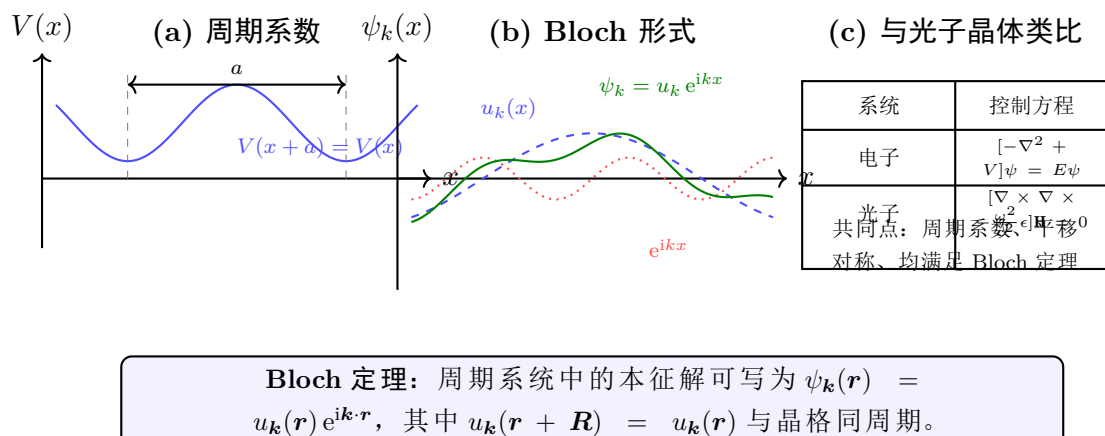


图 4.1: 一维 Bloch 定理直观演示。(a) 周期势场 $V(x)$ (黑色) 与 Bloch 波函数 $|\psi_k(x)|^2$ (彩色); (b) Bloch 波函数的分解: 包络函数 $|u_k(x)|^2$ (红色虚线, 具有与晶格相同的周期性) 乘以平面波因子 (蓝色振荡); (c) 电子系统与光子系统的类比: 尽管控制方程不同 (Schrödinger 方程 vs Maxwell 方程), 但由于两者都是线性波动方程且具有周期系数, Bloch 定理的数学形式完全相同。这使得我们可以将固体物理中的能带理论直接迁移到光子晶体。

Bloch 定理的本质是利用平移对称性把无限系统的解空间按表示分类, 而非经验性地假设“平面波乘周期函数”。

这里还隐含了一个对初学者很重要的前提: 本章讨论的是无损、无限、严格周期的理想背景, 因此我们把 $\mathbf{k}$ 视为实向量, 并要求

$$|\lambda(\mathbf{R})| = 1.$$

这个条件的物理意思是: 把模式平移一个晶格周期后, 只允许它获得相位, 而不允许幅值沿晶格方向系统性增长或衰减。若问题改成有限周期链、有损介质或倏逝 Bloch 解, 则复 Bloch 波矢同样可能出现; 但那已经超出了本章“无限周期模式分类”的任务边界。

数学基础: Bloch 定理的严格数学证明

设周期势场满足 $V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r})$, 其中 $\mathbf{R}$ 为任意晶格矢量。定义平移算符 $T_{\mathbf{R}}$ 使得 $T_{\mathbf{R}}f(\mathbf{r}) = f(\mathbf{r} + \mathbf{R})$ 。

第一步: 证明 $[H, T_{\mathbf{R}}] = 0$ 。 动能项 $-\frac{\hbar^2}{2m}\nabla^2$ 是空间导数算符, 与平移操作对易, 因此只需验证势能项:

$$\begin{aligned} T_{\mathbf{R}}[V(\mathbf{r})\psi(\mathbf{r})] &= V(\mathbf{r} + \mathbf{R})\psi(\mathbf{r} + \mathbf{R}) \\ &= V(\mathbf{r})T_{\mathbf{R}}\psi(\mathbf{r}), \end{aligned}$$

故 $HT_{\mathbf{R}} = T_{\mathbf{R}}H$ 。

第二步: 由量子力学基本原理, 可对易的算符拥有共同本征态。 设 ψ 为 H 和 $T_{\mathbf{R}}$ 的共同本征态:

$$T_{\mathbf{R}}\psi(\mathbf{r}) = c(\mathbf{R})\psi(\mathbf{r}).$$

第三步：确定 $c(\mathbf{R})$ 的形式。由于平移不改变波函数归一化，要求 $|c(\mathbf{R})|^2 = 1$ ，故可写 $c(\mathbf{R}) = e^{i\theta(\mathbf{R})}$ 。又因平移操作的群性质 $T_{\mathbf{R}_1}T_{\mathbf{R}_2} = T_{\mathbf{R}_1+\mathbf{R}_2}$ ，可得：

$$c(\mathbf{R}_1)c(\mathbf{R}_2) = c(\mathbf{R}_1 + \mathbf{R}_2).$$

该函数方程的唯一连续解为 $c(\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}$ ，其中 $\mathbf{k}$ 为与 $\mathbf{R}$ 无关的常矢量。

第四步：构造 Bloch 函数。令 $\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{r})$ ，代入本征方程：

$$T_{\mathbf{R}}\psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot(\mathbf{r}+\mathbf{R})}u(\mathbf{r} + \mathbf{R}).$$

另一方面，由本征值关系：

$$T_{\mathbf{R}}\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{R}}e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{r}).$$

比较两式立得 $u(\mathbf{r} + \mathbf{R}) = u(\mathbf{r})$ ，即 u 具有晶格周期性。证毕。

关键注记：上述证明中假设了 $|\lambda(\mathbf{R})| = 1$ ，这对应于无损耗的保守系统。若介质存在增益或损耗，或者考虑有限结构的倏逝模，则 Bloch 波矢可以是复数： $\mathbf{k} = \mathbf{k}' + i\mathbf{k}''$ 。此时 $|c(\mathbf{R})| = e^{-\mathbf{k}''\cdot\mathbf{R}} \neq 1$ ，但仍满足乘法群同态关系。这也是激光腔中有源区建模时需要推广的理论框架。

Bloch 定理讨论的是无限周期结构中的模式分类。它告诉我们如何组织候选模，但并不自动告诉我们有限器件中哪个模会激射，也不直接给出阈值电流。

4.3 晶格平移与倒格矢记号

若二维晶格的原胞基矢为 $\mathbf{a}_1, \mathbf{a}_2$ ，则倒格矢基矢 $\mathbf{b}_1, \mathbf{b}_2$ 定义为

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}. \quad (4.4)$$

这里写 2π 而不是 1，是为了保证任何晶格平移对应的相位因子 $\exp(i\mathbf{G}\cdot\mathbf{R})$ 对所有整数 m_i, n_i 都严格等于 1，从而 Bloch 条件与傅里叶展开可以直接用整数标记。部分固体物理教材采用 $\mathbf{a}_i \cdot \mathbf{b}_j = \delta_{ij}$ 的约定，两种写法等价但符号不同，查阅文献时需注意对照。任何晶格平移都可写为

$$\mathbf{R} = m_1\mathbf{a}_1 + m_2\mathbf{a}_2, \quad m_1, m_2 \in \mathbb{Z},$$

任何倒格矢都可写为

$$\mathbf{G} = n_1\mathbf{b}_1 + n_2\mathbf{b}_2, \quad n_1, n_2 \in \mathbb{Z}.$$

之所以在一开始就引入倒格矢，是因为 PCSEL 的许多耦合图像本质上都是“相差一个倒格矢的平面波分量之间发生散射”。如果读者不熟悉倒空间记号，后面讨论 Bloch 波、布里渊区、耦合波理论（CWT）和 RCWA 时会不断中断。

对于方格晶格，若 $\mathbf{a}_1 = a\hat{x}$ 、 $\mathbf{a}_2 = a\hat{y}$ ，则由式 (4.4) 立刻得到 $\mathbf{b}_1 = (2\pi/a)\hat{x}$ 、 $\mathbf{b}_2 = (2\pi/a)\hat{y}$ 。这个最简单的例子会在后文多次作为参考系出现。

4.4 为什么波矢只需取第一布里渊区

Bloch 波矢并非任意二维向量都互不相同。若 $\mathbf{G}$ 是倒格矢，则

$$e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{R}} = e^{i\mathbf{k}\cdot\mathbf{R}}e^{i\mathbf{G}\cdot\mathbf{R}} = e^{i\mathbf{k}\cdot\mathbf{R}},$$

因为 $\mathbf{G}\cdot\mathbf{R}$ 总是 2π 的整数倍。于是 $\mathbf{k}$ 与 $\mathbf{k}+\mathbf{G}$ 对应同一组 Bloch 相位条件。因此，我们只需要在倒空间中选取一个基本区域来代表所有不等价的 $\mathbf{k}$ ，这个区域就是第一布里渊区（first Brillouin zone）。

这一步的物理意义在于：它说明在周期结构中，物理上不等价的是波矢模掉倒格矢后的等价类，而非绝对波矢。后面我们画能带、讨论 Γ 点和高对称路径时，都是在这个意义下工作的。

4.4.1 约化区方案与扩展区方案

在固体物理和光子晶体文献中，能带图有两种常见的画法约定：

简约区方案 (Reduced Zone Scheme) 将所有能带都折叠到第一布里渊区内显示。这是最常用的方式，因为它紧凑且突出了布里渊区边界处的能隙 opening。当你看到一条连续的色散曲线在边界处似乎“断开”并跳到另一支能带时，那就是折叠效应的结果。

扩展区方案 (Extended Zone Scheme) 允许波矢超出第一布里渊区，在不同区域内画出相应的能带分支。这种方式更能反映自由粒子色散关系的连续性，但在比较不同材料的能带时不如简约区方便，因为你需要时刻记住哪些区域是等价的。

本书默认采用简约区方案，除非特别说明。理解这两种方案的区别对于正确解读文献中的能带图至关重要。

可能出错处

有些初学者会误以为布里渊区边界是能带的“终点”，或者认为超出边界的波矢就没有物理意义。正确的理解是：第一布里渊区只是一个方便的“基本区域”，任何 $\mathbf{k}$ 都可以通过加减倒格矢映射回这个区域，而物理状态保持不变。因此我们说 Bloch 波矢是在“环面”（torus）上定义的——走出边界会从另一边回来。

直觉框

简约区方案可理解为对等价波矢的代表元选择：越过布里渊区边界的波矢，可通过加减倒格矢映回等价点。

初学者在读图时还会遇到另一个常见困惑：为什么有时文献把所有波矢都折回第一布里渊区，有时又把色散画成跨越多个布里渊区的连续曲线？

这对应两种完全等价的记账方式。

- 约化区方案 (reduced-zone scheme)：把所有不等价的 $\mathbf{k}$ 都压回第一布里渊区，优点是最适合讨论对称性和能带编号；

- 扩展区方案 (extended-zone scheme): 允许 $\mathbf{k}$ 继续延伸到多个倒空间区域, 优点是更直观地显示“原来哪一支平面波被周期势折叠回来”。

阅读时可采用以下顺序: 先用约化区方案理解高对称点和模式分类, 再用扩展区方案建立“散射把哪几个平面波连到一起”的直觉。后者对耦合波理论尤其重要。

4.4.2 区折叠、Umklapp 过程与 Γ 点候选模

要把“画图约定”升级成“物理机制”, 关键是把扩展区中的波矢显式映射回第一布里渊区。对任意扩展区波矢 $\mathbf{k}_{\text{ext}}$, 总能找到某个倒格矢 $\mathbf{G}$ 使

$$\mathbf{k}_{\text{red}} = \mathbf{k}_{\text{ext}} - \mathbf{G}, \quad \mathbf{k}_{\text{red}} \in \text{BZ}_1. \quad (4.5)$$

式 (4.5) 就是区折叠 (zone folding) 的数学表达: 不同扩展区分支在折回后可能落到同一个约化区点, 从而在该点发生简并或近简并重组。

这一过程的物理图像是 Umklapp (折返) 散射: 晶格周期性允许“动量守恒到一个倒格矢”。最小写法是

$$\mathbf{k}_{\text{out}} = \mathbf{k}_{\text{in}} + \mathbf{G}, \quad (4.6)$$

其中 $\mathbf{G}$ 可由晶格提供 (或吸收)。因此在折叠语境下, 看似“不同方向/不同区”的平面波分量, 实际上可能在约化区内属于同一耦合家族。

这也是 Γ 点模式形成的核心入口: 当扩展区中的若干波分量经过式 (4.5) 同时映射到 Γ (即 $\mathbf{k}_{\text{red}} = \mathbf{0}$) 附近时, 它们会在周期耦合下重组为一组区中心候选模。对 PCSEL 来说, 这组模式之所以关键, 是因为区中心模式更容易与法向辐射通道建立相干叠加关系。

对常见方格晶格 PCSEL, 可把二阶 Bragg 工作点直观理解为“主导导模波数与一阶倒格矢同量级”, 即

$$\beta_{\text{slab}} \sim |\mathbf{G}| = \frac{2\pi}{a}. \quad (4.7)$$

在该条件附近, $\pm x/\pm y$ 方向的主导平面波分量经区折叠与周期耦合汇聚到 Γ 邻域, 形成后续第 5 章讨论的区中心模家族与模式选择问题。

例子

[本书算例] 从 940 nm 反推晶格常数。若先取 $a = 280 \text{ nm}$, 则主倒格矢 $G = 2\pi/a \approx 22.4 \mu\text{m}^{-1}$ 。再用 $\beta \approx 2\pi n_{\text{eff}}/\lambda$ 与 $\beta \approx G$, 可反推 $n_{\text{eff}} \approx \lambda/a \approx 3.36$ 。因此“940 nm 配 280 nm 晶格”对应的是一个清楚的折叠数量级, 而不是经验记忆; 真实 slab 仍需用纵向模和三维开放模型修正。

4.5 二维晶格、倒晶格与布里渊区

图 4.3 展示了二维实空间晶格和倒空间晶格的概念示意。图 4.2 进一步给出了正方晶格的完整实例, 包括实空间原胞构造 (左) 和对应的第一布里渊区及高对称路径 (右)。

表 4.1 总结了 PCSEL 中最常见的三种二维晶格类型及其基本性质。

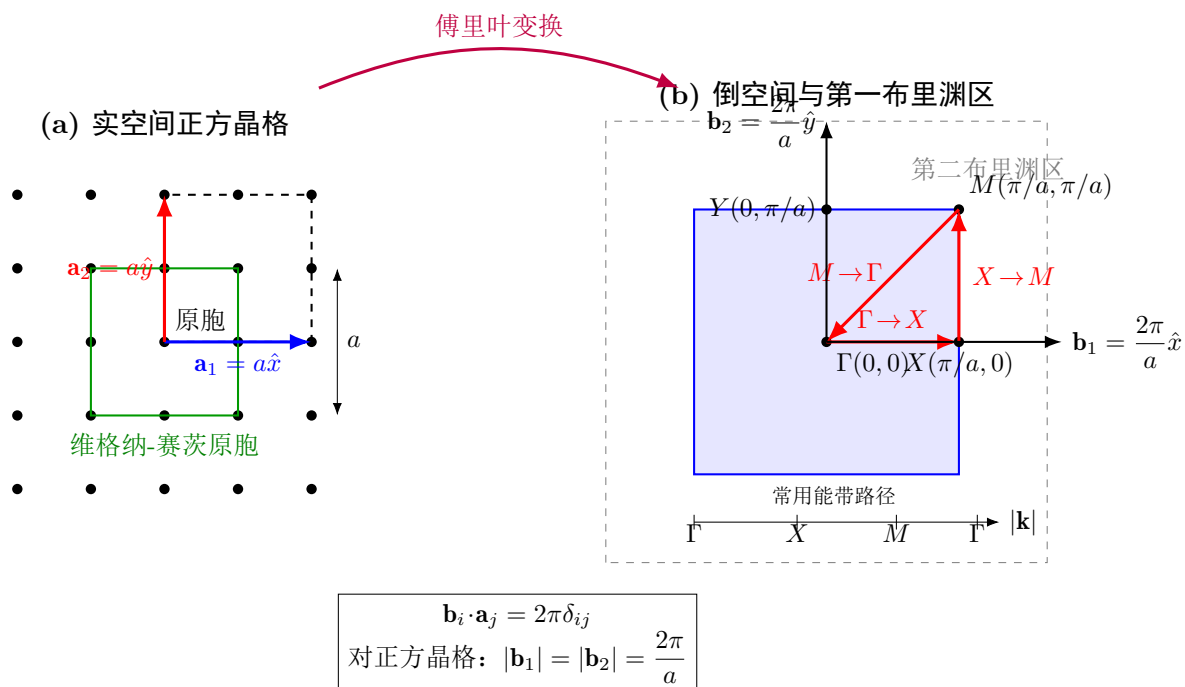


图 4.2: 正方晶格的实空间原胞与对应的第一布里渊区示意图。(a) 左图给出晶格基矢、散射体位置以及 Wigner-Seitz 原胞; (b) 右图给出倒空间中的第一布里渊区、高对称点 Γ 、 X 、 M 及常用能带路径 $\Gamma \rightarrow X \rightarrow M \rightarrow \Gamma$ 。

表 4.1: 常见二维晶格类型及其性质

晶格类型	正方	三角	蜂窝
基矢	$\mathbf{a}_1 = a\hat{x}$ $\mathbf{a}_2 = a\hat{y}$	$\mathbf{a}_1 = a\hat{x}$ $\mathbf{a}_2 = \frac{a}{2}(\hat{x} + \sqrt{3}\hat{y})$	$\mathbf{a}_1 = a\hat{x}$ $\mathbf{a}_2 = \frac{a}{2}(\hat{x} + \sqrt{3}\hat{y})$
倒格矢	正方形	六角形	六角形
第一布里渊区	正方形	正六边形	正六边形
高对称点	Γ, X, M	Γ, K, M	Γ, K, M
点群	C_{4v}	C_{6v}	C_{6v}
典型应用	大多数 PCSEL	三角晶格激光器	光子石墨烯

实空间基矢告诉我们孔阵在物理空间如何重复，倒空间基矢则告诉我们哪些傅里叶分量最先参与散射与耦合。

更正式地，二维倒格基矢由

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij} \quad (4.8)$$

定义。对任意二维基矢 $\mathbf{a}_1, \mathbf{a}_2$ ，可写成

$$\mathbf{b}_1 = 2\pi \frac{\hat{\mathbf{z}} \times \mathbf{a}_2}{|\mathbf{a}_1 \times \mathbf{a}_2|}, \quad \mathbf{b}_2 = 2\pi \frac{\mathbf{a}_1 \times \hat{\mathbf{z}}}{|\mathbf{a}_1 \times \mathbf{a}_2|}. \quad (4.9)$$

式 (4.9) 是后面处理三角晶格时应反复调用的工具，而不是只对正方晶格记一个“旋转九十度再乘 $2\pi/a$ ”的特殊结果。

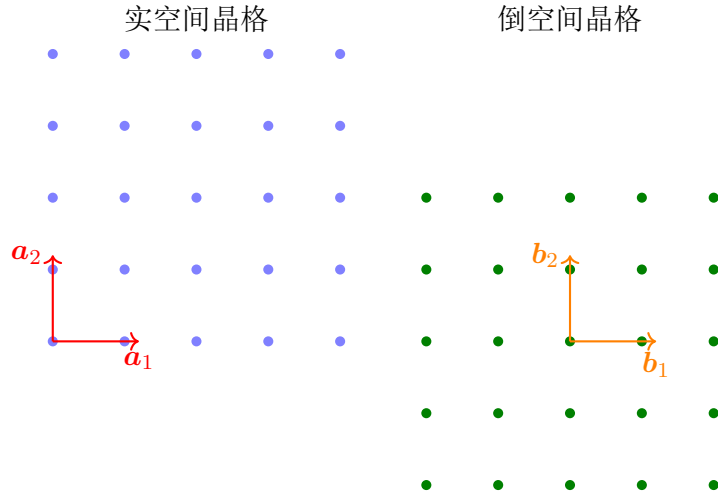


图 4.3: 二维实空间晶格与倒空间晶格示意。

第一布里渊区的构造最好也不要只记最终图形。它的定义是：在倒空间中取所有比其他倒格点更接近原点 Γ 的点所组成的区域，也就是倒空间里的 Wigner-Seitz 原胞。实际作图时，最稳妥的步骤是先画出离原点最近的一圈倒格点，再对连接原点与这些倒格点的线段作垂直平分线，最后取围住原点的最小闭区域。

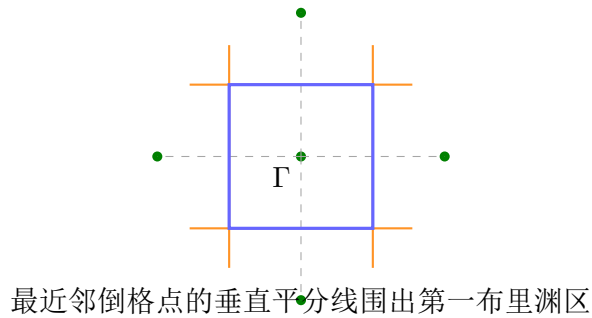


图 4.4: 方格晶格第一布里渊区的构造示意。蓝色区域是围住原点的 Wigner-Seitz 原胞。

图4.5 则给出了标准方格晶格高对称点标注。中心点 Γ 对应 $\mathbf{k} = \mathbf{0}$ ，而边中心 X 与角点 M 分别代表两类不同的特殊相位条件。后面讨论 PCSEL 时， Γ 点之所以格外重要，是因为它对应法向辐射、区中心模以及多组等价平面波分量之间的特殊耦合结构。

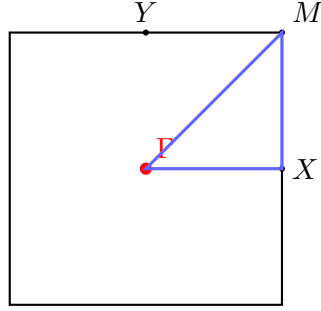


图 4.5: 方格晶格第一布里渊区与标准高对称点标注。常用路径为 $\Gamma \rightarrow X \rightarrow M \rightarrow \Gamma$ 。

4.6 从平面波展开理解能带为什么会出现

如果把上一章的 Maxwell 方程和本章的 Bloch 形式合并起来，可以把周期介质中的模式进一步写成

$$\mathbf{E}_{\mathbf{k}}(\rho, z) = \sum_{\mathbf{G}} \mathbf{c}_{\mathbf{k}+\mathbf{G}}(z) e^{i(\mathbf{k}+\mathbf{G}) \cdot \rho}.$$

这个展开式说明：在周期结构中，一个 Bloch 模并非单一平面波，而是许多相差倒格矢的平面波分量的相干叠加。若没有周期调制，这些平面波彼此互不耦合；一旦周期调制存在，它们就通过介电函数的傅里叶分量彼此混合，于是原本简单的自由色散被“切开”和“重排”，这就是能带结构出现的根本原因。

这一步容易被误解为单纯的“频率对波矢作图”。实际上，能带是平面波在周期介质中相互耦合后的本征值图谱。理解这种耦合图像后，“带隙”“带折叠”“带边”和“Gamma 点简并”都可回到“哪些平面波被耦合在一起”这一问题上。

对一维周期结构，最基本的带隙可以看成两个满足布拉格条件的反向平面波发生耦合后形成的分裂。二维 PCSEL 虽然复杂得多，但逻辑是一致的：只是参与耦合的波分量不再只有一对，而可能是一组等价方向的波。

4.7 二维光子晶体的常见晶格：方格与三角晶格

PCSEL 文献中最常见的二维晶格是方格晶格和三角晶格。选择哪一种，不是审美问题，而会直接影响：

- 高对称点的类型与简并结构；
- 等价散射方向的数量；
- 允许的偏振选择规则与远场工程路径。

方格晶格的优势是几何和对称性直观，很多最小耦合波图像都容易从这里建立；三角晶格则往往在六重对称条件下展现不同的模式简并与辐射特征。对初学者而言，先在方格晶格上建立直觉，再迁移到三角晶格，是比较稳妥的学习路径。

4.7.1 三角晶格速览：后面为什么会频繁出现 K 点与 C_{6v}

三角晶格的一个常用基矢选择是

$$\mathbf{a}_1 = a\hat{\mathbf{x}}, \quad \mathbf{a}_2 = a\left(\frac{1}{2}\hat{\mathbf{x}} + \frac{\sqrt{3}}{2}\hat{\mathbf{y}}\right). \quad (4.10)$$

将它代入式 (4.9)，得到

$$\mathbf{b}_1 = \frac{2\pi}{a}\left(1, -\frac{1}{\sqrt{3}}\right), \quad \mathbf{b}_2 = \frac{2\pi}{a}\left(0, \frac{2}{\sqrt{3}}\right). \quad (4.11)$$

此时第一布里渊区是正六边形，最常见的高对称点是区中心 Γ 、边中心 M 和顶点 K 。在标准笛卡尔坐标系下，这些点的坐标通常取为

$$\Gamma = (0, 0), \quad M = \left(\frac{\pi}{a}, \frac{\pi}{\sqrt{3}a}\right), \quad K = \left(\frac{4\pi}{3a}, 0\right). \quad (4.12)$$

注意 M 点是第一布里渊区六边形边的中点，而 K 点是顶点。有时也会见到用倒格矢基底的约化坐标表示，例如 $K = (1/3, 1/3)$ ，这取决于具体约定。这一节的任务是给后文一个最小几何坐标系：当第八章、第九章讨论 C_{6v} 小群、双重简并和六重对称散射时，读者至少知道这些标签对应的倒空间位置是什么，而非立刻把三角晶格全部算透。

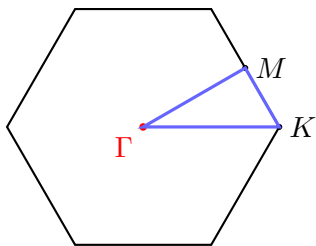


图 4.6: 三角晶格第一布里渊区示意。最常用的高对称路径是 $\Gamma \rightarrow M \rightarrow K \rightarrow \Gamma$ 。

4.7.2 初读高对称路径时应看什么

很多书或软件会直接给出一条高对称路径，比如方格晶格中的 $\Gamma \rightarrow X \rightarrow M \rightarrow \Gamma$ 。常见误区是把注意力全部放在“哪条曲线高、哪条曲线低”，却忽略了更重要的问题：

1. 哪些点是高对称点，为什么只看这些点；
2. 某条带在这些点附近是否变平，意味着哪些耦合增强；
3. 哪些带在高对称点重合，意味着可能存在简并与后续可控分裂。

因此，第一次看高对称路径，不要急着评价“哪条带更好”，而要先学会读出“哪里在发生模式重组”。

若在方格晶格中只保留最主要的四个平面波分量，许多 Γ 点耦合现象就能得到很清楚的图像。这也是为什么很多教材和早期解析模型喜欢从方格晶格开始，而不是一上来就讨论所有可能晶格。

4.8 把二维理想图像放回真实 PCSEL：slab 结构会带来什么变化

到目前为止，我们讨论的 Bloch 理论主要是“平面内无限周期”的语言。但真实 PCSEL 是具有有限厚度的 slab 结构，而非严格二维材料。也就是说，场不仅在平面内受周期调制，还在 z 方向受外延层栈约束。这一事实会带来两个关键后果。

第一，实际模式是全矢量三维模式。即使在某些参数区间内人们习惯说“TE-like”或“TM-like”，那也通常意味着“主要电场分量或主要磁场分量具有某种偏向”，而不是二维教科书中严格分离的纯 TE/TM 解。

第二，slab 的存在允许导模、泄露模和辐射模共存。某个在二维理想模型中看似被很好分类的 Bloch 模，放到有限厚度 slab 中以后，可能会通过垂直辐射通道与自由空间耦合，从而成为 PCSEL 需要追踪的候选模。

二维 Bloch 理论的作用是告诉我们“平面内有哪些组织良好的模式家族”；slab 结构则提醒我们“这些家族在真实三维器件里会怎样与纵向约束和辐射通道重新混合”。PCSEL 正是在这两层之间工作。

4.9 TE/TM 分类为什么常常只是近似

在严格二维、无厚度、标量化处理的模型中，TE 与 TM 可以作为清晰的极化分类。但在真实 PCSEL slab 中，下列因素都会破坏这一理想图像：

- 高折射率对比导致边界处矢量分量强烈耦合；
- 周期孔形和垂直不对称使场分量不再简单分离；
- 辐射通道要求场同时满足面内周期和垂直出射条件。

因此，后文若使用“TE-like”或“TM-like”这样的词，应理解为工程上的近似标签，而不是严格数学分类。本书采用一个可操作的判别标准：若模式电场能量中 E_z 分量占比 $< 5\%$ ，称为 **TE-like**；反之若 H_z 分量占比 $< 5\%$ ，称为 **TM-like**。这里的“like”提醒读者：只要 slab 上下不对称， E_z 和 H_z 都不可能严格为零，真实模式是混合极化的。这个提醒很重要，因为许多关于偏振和远场的误判，都源自把近似 TE/TM 标签当成了真实矢量模的完整描述。

4.9.1 一个对初学者更安全的学习顺序

如果读者刚接触 PCSEL，不必一上来就放弃所有 TE/TM 直觉。更安全的做法是分三步：

1. 先用二维标量或 TE-like 图像理解带结构、带隙和高对称点；
2. 再用全矢量 slab 模型检查哪些结论仍然成立，哪些开始偏移；
3. 最后在有限尺寸和开放边界模型中验证偏振与远场。

这种“先建立直觉，再主动拆掉过强近似”的路径，比一开始就把所有三维矢量细节压到新手身上更有效。

易错点

把二维平面波展开中的 TE/TM 标签直接搬到真实有限厚度电注入器件上，是 PCSEL 建模里非常常见的错误。尤其当外延不对称、金属邻近损耗或偏振工程很重要时，这种错误会直接影响模式判定。

常见错误：把二维 TE/TM 当成三维 slab 的严格本征极化

常见误读：在二维模型里标了 TE/TM，就直接把它当成真实器件中的“纯 TE/纯 TM 模”。

正确做法：把二维结果仅作为 TE-like/TM-like 的初筛标签，再用全矢量三维模型（含开放边界与有限尺寸）复核偏振、远场和阈值排序。

典型后果：

- 把混合极化模误判为“纯偏振模”
- 误判法向辐射允许性与偏振选择规则
- 误把二维排序当成器件阈值排序

4.10 Bloch 理论能告诉我们什么，又不能告诉我们什么

到这一章结尾，必须把 Bloch 理论的能力边界说清楚。

它能告诉我们：

- 周期结构中模式如何按 $\mathbf{k}$ 和对称性分类；
- 第一布里渊区内哪些高对称点值得重点关注；
- 某些简并与选择规则为何出现。

它不能单独告诉我们：

- 有限器件边缘效应会不会改变模式排序；
- 有源区增益分布是否偏向目标模；
- 电注入和温升下哪一个模最终先起振并保持稳定。

因此，Bloch 理论是模式分类的起点，而不是 PCSEL 器件结论的终点。

4.11 通向下一章：为什么 Gamma 点成为重点

本章已经说明：二维周期结构中的模式可以用 Bloch 波矢分类，而 Γ 点是布里渊区中心的特殊位置。下一章将进一步回答：为什么在 PCSEL 中， Γ 点附近的模式经常成为重点关注对象？简并、对称性、辐射耦合和有限尺寸修正又会怎样共同决定候选激射模？

回看本章

Bloch 定理把无限周期结构的求解组织成了按波矢分类的问题，并由此引出倒格矢、布里渊区和高对称点等核心概念。对 PCSEL 而言，这一步的意义在于建立候选模的“目录系统”。但这还只是目录，目录之后还要讨论哪些模式与辐射、增益和器件工作条件相匹配。

练习题

1. [基础] 对方格晶格和三角晶格，分别写出一组可能的实空间基矢与倒格矢基矢。
2. [基础] 从平移算符的角度，重述一次 Bloch 定理的证明思路。
3. [基础] 证明 $\mathbf{k}$ 与 $\mathbf{k} + \mathbf{G}$ 对应相同的 Bloch 相位条件。
4. [进阶] 解释为什么 Γ 点的重要性不仅仅来自“群速度小”，还来自模式对称性和辐射通道。
5. [综合]（思路提示：先定层级，再列关键式）给出一个 slab 结构中 TE/TM 近似可能失效的具体场景。

延伸阅读

- [28, 57] 中关于 Bloch 定理与二维光子晶体的基础章节。
- 读完下一章后再回看本章，通常会更容易理解 Γ 点为何成为 PCSEL 模选择的核心位置。

第五章 光子能带基础、Gamma 点物理与模选择

读完本章，你能

- 从零理解什么是光子晶体能带，明确它为什么是周期 Maxwell 本征问题的结果，而不是一张“频率随波矢变化”的普通曲线图。
- 准确区分光子晶体能带与半导体材料电子能带，理解二者的联系、差异和在 PCSEL 中重新相遇的位置。
- 学会阅读能带图，知道应该看哪些信息，哪些信息不能从能带图中直接读出。
- 弄清楚 Γ 点、带边、简并、辐射耦合、被动共振模和真实激射模之间的关系。
- 为第十三章的平面波展开法 (PWEM) 与后续 RCWA/FDTD/FEM 建模建立清晰的概念入口。

读前准备 建议已掌握第三章、第四章中关于开放 Maxwell 问题、Bloch 定理、布里渊区和 slab 近似的内容。

阅读主线 本章所处模型层级：L1

上一章已经说明，二维周期结构中的场是按照 Bloch 波矢和对称性组织起来，而非随意分布的。现在自然要追问：既然每一个 Bloch 波矢都对应一组允许模式，那么把这些模式的本征频率沿倒空间路径排出来之后，我们到底得到了什么？这就是光子能带。为避免新手把能带图误读成器件工作图，本章先从能带的最基础概念讲起，解释它和电子能带在数学上相似、在物理上又有哪些关键区别；随后用一个正方晶格圆孔二维光子晶体的最小例子建立直觉，说明高对称路径、归一化频率和带隙是什么意思；再专门讲“怎么看能带图”，把群速度、简并、带边、光锥与 PCSEL 候选模筛选联系起来；最后明确指出：能带图只能给出候选模目录，不能直接给出阈值电流、热稳定性和真实激射模。第十三章将把这里的概念完全落实到 PWEM 计算流程中。

5.1 为什么在 Bloch 之后必须讲能带

Bloch 定理告诉我们，若介质在平面内满足周期性，那么模式可按波矢 $\mathbf{k}$ 分类。到这里为止，我们已经知道怎样给无限周期结构中的候选模“命名”。但命名还不够。接下来必须知道，

在每一个 $\mathbf{k}$ 处，到底允许哪些本征频率出现；随着 $\mathbf{k}$ 变化，这些频率怎样连续演化；在哪些高对称点会出现简并、分裂和平带。把这些问题合在一起，得到的就是光子能带结构。

因此，能带图是参数化本征值问题的可视化。参数是 Bloch 波矢，输出是本征频率。难点不在于这个定义本身，而在于它在 PCSEL 里意味着什么。很多初学者一看到能带图，就会下意识地问“哪条带最适合器件工作”“哪条带最低阈值”。这个问题跨越了模型层级。能带图首先回答的是哪些模式可作为候选，而不是哪一支最终激射。

直觉框

能带图可作为候选模目录：它给出频率、对称性和简并关系，但不能替代辐射损耗、增益重叠、电流路径和热反馈分析。

5.2 先把两个“能带”分开：材料电子能带与光子晶体能带

“能带”这个词在 PCSEL 教材里会反复出现，但它至少指向两类不同对象。第一类是半导体材料的电子能带：电子和空穴在原子周期势、异质结台阶、量子阱与应变作用下形成的可占据能级结构。它决定带隙、载流子束缚、跃迁能量、TE/TM 增益和温度漂移。第二类是光子晶体的光子能带：电磁场在介电常数周期调制中形成的允许模式频率结构。它决定 Bloch 模、带边、群速度、带隙、辐射通道和 PCSEL 的候选腔模。

两者的联系是真实的：它们都来自周期算符，都可用 Bloch 定理把无限周期问题分解成由波矢标记的本征问题；带隙也都可从 Bragg 散射、简并解除和本征值分裂的角度理解。因此，固体物理中的电子能带直觉对理解光子晶体很有帮助。

但两者的差异同样关键。材料电子能带的本征值是能量 $E_n(\mathbf{k})$ ，本征态会被电子按费米统计占据；光子晶体能带的本征值是频率 $\omega_n(\mathbf{k})$ ，描述的是电磁模式是否允许存在、如何传播、如何辐射，并不表示某些“光子态已经被填满”。更直白地说：电子能带回答“载流子能待在哪里、如何跃迁”，光子能带回答“电磁场能以什么模式存在、如何反馈和出光”。

表 5.1: 本书中两类“能带”的最小辨析表。

对比项	半导体材料电子能带	光子晶体光子能带
控制方程	周期势场中的 Schrödinger / $k \cdot p$ / 有效质量方程	周期介电结构中的 Maxwell 本征方程
本征对象	电子、空穴波函数或包络函数	电磁场分布 $\mathbf{E}_{n\mathbf{k}}, \mathbf{H}_{n\mathbf{k}}$
纵轴物理量	能量 E ，常以 eV 表示	频率 ω 或归一化频率 $\Omega = a/\lambda$
占据意义	有费米能级、载流子分布、Pauli 排斥和准费米能级分离	无费米面和“填满”的概念；重点是模式存在、传播、驻留与辐射
PCSEL 中的角色	决定量子阱增益峰、带阶、载流子泄漏、TE/TM 增益和温度敏感性	决定候选 Bloch 模、 Γ 点带边、群速度、光锥耦合和面内反馈
典型误读	把光子晶体带隙误当成半导体材料带隙	把一条漂亮的光子能带直接当成最终激射模

在本书中，若讨论“量子阱能带”“导带/价带台阶”“准费米能级分离”，默认指材料电子

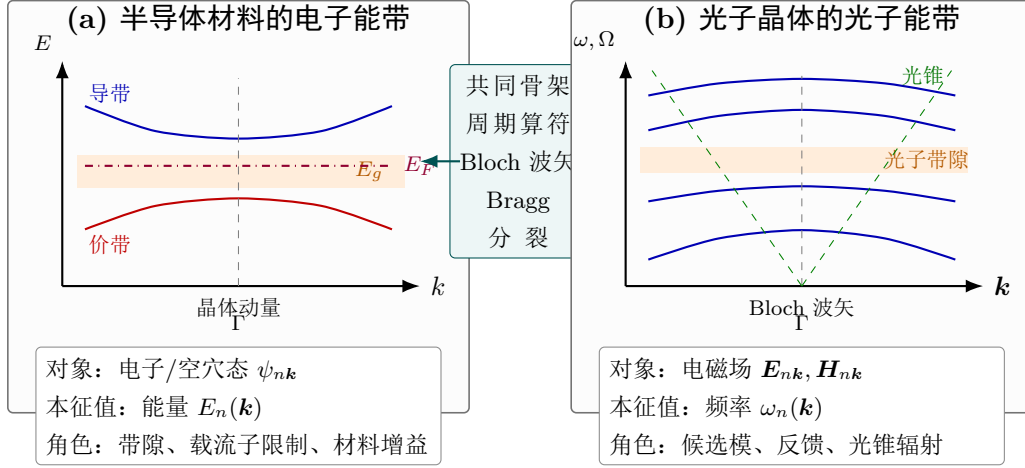


图 5.1: 材料电子能带与光子晶体光子能带的联系和区别。二者共同依赖周期性与 Bloch 定理，但本征对象、纵轴物理量、占据意义和器件作用不同。对 PCSEL 而言，材料电子能带主要通过量子阱增益峰、载流子限制和偏振增益进入问题；光子晶体能带则通过 Γ 点候选模、带边平坦度、光锥和辐射通道进入问题。真实激射只有在这两条链与损耗、注入、热反馈同时匹配时才出现。

能带，详见第十七章至第十九章。若讨论“ Γ 点带边”“光子带隙”“光锥”“PWEM 能带图”，默认指光子晶体能带。PCSEL 的难点恰恰在于二者不能割裂：光子能带给出一个候选光学频率，材料电子能带必须在同一频率附近提供足够净增益，最后还要通过有限尺寸、注入和热反馈检验。

5.3 什么是光子晶体能带

设周期 Maxwell 本征问题写成

$$\hat{\mathcal{M}}(\mathbf{k}) \Psi_{n\mathbf{k}} = \omega_n(\mathbf{k}) \Psi_{n\mathbf{k}}, \quad (5.1)$$

其中 $\hat{\mathcal{M}}(\mathbf{k})$ 表示在给定 Bloch 波矢 $\mathbf{k}$ 下的本征算符， $\Psi_{n\mathbf{k}}$ 代表第 n 条本征模， $\omega_n(\mathbf{k})$ 是对应本征频率。对于每一个固定的 $\mathbf{k}$ ，通常存在一串离散本征值；把这些本征值在倒空间中连续追踪，就得到一组函数 $\omega_n(\mathbf{k})$ 。每一个这样的函数分支就称为一条光子能带 (photonic band)。

与上一节的辨析对应，这里的“带”是 Maxwell 模式目录，而非载流子占据带。它受介电常数的空间周期、极化、矢量性和开放辐射边界共同控制。特别是对 PCSEL 的 slab 结构而言，光锥和垂直辐射通道会让某些带与自由空间耦合，材料电子能带图中没有这一开放辐射对应物。

光子晶体能带常采用归一化频率表示，而不是用绝对能量。最常见的写法是

$$\Omega = \frac{\omega a}{2\pi c} = \frac{a}{\lambda}, \quad (5.2)$$

其中 a 是晶格常数， λ 是真空波长。式 (5.2) 的根源是 Maxwell 方程的几何相似性：若将结构所有尺寸按比例缩放 $\mathbf{r} \rightarrow s\mathbf{r}$ ，而材料分布不变，则频率必须按 $\omega \rightarrow \omega/s$ 缩放，从而无量纲量

$\omega a/c$ 保持不变。这意味着同一张 Ω - $\mathbf{k}$ 能带图可直接用于任意尺度的几何相似结构，只需把 Ω 换算回实际 a/λ 即可。

使用该定义时，应能在归一化频率与实际几何尺度之间互相换算。比如若某一候选模对应

$$\Omega = 0.29,$$

而目标真空波长是 966 nm，则需要的晶格常数约为

$$a = \Omega \lambda \approx 0.29 \times 966 \text{ nm} \approx 280 \text{ nm}.$$

反过来，若先给定 $a = 280 \text{ nm}$ ，则这条归一化频率对应的目标波长就是

$$\lambda = \frac{a}{\Omega} \approx \frac{280 \text{ nm}}{0.29} \approx 966 \text{ nm}.$$

因此能带图虽然看上去无量纲，却能直接服务于器件几何尺寸的前期反推。

5.4 一个一维玩具模型：带隙为什么会打开

为说明周期耦合如何打开带隙，可先看一维 nearly-free-photon 模型。设介电函数只含一个主导傅里叶分量

$$\varepsilon(x) = \bar{\varepsilon} + \Delta\varepsilon_G e^{iGx} + \Delta\varepsilon_{-G} e^{-iGx},$$

并考虑接近布里渊区边界 $k \approx G/2$ 的两支平面波

$$e^{ikx}, \quad e^{i(k-G)x}.$$

在这一最小两波基底里，周期调制把原本简并的两支平面波耦合成一个 2×2 本征问题：

$$\begin{bmatrix} \bar{\omega}(k) & \kappa \\ \kappa^* & \bar{\omega}(k-G) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \omega \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}. \quad (5.3)$$

这里的 κ 是周期介电调制在这两支反向波之间产生的矩阵元，而非凭空命名的常数；在这个最小玩具模型里，它正比于主导 Fourier 分量 $\Delta\varepsilon_G$ 与两支基波的重叠。

在精确 Bragg 条件 $k = G/2$ 处，两条未耦合频率相等，记为 $\bar{\omega}_B$ 。此时耦合系数可以用介电函数的主导 Fourier 分量显式估计：对 $\varepsilon(x) = \bar{\varepsilon} + 2\Delta\varepsilon \cos(Gx)$ ，两波基底 $\{e^{ikx}, e^{i(k-G)x}\}$ 之间的矩阵元为

$$\kappa = \frac{\omega_B^2}{2k^2 c^2} \Delta\varepsilon \approx \frac{\Delta\varepsilon}{2\bar{\varepsilon}}, \quad \bar{k}^2 = \frac{\omega_B^2 \bar{\varepsilon}}{c^2}.$$

因此带隙宽度

$$\Delta\omega_{\text{gap}} = 2|\kappa| \approx \frac{\omega_B \Delta\varepsilon}{\bar{\varepsilon}} = \omega_B \frac{\Delta n}{\bar{n}},$$

这里最后一步使用 $\Delta\varepsilon \approx 2\bar{n}\Delta n$ 。这说明带隙与折射率对比度成正比，也解释了为何高折射率对比结构的带隙更宽。

此时本征值来自特征行列式

$$\det \begin{bmatrix} \bar{\omega}_B - \omega & \kappa \\ \kappa^* & \bar{\omega}_B - \omega \end{bmatrix} = (\bar{\omega}_B - \omega)^2 - |\kappa|^2 = 0. \quad (5.4)$$

于是本征值直接分裂为

$$\omega_{\pm} = \bar{\omega}_B \pm |\kappa|. \quad (5.5)$$

这就是带隙最朴素的来源：原先在区边界相遇的两支反向波，不再直接穿过，而是因周期耦合而分裂成上下两支。更进一步，耦合强度 $|\kappa|$ 与介电函数的主导傅里叶分量成正比，因此带隙宽度满足

$$\Delta\omega_{\text{gap}} \propto |\Delta\epsilon_G|. \quad (5.6)$$

后面二维 PCSEL 的带隙、Gamma 点分裂和耦合波矩阵会包含更多通道，但基本机制相同：原本简并或近简并的平面波分量，被周期结构的傅里叶分量耦合并分裂。

5.5 Γ 点四带与 ABCD 模式分类

对 PCSEL 来说，最核心的能带特征出现在布里渊区中心 Γ 点。本节解释 Γ 点处四条带边模（A、B、C、D）的起源与物理含义 [41]。完整的 PWEM 矩阵本征问题推导、圆孔傅里叶系数的 Bessel 表达式以及二维等效模型的 slab 修正，统一放在 第十三章；本章只需要它们的两个输出： Γ 点附近的候选带边频率，以及各支带边模由哪四个平面波分量构成。

对称性速成

对称性速成：如何用点群快速分类 PCSEL 模式

初次接触 A_1 、 B_1 、 E 这些符号时，容易觉得抽象。其实它们的物理含义非常直观，而且可以直接用来预测模式的辐射特性和简并度。下面用最直白的语言解释如何快速上手。

第一步：认清你的晶格属于哪个点群 对方波晶格圆孔或方孔，若晶胞完美正方形且无倾斜变形，则属于 C_{4v} 点群（也叫 4mm 群）。这个群有 8 个对称操作：

- 恒等操作 E （什么都不做）
- 绕中心旋转 90° (C_4)、 180° (C_2)、 270° (C_4^3)
- 对两个对角线镜面反射 (σ_v)
- 对两条中线镜面反射 (σ_d)

关键是：如果一个模式在经过这些操作后保持某种一致的行为，就可以给它贴上对应的不可约表示（irrep）标签。

第二步：看懂特征标表的三行核心信息 C_{4v} 的特征标表有三类重要的不可约表示：

C_{4v}	E	$2C_4$	C_2	$2\sigma_v$	$2\sigma_d$	典型基函数
A_1	1	1	1	1	1	$z, x^2 + y^2$
B_1	1	-1	1	1	-1	$x^2 - y^2$
E	2	0	-2	0	0	$(x, y), (xz, yz)$

表中数字的含义是：当你对该模式施加某个对称操作后，它的场分布会乘以这个数字（特征标）。例如：

- A_1 ：所有操作都是 +1，说明这个模式对所有对称操作都完全对称。它在数学上类似于 s 轨道，远场通常是圆形主瓣。
- B_1 ：对 90° 旋转变号 (-1)，但对 180° 又变回 +1。这意味着它具有四叶草形状的近场，相邻象限反相。它在数学上类似于 $d_{x^2-y^2}$ 轨道。
- E ：这是二维表示（特征是 2 维矩阵而非单个数字），对应一对简并的模式。它们在数学上类似于 (p_x, p_y) 轨道，天然支持定向辐射。

第三步：用对称性预测辐射特性 一旦你知道某个模式属于哪个 irrep，就能立即回答几个关键问题：

1. **是否简并？** 一维表示 (A_1 、 B_1) 意味着非简并单模；二维表示 (E) 意味着必然存在一对简并模。这对设计单模激光器至关重要——如果你想避免模式竞争，就要选一维表示的模式。
2. **能否法向辐射？** 根据互易定理，只有那些与自由空间平面波具有相同对称性的模式才能有效耦合到法向。 A_1 模式因为全对称，原则上可以辐射，但在某些特定结构（如纯二阶光栅）中可能因相位抵消而被抑制。 B_1 模式由于四象限反相，exact Γ 点的法向辐射会被对称性严格禁止（这就是 BIC 的来源）。 E 模式天然亮模。
3. **偏振特性如何？** A_1 和 B_1 都是一维不可约表示（非简并）；其中 A_1 是全对称表示， B_1 具有 $x^2 - y^2$ 型各向异性。 E 模的两个简并分量可以组合成任意方向的线偏振，满足相位条件时也可形成圆偏振。

第四步：理解对称性破缺的后果 实际器件往往不会完美保持 C_{4v} 对称性。常见的对称性降低场景包括：

- **椭圆孔：**从 C_{4v} 降到 C_{2v} ，此时 E 双重态会分裂成两个非简并的单重态（通常标记为 A_1 和 B_1 在新群下），这就解释了为什么椭圆化可以实现单模。
- **矩形晶格：**同样破坏四重旋转对称性，解除 E 模简并。
- **倾斜入射或非零 $k_{\parallel}$ ：**离开 Γ 点后，对称性进一步降低，原本禁戒的辐射通道可能被打开。

记住一条黄金法则：对称性越高，简并越强，选择定则越严；对称性越低，简并解除，原来禁戒的过程可能被允许。

实战技巧 在实际读论文或分析仿真结果时，可以用以下方法快速识别模式的 irrep:

- 看近场：数一下有几个“花瓣”，相邻花瓣是否反相。0 节点圆对称 $=A_1$ ；4 叶片且邻瓣反相 $=B_1$ ；双叶片且有方向性 $=E$ 。
- 看远场：圆形主瓣可能是 A_1 ；十字形或四瓣形可能是 B_1 ；双瓣定向辐射很可能是 E 。
- 看色散：如果在 Γ 点有两支模频率完全重合，那几乎肯定是 E 双重态。
- 做微扰：故意加一点椭圆度，如果某支模立刻分裂成两支，原来的那个就是 E 模。

最后需要说明：对称性是分类工具，不是器件约束的全部。实际器件总会包含对称性破缺因素；掌握理想分类，有助于判断偏离的来源以及如何利用这种偏离。

5.5.1 Γ 点四带的起源与 ABCD 分类

在正方晶格二阶 Bragg 条件附近， Γ 点处存在四支带边模，按频率从低到高通常标记为 A、B、C、D [41]。这四支模来自四个主导波分量 $(\pm\beta_{\text{slab}}, 0)$ 和 $(0, \pm\beta_{\text{slab}})$ 之间的相干耦合重组。在 C_{4v} 对称群框架下，它们对应不同的不可约表示，辐射特性也截然不同：

- **模式 A、B**（频率较低的两支非简并模）：更准确地说，A 模对应 B_1 （具有 $x^2 - y^2$ 类变换特征），B 模对应 A_1 （全对称）；它们都不应被简化成“关于孔中心反对称”这一句。对模式判据而言，关键是 irrep 标签（ B_1/A_1 ）而不是口语化形容词。二者在 exact Γ 附近与法向自由空间平面波的直接耦合通常受对称性抑制，辐射损耗 α_{rad} 相对较小，对应 1D-DFB 激光器中的非辐射模类比；这两支模更容易先达到阈值，是 PCSEL 最常利用的候选激射模。
- **模式 C、D**（频率较高的简并对）：属于 E 表示，对称性允许直接法向辐射，辐射损耗较大，对应 1D-DFB 中的对称（辐射）模类比。

这里还必须再加一层命名边界。A/B/C/D 首先是带图里的分支标号，它的作用是让读者在某一篇具体文献里快速认出“哪一支是哪一支”；而 $A_1/B_1/E$ 等则是 little group 下的对称性标号。两者相关，但不是同一套语言。更稳妥的表述应是：某篇文章中的某一支 A、B、C 或 D 带边家族，因为其 irrep、辐射损耗和增益重叠更有利而成为优选模；而不能把“叫作 A 的那支”误解成跨文献永远最优。

若采用本书第 4 章的最小四波基底与标准频率排序，则这四支分支和 C_{4v} 不可约表示之间可作如下对应：低频两支 singlet 分别对应 B_1 与 A_1 ，高频两支构成同一个 E doublet，因此常写作

$$A \leftrightarrow B_1, \quad B \leftrightarrow A_1, \quad C/D \leftrightarrow E.$$

这一定义的重点不在于“哪一支永远叫 A”，而在于给出一条可复用的接口：后文若要判断 BIC、偏振分裂或法向辐射允许性，应用作选择定则判据的是 $A_1/B_1/E$ 这些 irrep 标签，

表 5.2: 本书采用的 ABCD 模标注与 C_{4v} 不可约表示的标准对应关系。后续所有章节在引用模式标注时, 应优先使用 irrep 标签 (A_1, B_1, E); A/B/C/D 仅作为带边序号使用, 且在不同文献中可能因频率排序而互换。

模标注	C_{4v} irrep	频率相对高低	法向辐射特征
A	B_1	最低	symmetry-protected dark (BIC 候选)
B	A_1	次低	symmetry-protected dark (BIC 候选)
C	E	较高 (简并对)	亮模, direct normal radiation
D			

而不是 A/B/C/D 这种文献内排序标签。若某篇文献因孔形、频率排序或记号习惯交换了 A 与 B, 只要 singlet/doublet 与 irrep 的关系保持一致, 后续的对称性判断并不会改变。

需要强调的是, 这里的辐射损耗大小排序仍然是二维无限周期的结论。进入真实三维有限尺寸器件后, 上下辐射不对称、有限孔阵边缘散射和注入分布都会进一步改变模式排序, 因此不能把“A 模辐射最低”直接等同于“A 模阈值电流最低”。图 5.2 对比了 Γ 点模式与非 Γ 点模式的近场分布和远场辐射特性。可以看出, Γ 点模式下所有原胞同相位振动, 远场相干叠加形成法向主瓣, 这也是 PCSEL 选择在 Γ 点附近工作的根本原因。

5.5.2 能带计算的收敛性测试：一个可复现的最小例子

初次使用 PWEM 时, 常见误区是默认“图能画出来就可信”。这在 PCSEL 设计里风险很高。更稳妥的做法是先固定几何参数, 只扫描平面波截断阶数 N , 观察关键观测量是否进入平台区。这里给出一个和本书脚本一致的最小例子: 正方晶格圆孔, $r/a = 0.30$, $\varepsilon_b = 3.4^2$, 比较 X 点第 2/3 带间隔

$$\Delta\Omega_{23}(X) = \Omega_3(X) - \Omega_2(X)$$

随 $N_{PW} = (2N + 1)^2$ 的变化。

图 5.3 的数据来自本书附带脚本 (`code/pwem_square_hole/run_pwem_square_hole.py`) 的逐级截断扫描。结果很直观: 当 $N_{PW} = 9$ 时误差仍很大; 提高到 $N_{PW} = 49$ 后误差已降到约 1%; 到 $N_{PW} = 121$ 时已进入 0.3% 量级。这一趋势与经典 PWEM 文献中“先做截断收敛再给结论”的基本规范一致 [29]。

常见错误：把“候选带排序”建立在未收敛基底上

如果在低截断下就直接比较“哪一条带更平、哪一条带更适合做目标模”, 最常见后果是: 随着 N_{PW} 提高, 带间距和排序发生可见漂移, 前面的设计判断被整体推翻。对 PCSEL 来说, 这类错误会沿后续 CWT/RCWA/FDTD 链条被放大, 代价远高于多做几轮收敛扫描。

5.5.3 为什么对称圆孔常给出中心发暗的远场

即使某一模式确实位于二阶 Γ 点附近, 也并不意味着它的远场中心一定最亮。对理想方格晶格、圆形孔和完美对称晶胞而言, 某些 Γ 点带边模式在法向中心会出现严格或近似的相

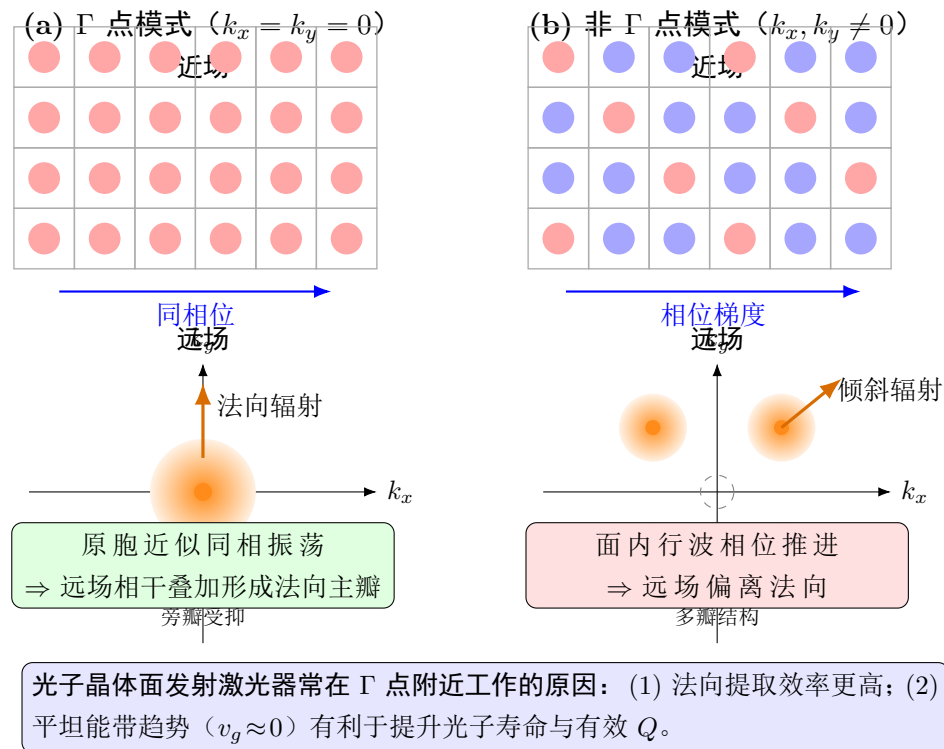


图 5.2: Γ 点模式与非 Γ 点模式的近场分布和远场辐射对比。(a) Γ 点模式 ($k_x = k_y = 0$): 所有原胞以相同相位振动, 远场相干叠加形成法向主瓣 ($\theta = 0^\circ$), 适合表面发射。(b) 非 Γ 点模式: 近场呈现行波特征, 相位沿传播方向渐变, 导致远场辐射方向图倾斜。图中 $\theta \approx 45^\circ$ 仅作 [教学示意] 角度, 真实角度依赖结构参数与工作点。底部总结了 PCSEL 选择在 Γ 点附近工作的两个核心原因: 法向辐射效率高, 以及平坦能带导致群速度趋零、光子寿命延长、 Q 值提升。

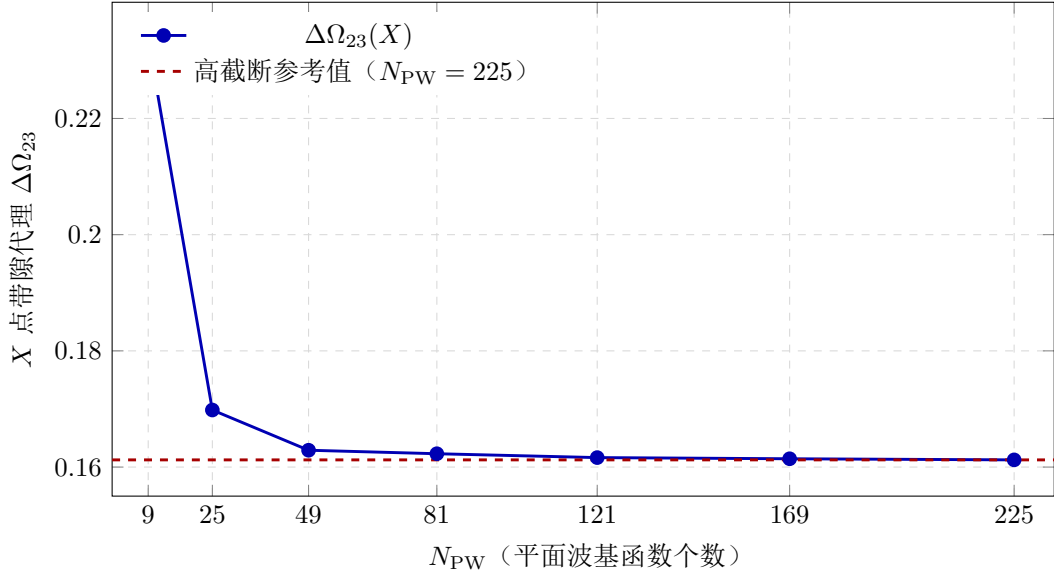


图 5.3: PWEM 截断收敛示例（本书附带脚本计算）。横轴为截断后平面波个数 $N_{PW} = (2N + 1)^2$ ，纵轴为正方晶格圆孔教学模型在 X 点的带隙代理 $\Delta\Omega_{23}$ （第 3 带与第 2 带频率差）。当 N_{PW} 从 25 提升到 121 后，曲线已进入平台区，说明“先做收敛检查再读带隙趋势”是必要步骤。

消干涉，因此在有限尺寸器件里常表现为环形或中心发暗的远场图案。

这并非说该模式“没有垂直辐射”，而是说：在 exact Γ 的 $k_{\parallel} = 0$ 位置，辐射振幅可能被对称性和相位关系压成零；有限器件中由于阵列尺寸有限，角谱不可避免地扩展到 Γ 点附近的非零 $k_{\parallel}$ 区域，于是这些邻近动量分量仍能辐射，最终形成中心发暗但外围有能量的远场图案。

这条结论在本章提前说清楚有两个好处。第一，它让读者尽早建立“存在带边谐振不等于法向中心一定亮”的边界感。第二，它把后续通过椭圆孔、双晶格、多孔单胞或 quasi-BIC 扰动把环形远场改成单主瓣远场这件事，重新还原成“同一带边家族的辐射相位关系被重新组织”，而不是凭空出现的新主题。

以方格晶格圆孔（填充比 $f = 12\%$ ， $n_b = 3.5$ ）为例，[本书算例] 中修正等效介电常数后的二维 PWEM 带隙位置与三维 FDTD 差异约为 1-3%；A/B/C/D 四模的相对频率排序在该算例下保持一致。该数值不应外推为普适误差上限。辐射常数（radiation constant） r 的绝对值需要完整的三维 CWT（见第十章、第十一章）才能精确给出，二维 PWEM 本身不直接提供此信息。

PWEM 矩阵本征方程（第十三章的式 (13.12)）是厄米本征问题，本征值为实数，这对应无损、封闭、无限周期情形。真实 PCSEL slab 的开放辐射使本征频率变为复数；在本书 $e^{-i\omega t}$ 约定下，空间辐射常数应写成 $r = -2\text{Im}(\tilde{\omega})/v_g$ 。该量只能通过包含辐射通道的开放系统模型（3D CWT、RCWA 或 FDTD）给出。PWEM 的贡献是给出频率排序与场型分类，不是给出损耗排序。

5.6 一个最小例子：正方晶格圆孔二维光子晶体

为了让概念尽快落地，先考虑一个教学上最常用的理想二维模型：背景介质折射率为 n_b ，其中刻蚀出半径为 r 的空气圆孔，圆孔按正方晶格排列，晶格常数为 a 。这是一个比真实 PCSEL slab 更理想化的模型，但它的教学价值很高，因为它已经包含了周期性、倒格空间、高对称点和带图读取的大多数关键元素。

对正方晶格，实空间基矢可取为

$$\mathbf{a}_1 = a \hat{x}, \quad \mathbf{a}_2 = a \hat{y}, \quad (5.7)$$

对应的倒格矢基底为

$$\mathbf{b}_1 = \frac{2\pi}{a} \hat{x}, \quad \mathbf{b}_2 = \frac{2\pi}{a} \hat{y}. \quad (5.8)$$

在这种约定下，第一布里渊区中的三个高对称点通常记为

$$\Gamma = (0, 0), \quad X = \left(\frac{\pi}{a}, 0\right), \quad M = \left(\frac{\pi}{a}, \frac{\pi}{a}\right). \quad (5.9)$$

若只想快速建立直觉，最常画的高对称路径就是

$$\Gamma \rightarrow X \rightarrow M \rightarrow \Gamma. \quad (5.10)$$

这条路径并非说系统只在这些方向上有物理，而是因为它能以最紧凑的方式展示主要对称性变化和带的代表性重组过程。

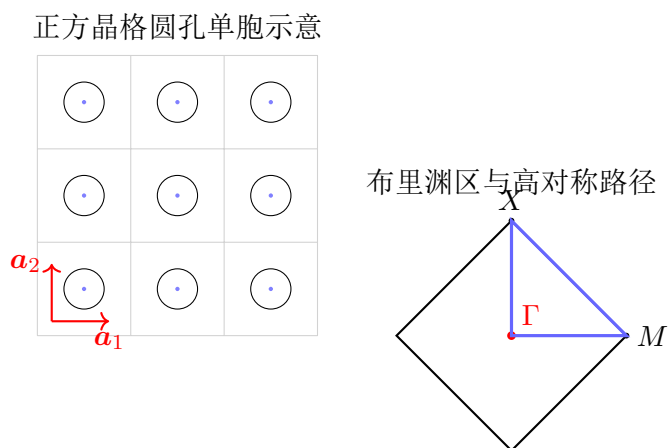


图 5.4: 正方晶格圆孔二维光子晶体的教学型几何示意，以及对应的高对称路径 $\Gamma \rightarrow X \rightarrow M \rightarrow \Gamma$ 。

5.7 怎么看一张能带图

阅读能带图时，不应只记录曲线数量，而应判断每个图形特征对应的物理信息。下面列出五个基本问题。

第一问：横轴和纵轴到底是什么

横轴通常是沿高对称路径的倒空间路径坐标，而非“实空间距离”；纵轴通常是像式 (5.2) 那样的归一化频率，而非绝对频率。因此，同一张图可以同时服务于“几何缩放设计”和“模式对称性阅读”。如果这一步没弄清，就很容易把不同尺寸器件的结果错拿来直接比较。

第二问：是否存在完整带隙

若在某一个频率区间内，沿整条高对称路径都没有本征带穿过，那么这个区间可能对应某一极化下的带隙。注意这里的关键词是“沿整条相关路径都没有本征值”。只在某一段路径上没看到曲线，并不等价于全局带隙；它可能只是局域频带稀疏或路径选择导致的视觉空白。

第三问：哪些点出现简并或近简并

高对称点附近的简并尤其重要，因为它们往往意味着模式可被微扰重新组织。对 PCSEL 来说，简并本身不是目标，如何利用对称性破缺让目标模和竞争模在损耗、偏振或远场上分开，才是关键。也正因此，进行器件设计时，读带图不应只记频率，还应记“哪几条带在 Γ 点附近缠在一起”。

第四问：哪些带在目标频率附近变平

带变平意味着群速度减小。更正式地，一维路径上的群速度定义为

$$v_g = \frac{\partial \omega}{\partial k}, \quad \text{更一般地} \quad \mathbf{v}_g = \nabla_{\mathbf{k}} \omega(\mathbf{k}). \quad (5.11)$$

因此，带图上某一段曲线越平，表示该方向上的群速度投影越小，局域态密度往往越高。对很多光子晶体问题，这会让人立刻联想到带边效应、场增强和长驻留时间。但这里必须谨慎：平带不等于最终低阈值。它只说明色散特征上可能存在有趣的候选模。若该模同时与自由空间过强耦合、与有源区重叠差、或在器件工作中受热漂移严重，那么它仍然可能不是最终激射模。

第五问：这张图是否已经告诉了我辐射信息

对理想二维无限周期问题，能带图主要告诉你 Bloch 模存在性和色散。若要判断该模在真实 slab PCSEL 中是否容易向法向辐射，还要进一步考虑光锥、纵向模、上下非对称层栈和开放边界。也就是说，能带图本身还不同于辐射图、远场图或阈值图。

5.8 从能带图中我们究竟能得到什么

若把能带图的用途说得更系统一些，它至少提供五类高价值信息。

第一，**候选模频率窗口**。这决定目标工作波段附近是否存在值得继续分析的 Γ 点模或带边模。

第二，**高对称点简并结构**。这决定后续微扰、对称性破缺和偏振工程可能怎样起作用。

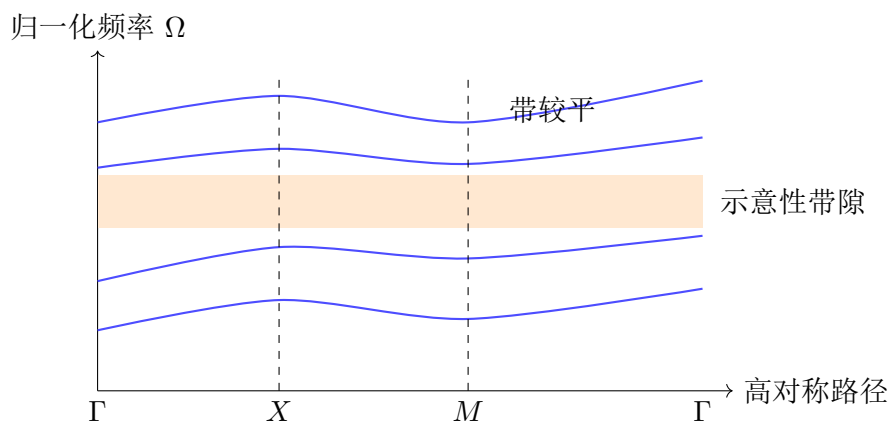


图 5.5: 能带图读图示意。橙色区域表示某一极化下可能的带隙，曲线局部变平提示群速度减小，但这些都只是候选模筛选信息，不直接等于激射结论。

第三，色散平坦程度与群速度趋势。这能帮助判断哪些模式可能具有较长驻留时间或较强局域化倾向。

第四，带隙与频带重排。这能帮助理解周期结构如何阻止或允许某些传播通道。

第五，给后续数值与半解析模型提供候选对象。第十章和十三至十五都需要先知道“应该盯住哪几条带”。

但同样重要的是知道它不能直接给什么。能带图不能直接给出：

- 有限孔阵边界下的真实腔模排序；
- 模与光锥耦合后的定量辐射损耗；
- 与量子阱重叠和内部吸收共同决定的净模增益；
- 电注入、热漂移和模式竞争下的阈值电流与工作窗口。

这条边界需要在本章明确。否则，后面关于 RCWA、FDTD/FEM、外延层、电学和热学的章节容易被误读为附加修正，而不是完成器件级结论所必需的部分。

可概括为：**band diagram 给出候选资格，不给出器件级排序**。器件级排序还需要加入辐射、有限尺寸、增益、注入和热反馈。

图5.6 用一张守门图总结这条边界。读能带图时应先把“可以直接读出的量”和“必须继续计算的量”分开放置；只要还没有进入右侧那些后续证据，就不能把候选模目录升级成器件级结论。

5.8.1 带边模式、缺陷模式与 PCSEL 候选模的区别

在光子晶体激光器中，最常见的两类谐振思路是缺陷模式与带边模式。缺陷模式对应在本周期完好的光子晶体中人为引入点缺陷或线缺陷，使某一频率位于禁带中的局域态被困在缺陷附近；它的关键词是“局域化”和“缺陷腔”。带边模式则不同，它并不依赖单独的几何缺陷，而是依赖色散关系在某个高对称点或布里渊区边界处出现极值，使群速度趋于零、态

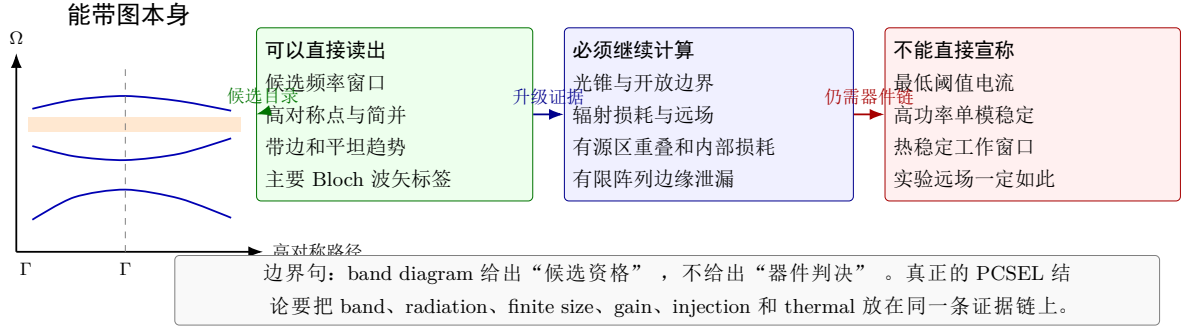


图 5.6: 能带图的证据边界。能带图可以直接给出候选频率、高对称点、简并与平坦趋势；但辐射损耗、远场、有限尺寸、净模增益、注入和热反馈都需要后续模型补齐。

密度增大，并在有限尺寸结构中形成长寿命共振。它的关键词是“色散极值”和“二维分布反馈”，而非“缺陷”。

这一差别对 PCSEL 特别重要。PCSEL 追求的是大面积二维相干振荡与法向面发射，因此更自然的工作方式是让一大片周期结构在带边附近共同参与反馈，而非把光强行困在一个很小的点缺陷里。也正因此，带边模式在 PCSEL 中不只是一个能带术语，而是连接二维分布反馈、法向辐射和大面积单模输出的核心概念。

但这里必须把层级写清楚：

1. 带边模式首先是一个**被动色散概念**，它说明模式位于色散极值附近；
2. PCSEL 候选模是在此基础上进一步要求该模式能够与法向辐射通道、纵向波导模和有源区耦合；
3. 真实激射模还要再经过有限尺寸、净模增益、注入分布和热回写的筛选。

因此，带边模式是重要起点，但绝不是器件结论本身。

5.8.2 有限尺寸下带边模式的品质因子为什么会随群速度降低而增大

带边模式之所以重要，最常见的物理直觉是：在色散极值附近，群速度

$$\mathbf{v}_g = \nabla_{\mathbf{k}} \omega(\mathbf{k}) \quad (5.12)$$

趋近于零，光在结构中的驻留时间变长，因此更容易形成高品质因子共振。这个说法方向上是对的，但还需要加上有限尺寸这一层。

对于有限尺寸的周期结构，Bloch 波矢并不能像无限周期那样被无限精确地确定。若器件的有效横向尺度为 L ，则在倒空间中总会存在一个有限的波矢不确定度

$$\Delta k \sim \frac{\pi}{L}. \quad (5.13)$$

若某一候选模围绕带边波矢 $\mathbf{k}_0$ 展开，其色散可写成

$$\omega(\mathbf{k}_0 + \mathbf{q}) = \omega_0 + \mathbf{v}_g \cdot \mathbf{q} + \frac{1}{2} \mathbf{q}^T H \mathbf{q} + \dots, \quad (5.14)$$

其中 H 是局部曲率张量。于是，有限尺寸导致的谱宽至少会受到两层控制：

1. 若工作点只是接近带边而未精确落在极值点，则有一阶展宽

$$\Delta\omega_{\text{lin}} \sim |v_g| \Delta k; \quad (5.15)$$

2. 若工作点已在真正极值点，使一阶项消失，则剩余展宽由二阶曲率主导

$$\Delta\omega_{\text{quad}} \sim \|H\|(\Delta k)^2. \quad (5.16)$$

因此，“带越平， Q 往往越高”的更准确表述应是：更小的群速度和更小的局部曲率，都会压缩由 Δk 引起的频率展宽，从而提高由结构色散决定的有效品质因子。

但边界同样必须说清楚：带平并不自动等于最高 Q ，更不自动等于最低阈值。真实 Q 还会同时受到垂直辐射、边缘泄露、工艺粗糙和吸收损耗控制。也就是说，带边平坦度是影响 Q 的重要因子，但绝不是唯一因子。

5.8.3 光锥、引导共振与为什么不是所有带边都适合面发射

对理想二维光子晶体来说，能带图只是在回答“给定 Bloch 波矢下允许哪些本征频率”；但 PCSEL 的真实结构是有限厚度的光子晶体平板或多层半导体波导，因此带边模式还必须再经过一关：它是否能够耦合到垂直辐射连续谱。

设外部介质折射率为 n_{out} ，则某个 Fourier 分量 $(\mathbf{k}_{\parallel} + \mathbf{G})$ 若满足

$$|\mathbf{k}_{\parallel} + \mathbf{G}| \leq k_0 n_{\text{out}}, \quad (5.17)$$

便存在沿法向传播的自由空间解。所有满足这类条件的动量点构成光锥。于是，对 slab-PCSEL 而言：

1. 位于光锥之下的模式更接近导模，垂直泄露弱；
2. 位于光锥之上、但因周期折叠而进入简约布里渊区的模式，则可能成为引导共振 (guided resonance)，既保留导模特征，又能向外辐射；
3. PCSEL 所需的工作带边，是既能提供二维反馈、又能以合适方式打开法向辐射通道的那一类带边，而非任意带边。

这也是为什么本章不能只说“带边群速度低，所以适合做激光器”，而必须再补一句“还要检查它相对于光锥和辐射通道的位置”。否则读者会误以为只要找到最平的一段带就足够了。

5.8.4 用三角晶格 I–IV 的布拉格图像理解为什么只有某些带边能成为面发射候选模

为说明“不是所有带边都适合 PCSEL”，可再考察一个经典的三角晶格教学图像。它的价值不在于规定未来一定要采用三角晶格，而在于说明：不同带边点对应的耦合回路和辐射方向本来就不同。

以三角晶格 TE 模为例，在文献中常把若干典型带边点记为 I、II、III、IV。它们的差别可概括为：

1. 在点 I，主要发生的是 180° 反向耦合；由于存在六个等价方向，腔内可对应多个方向不同的谐振家族；
2. 在点 II，相互夹角为 120° 的波之间可以实现耦合，因此耦合图像已不同于单纯的反向反馈；
3. 在点 III，六个等价方向上的波可以同时耦合，并且还能通过一阶 Bragg 衍射把腔内谐振场耦合到法向辐射通道，因此它最适合拿来解释 PCSEL 的面发射原理；
4. 在更高频的点 IV，除了垂直方向外还常伴随倾斜的面外衍射，因此虽然它也能辐射，但并不天然是最优的法向面发射选择。

这段内容的教学目的非常明确：PCSEL 设计不是“先找一个带边，再默认它能法向出光”，而是必须检查这个带边对应的 Bragg 散射网络究竟把光送回了哪里、又把光送出了哪里。

5.9 为什么 Γ 点会成为 PCSEL 的重点

现在可以把注意力集中到 PCSEL 最核心的倒空间位置： Γ 点。它之所以重要，是因为它对 PCSEL 的三个关键目标同时敏感，与其在图上的居中位置无关。

第一，PCSEL 追求近法向面发射。对倒空间而言，法向出射对应零面内波矢，因此 Γ 点天然与近法向辐射联系最紧密。

第二， Γ 点往往汇聚多个等价基本波分量。以正方晶格为例，最小耦合图像常涉及沿 $\pm x$ 和 $\pm y$ 方向传播的四个基本波。它们在 Γ 点附近发生强耦合，这也是 PCSEL 耦合波理论的起点。

第三， Γ 点通常是对称性标签最清晰的地方。模式的奇偶性、旋转对称性和简并结构在这里最容易被分类，而后续偏振选择和远场工程恰恰依赖这些标签。

因此，PCSEL 之所以一再回到 Γ 点，是因为“法向辐射、基本波耦合、对称性工程”这三件事都在这里交汇，而非出于习惯。新手如果抓住这三个目标的交汇关系，后面再看面内反馈、面发射机制和偏振整形就不会断裂。

5.10 带边模、 Γ 点模、被动共振模和真实激射模的区别

这一节必须写得非常硬，因为它关系到整本书能否避免概念偷换。

带边模 (band-edge mode)

带边模强调的是色散性质，即模式位于色散曲线的极值附近，群速度趋近于零。它是一个被动色散概念。

Γ 点模 (Gamma-point mode)

Γ 点模强调的是倒空间位置，即模式位于布里渊区中心附近。它是一个波矢位置概念。

被动共振模 (passive resonance)

被动共振模强调的是在无增益条件下、带有开放边界和损耗时的共振特征。它是一个开放系统光学概念，已经比理想能带更接近真实器件，但仍然没有引入增益和输运。

真实激射模 (lasing mode)

真实激射模强调的是在给定注入和热反馈下，哪个模先达到阈值并在工作窗口内维持主导输出。它是一个器件工作状态概念。

这四者可能重合，但绝不能默认重合。某个模式可以同时是 Γ 点模和带边模，但仍不是最终激射模；某个被动共振模可以拥有较高 Q ，但若与有源区重叠差或在热回写后排序翻转，它仍可能输给别的模。

“某模式在 Γ 点”“某模式位于带边”“某模式被动 Q 较高”都只是候选资格，不是最终器件判决。器件级判断必须等到净模增益、有限尺寸、注入分布和热回写都进入模型之后才成立。

5.11 为什么能带图好看不等于器件就会好用

这句话必须反复说，因为它对应的误判太常见。假设某条带在 Γ 点附近很平、频率也落在目标波段上，看起来非常“像一个好模式”。这当然值得兴奋，但还必须立刻追问至少四个问题。

第一，它在真实 slab 结构中是否落在有意义的光锥与导模区间里。

第二，它与量子阱和损耗层的重叠是否有利。

第三，它在有限孔阵中是否仍维持正确排序，而不会被边缘泄漏或横向包络效应推翻。

第四，它在电流注入和温升写回之后是否仍保持净模增益优势。

只要其中任一项答案不清楚，就不能把“能带上很好看”上升为“器件设计已成功”。这也是为什么本章后面必须接第六章、十三至十五和十七至二十。

若某个 Γ 点候选模在能带图上位置很理想，但 RCWA 显示它主要耦合到不希望的辐射通道，或者有限阵列 FDTD 显示它对边缘极其敏感，那么这个模式仍然只适合作为“理论上有趣的候选模”，而不是器件设计中的首选目标。

PCSEL 的能带设计是在寻找“既位于合适带边、又能通过正确的 Bragg 回路形成二维反馈、并且能以可控方式接入法向辐射通道”的那一类候选模，而非在寻找“最平的一条带”。

5.12 从本章到 PWEM：为什么还需要平面波展开法

到这里，读者应该已经明白“为什么要看能带图”以及“能带图能告诉我们什么”。但还没有回答另一个很实际的问题：这些图到底是怎么来的？最常用的数值入口之一，就是平面波展开法，也就是 PWEM。第十三章 将把本章的概念完全落实到数值过程：从周期 Maxwell 本

征问题出发，如何展开到平面波基底，如何构造正方晶格圆孔二维光子晶体的矩阵本征问题，如何沿 $\Gamma \rightarrow X \rightarrow M \rightarrow \Gamma$ 计算带图，以及如何把计算结果读成候选模筛选信息。

若你希望把本章的正方晶格圆孔例子立刻落到真实可运行的计算，而不是停留在概念层面，那么可以直接跳转到 [第十三章“平面波展开法 \(PWEM\) 与光子能带计算”](#)。在那里，本书项目附带的 Python 脚本会实际计算一张教学型带图，并把结果直接插回书稿。

但在进入数值之前，本书仍然按物理逻辑先走第六章至第十二章。原因很简单：在 PCSEL 里，模式为什么面发射、为什么会竞争、为什么阈值不是单看带图就能决定，这些物理问题必须先讲透。否则第十三章里的 PWEM 很容易再次被误读成“算图软件教程”。

把能带图当成“只要会算就会设计”的工具，是软件化误读。教材顺序应先说明图像与边界，再进入算法与求解；否则容易只记住出图流程，而不知道曲线对应的物理量。

回看本章

光子晶体能带是周期 Maxwell 本征问题的参数化本征谱。它与半导体材料电子能带共享 Bloch 定理、倒空间和 Bragg 分裂这些数学骨架，但本征对象和器件含义不同：电子能带描述载流子可占据能量和跃迁条件，光子能带描述电磁模式允许频率、传播、反馈和辐射。对 PCSEL 而言，能带图的价值在于给出候选模目录：它告诉我们目标波段附近有哪些 Γ 点模、带边模和简并结构值得继续追踪。进一步的亮模/暗模 (bright/dark) 区分来自模式对称性与辐射通道的匹配：若近场对称性允许与自由空间平面波耦合，则为亮模；若所有辐射通道因对称性相消，则为暗模 (BIC)。定量判据将在第八章用群论语言给出。但它不能直接替代开放结构分析、阈值分析和器件级热电建模。 Γ 点之所以成为重点，是因为法向辐射、基本波耦合和对称性工程都在这里交汇；后面第九章会进一步把这条线翻译成 BIC / quasi-BIC 的辐射语言，而第七章、第二十三章则会把同一批候选模继续写成阈值排序、动态稳定性和线宽来源的问题。沿着这条线阅读，后面的 CWT、PWEM、RCWA 和器件模型才能保持连续。

练习题

1. [基础] 用你自己的话解释“光子能带是参数化本征谱”这句话，并说明这里的参数是什么、输出是什么。
2. [基础] 分别用一句话说明“半导体材料电子能带”和“光子晶体光子能带”各自回答什么问题。
3. [基础] 解释为什么高对称路径上的视觉空白并不自动等于真正的带隙。
4. [进阶] 说明带边模、 Γ 点模和真实激射模之间为什么不能直接画等号。
5. [综合] (思路提示：先定层级，再列关键式) 假设某个候选模在 Γ 点附近很平且频率理想，写出至少三个仍然必须继续验证的后续问题。

延伸阅读

- 下一章将从能带目录继续向前推进，讨论面内基本波如何耦合、垂直辐射如何产生，也就是为什么这些候选模会成为面发射候选模。
- 若你已经意识到“ Γ 点候选模”的命运并不只由带图决定，那么可以带着本章的候选模清单继续阅读 第九章、第二十三章：前者回答它如何通过辐射通道进入 BIC / quasi-BIC 语言，后者回答它跨过阈值之后如何进入调制、稳定性和噪声问题。
- 第十三章 将把本章的几何和能带概念变成实际可计算的 PWEM 流程，特别是正方晶格圆孔二维光子晶体这一教学型例子，并给出本书项目中附带的 Python 计算脚本与实际带图。

第二部分

对称性、BIC 与降阶理论

第六章 面发射机制与面内耦合

读完本章，你能

- 从面内导模出发理解 PCSEL 的反馈为何来自二维布拉格散射网络；
- 解释同一组周期散射为什么既能提供反馈，又能打开垂直辐射通道；
- 建立“晶格傅里叶分量 $\rightarrow$ 耦合系数 $\rightarrow$ 阈值、偏振与远场”的连续图像；
- 明确本章如何把第 5 章的 Γ 点候选模推进到后续的阈值、BIC 与 CWT 模块。

读前准备 第 4 章、第 5 章中的 Bloch 模、 Γ 点、模式分裂与候选激射模概念。

阅读主线 本章所处模型层级：L1 $\rightarrow$ L2

上一章已经回答了“哪些模式值得关注”。本章进一步回答“这些模式靠什么在面内形成反馈，又为什么能够向法向辐射”。我们先从普通 slab 导模出发，再引入二维周期散射，随后建立最小四波图像，最后讨论反馈与提取之间的张力，以及它如何成为后续阈值、偏振、BIC 与 CWT 的共同基础。

6.1 先从最朴素的问题开始：如果没有光子晶体，会发生什么

想象一个没有二维周期结构的普通有源 slab 波导。若波导中存在横向传播的导模，它们会沿面内传播，并由端面反射、杂散散射或额外设置的反馈结构来形成腔。此时，模式的主要命运是“继续在平面内走”。

PCSEL 的关键变化恰恰在于：我们把一层二维周期折射率调制嵌入了这个面内传播环境。结果是传播图像被根本改写，而非简单地“多了一点散射”。原本孤立的平面内波分量，开始通过倒格矢连接成网络；某些原本只会在面内传输的导模，也因此获得了向自由空间辐射的机会。

因此，理解面发射机制的第一步，是先把面内传播这件事想清楚，而非直接去看远场。PCSEL 的面发射并非孤立产生的，它来自面内导模在二维周期介质中的重新组织。

直觉框

PCSEL 的“反馈”与“出光”是同一个散射网络的两种后果，而非先后发生的两件事。对一个给定模式而言，散射既可能把它送回别的面内波分量，也可能把它送进光锥成为

辐射。设计 PCSEL，本质上是在管理这两条路径的相对强弱。

6.2 面内反馈从哪里来：从倒格矢散射看布拉格回路

设一束面内导模具有主导波矢 $\mathbf{k}_{\parallel}$ 。若介电函数存在平面内周期调制，其傅里叶展开写为

$$\varepsilon(\rho, z) = \sum_{\mathbf{G}} \varepsilon_{\mathbf{G}}(z) e^{i\mathbf{G} \cdot \rho}.$$

那么 Maxwell 方程中的介电函数乘场项会自动引入不同平面波分量之间的耦合：

$$\mathbf{k}_{\parallel} \longleftrightarrow \mathbf{k}_{\parallel} + \mathbf{G}.$$

把它写成更接近散射论语言的形式，就是

$$\mathbf{k}_j - \mathbf{k}_i = \mathbf{G}, \quad (6.1)$$

其中 $\mathbf{k}_i$ 与 $\mathbf{k}_j$ 分别是耦合前后的面内波矢。式 (6.1) 给出的正是二维周期结构中的最小 Bragg 条件：只有当两个波分量之差恰好由某个倒格矢补偿时，它们才会被同一个晶格 Fourier 分量有效连接。

这一条式子就是 PCSEL 面内反馈的源头。它说明：周期结构中的反馈是通过倒格矢把一个传播方向耦合到另一个传播方向，再由这些方向之间的闭合耦合回路形成分布式反馈，而非端面“反弹回来”。

在方格晶格中，最直观的例子是四个主要平面波分量：

$$(+\beta_{\text{slab}}, 0), \quad (-\beta_{\text{slab}}, 0), \quad (0, +\beta_{\text{slab}}), \quad (0, -\beta_{\text{slab}}),$$

它们分别对应沿 $\pm x$ 和 $\pm y$ 传播的导模分量。若波长和晶格常数满足近布拉格条件，这四个分量会通过最主要的倒格矢彼此耦合，构成一个最小反馈网络。这个网络正是许多 PCSEL 解析模型的起点。

这里先给出一个最小记号约定：本章的传播常数统一写作 β_{slab} ，并满足

$$\beta_{\text{slab}} = n_{\text{eff}} k_0.$$

其中 n_{eff} 是把纵向层栈压缩到二维平面模型时的等效折射率。它不是“凭经验凑出来”的参数，而来自纵向模剖面与边界连续条件；完整推导见 第十七章。

把“近布拉格条件”写成式子，会更容易和后面的 CWT 对上。对一维前后向耦合，最基本的条件是

$$2\beta_{\text{slab}} \approx mG, \quad (6.2)$$

其中 $G = 2\pi/a$ 是主导倒格矢大小， m 为布拉格阶次。这里必须强调一阶与二阶布拉格的物理区别：一阶条件 $2\beta_{\text{slab}} = G$ （半波长周期）主要形成 $+\beta_{\text{slab}} \leftrightarrow -\beta_{\text{slab}}$ 的前后向耦合，折叠回 Γ 点的分量满足 $k_{\parallel} = \beta_{\text{slab}} - G = -\beta_{\text{slab}} \neq 0$ ，因此通常仍在光锥外，难以直接法向辐射；二阶条件 $2\beta_{\text{slab}} = 2G$ （全波长周期）则允许单次散射后 $k_{\parallel} = \beta_{\text{slab}} - G \approx 0$ 的分量落入光锥

中心，打开法向出射通道。也因此 PCSEL 更强调二阶布拉格或等效的 Γ 点折叠。对 PCSEL 常见的区中心工作图像，更有用的说法是：原本在扩展区方案中位于 $\pm\beta_{\text{slab}}$ 的几支导模分量，经由倒格矢散射折叠回 Γ 点附近，因此常把

$$|\mathbf{k}_{\parallel} + \mathbf{G}| \approx 0 \quad (6.3)$$

视作区中心耦合条件。教材里说“PCSEL 工作在二阶布拉格附近”，强调的是这类折叠回路，而不是孤立的整数阶次标签。

例子

[本书算例] 940 nm 器件的第一轮几何起点。取 $n_{\text{eff}} \approx 3.4$ ，区中心折叠条件 $\beta_{\text{slab}} \approx G$ 给出

$$a \approx \frac{\lambda_0}{n_{\text{eff}}} = \frac{940 \text{ nm}}{3.4} \approx 276 \text{ nm}.$$

所以 280 nm 左右是合理的初始晶格常数。它只是把纵向导模传播常数折回 Γ 点的尺度估计；孔形、刻蚀深度和开放辐射会继续移动真实候选频率。

例子

为什么 PCSEL 总是强调二阶布拉格，而不是一阶布拉格？

一阶布拉格最擅长做的是导模之间的前后向反馈，它确实能形成分布式反射，但通常还不足以把主导波分量直接送进法向出射通道。二阶布拉格或等效的 Γ 点折叠则不同：同一组倒格矢既能维持面内反馈网络，又能把部分面内动量压缩到 $k_{\parallel} \approx 0$ ，于是法向辐射被打开。对 PCSEL 来说，关键不在“整数阶次”标签本身，而在是否存在同时提供反馈和出射的散射路径。

扩展区方案与简约区方案的对应关系 一阶布拉格条件 $2\beta_{\text{slab}} = G$ 在扩展区方案中对应导模在 $\pm\beta_{\text{slab}}$ 处的前后向耦合；而在简约区方案中，这两个状态被折叠回第一布里渊区边界 ($k_x = \pi/a$)，形成带隙。二阶布拉格 $2\beta_{\text{slab}} = 2G$ 在扩展区中对应 $\beta_{\text{slab}} \approx G$ 的导模经散射后到达 $k_{\parallel} \approx 0$ ；在简约区中则表现为 Γ 点附近的模式折叠。PCSEL 的法向辐射要求简约区图像中模式位于 Γ 点 ($k_{\parallel} \approx 0$)，这对应扩展区中 $\beta_{\text{slab}} \approx G$ 的分量经一次倒格矢散射后到达光锥中心。如果把式 (6.1) 与式 (6.3) 放在一起看，会更容易理解本章后面要不断出现的两件事。第一，面内反馈来自“导模 $\rightarrow$ 导模”的闭合散射回路；第二，法向辐射来自“导模 $\rightarrow$ 光锥内连续谱”的开放散射回路。两条路的区别不在于是否经过周期结构，而在于散射后的终态仍在导模族里，还是已经掉进了辐射连续谱里。

严格表述（难度）

若只保留若干主导波分量，设其包络组成向量 $\mathbf{a}$ ，则最小线性模型可写成

$$\frac{d\mathbf{a}}{dt} = (i\mathbf{\Omega} - \mathbf{\Gamma} + \mathbf{K}) \mathbf{a},$$

其中 $\mathbf{\Omega}$ 表示未耦合时的本征频率项， $\mathbf{\Gamma}$ 表示总损耗， $\mathbf{K}$ 表示由晶格傅里叶分量引入的耦合矩阵。本章并不要求立即求解这个矩阵，而是先弄清每个矩阵元在物理上对应哪一

条散射路径。第 10 章会把这一步系统化。

6.3 为什么说 PCSEL 的反馈不是一维 DFB 的简单二维版

把 PCSEL 说成“二维 DFB”在启发式上并非完全错误，因为两者都依赖周期结构提供分布式反馈。但如果把这句话理解得太字面，就会忽略 PCSEL 的特有之处。

一维 DFB 的核心是：沿单一传播方向的前后波耦合。而 PCSEL 中，不同面内方向上的波同时参与耦合，模式不再只是“前向波和后向波”的简单组合，而可能是多个等价传播方向的相干叠加。于是：

- 模式对称性变得更丰富；
- 偏振选择不再只是副效应，而是内建在耦合网络中；
- 垂直辐射可以和面内反馈共享同一组散射傅里叶分量。

换句话说，一维 DFB 的直觉可以帮助入门，但不能替代对二维散射网络的理解。

Noda 2023 tutorial 给出的反馈网络设计路线

Noda 等人的 tutorial 对本章最有教学价值的一点，是把“二维反馈网络”从概念图像推进成可放大的设计路线 [50]。如果只看结果数字，很容易把路线误读为“孔形越来越复杂、功率越来越高”；但按本章语言重写，它实际上是一条反馈网络逐步可控化的链条。

第一步是高折射率反差方格或三角晶格：它证明 $\Gamma^{(2)}$ 点附近的多束 Bloch 波可以形成二维相干共振，而不是一维 DFB 的简单并列。第二步是非对称单胞，例如三角孔或直角三角孔：它不再只追求高 Q ，而是有意改变各带边模的辐射常数和远场干涉，使目标模获得可用的法向出射通道 [20]。第三步是双晶格：两组晶格点的相对位移重排 180° 衍射相位，使基模横向包络展开，同时让高阶横向模更容易在有限边界承担相对更大的损耗 [75]。第四步是 Hermitian/非 Hermitian 联合调控：面内 $180^\circ/90^\circ$ 衍射不再单独优化，而要和经由辐射波产生的耗散耦合一起调节，使毫米级口径下仍然保留模式裕量 [74, 50]。

这条路线说明，PCSEL 的“反馈设计”不能只问某个耦合强不强，而要问该耦合在目标模和竞争模之间如何分配、在法向输出和边缘泄漏之间如何分配、在冷腔和热工作点之间是否仍保持同样排序。这也是本书后面用模裕量和工作窗口语言重写高功率 PCSEL 的原因。

6.4 垂直辐射为什么会出现：从光锥条件理解面发射

到这里还需要回答一个核心问题：如果面内耦合形成了反馈，那为什么光又能从法向出去？

关键在于自由空间辐射通道的动量匹配。对于频率为 ω 的自由空间波，平面内波矢必须满足

$$|\mathbf{k}_{\parallel}| \leq n_{\text{out}} \frac{\omega}{c}, \quad (6.4)$$

其中 n_{out} 是目标输出侧外部介质的折射率；若上方为空气，则 $n_{\text{out}} \approx 1$ 。现实 PCSEL 不只有一个光锥，而是有上下两个光锥：向上空气侧的光锥较窄，向下衬底侧的光锥更宽。以下方 GaAs 衬底为例，常有 $n_{\text{sub}} \approx 3.5$ ，因此许多对空气侧来说仍偏“导模”的分量，对衬底侧却已经是允许辐射的波。这也是上下辐射天然不对称的第一原因。满足式 (6.4) 的区域称为光锥 (light cone)。普通 slab 导模往往位于光锥之外，因此难以直接向自由空间辐射。但周期散射可以改变面内动量：一个原本位于光锥外的导模分量，经由倒格矢散射后，可能被折回到上方或下方光锥内，于是对应的场就能向自由空间泄露。

这就是 PCSEL 面发射的核心机制：面内导模先存在，周期结构再把它的一部分动量“搬运”到可辐射区域。因而面发射是由晶格动量补偿打开的辐射通道，而非对原有导模的简单泄漏。

如果把这句话写成更像“可计算判据”的形式，就是：先有 slab 导模色散

$$\omega = \omega_{\text{guided}}(\mathbf{k}_{\parallel}),$$

它在折叠前通常位于光锥外；再经由倒格矢散射把面内动量改写为 $\mathbf{k}_{\parallel} + \mathbf{G}$ ；最后检查是否满足

$$|\mathbf{k}_{\parallel} + \mathbf{G}| \leq n_{\text{out}} \frac{\omega}{c}. \quad (6.5)$$

满足式 (6.5) 的分量就进入辐射连续谱，不满足的分量则继续留在导模家族中。于是，“面内反馈”和“法向耦出”才会成为同一个散射网络的两面。

更进一步，某个具体辐射通道的强弱并非一句“落入光锥就会辐射”就结束了。若把目标辐射通道记为 ν ，其平面内辐射波矢记为 $\mathbf{k}_{\parallel,\nu}^{\text{rad}}$ ，则导模到该通道的耦合振幅在最小投影层级上可写成

$$d_{\nu} \propto \int_V \Delta\varepsilon(\rho, z) \mathbf{E}_{\nu,\text{rad}}^*(\rho, z) \cdot \mathbf{E}_{\text{guided}}(\rho, z) dV. \quad (6.6)$$

若再把周期扰动展开到倒格空间，且只保留满足

$$\mathbf{k}_{\parallel,\nu}^{\text{rad}} - \mathbf{k}_{\parallel,\text{guided}} = \mathbf{G}_{\nu}$$

的主导 Fourier 分量，则式 (6.6) 会压缩成

$$d_{\nu} \propto \int \Delta\varepsilon_{\mathbf{G}_{\nu}}(z) \mathbf{e}_{\nu,\text{rad}}^*(z) \cdot \mathbf{e}_0(z) dz. \quad (6.7)$$

这是本章用于估计第一步散射通道的核心公式。它说明某一辐射通道的幅度由两部分共同决定：一是恰好匹配该动量转移的结构 Fourier 分量，二是导模与辐射通道在纵向上的场重叠。因此，面内反馈和法向辐射是同一套 Fourier 散射网络在不同终态上的投影，而非两套无关物理。

笔记

式 (6.7) 给出的是比例关系而非严格等式。严格的辐射振幅还需要包含纵向波矢失配因子 $1/k_{z,\nu}$ 以及来自模式归一化的前置系数。在定性分析通道选择时，这些因子通常不影响主导物理图像，但在进行定量计算时必须补全。

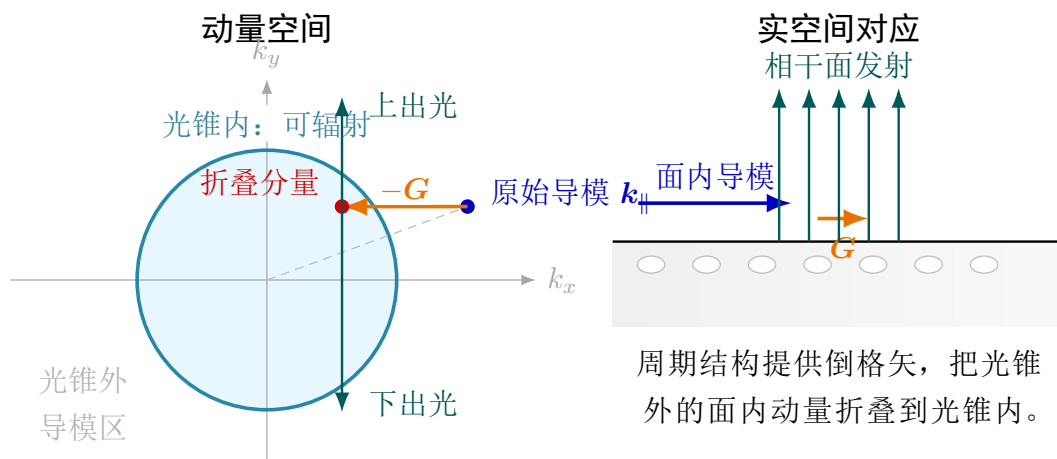


图 6.1: 动量折叠把面内导模分量送入辐射连续谱的示意图。蓝点表示原始导模分量，橙色箭头表示由倒格矢提供的动量转移，红点表示折叠后的分量；一旦落入光锥，便可向上下自由空间通道辐射。

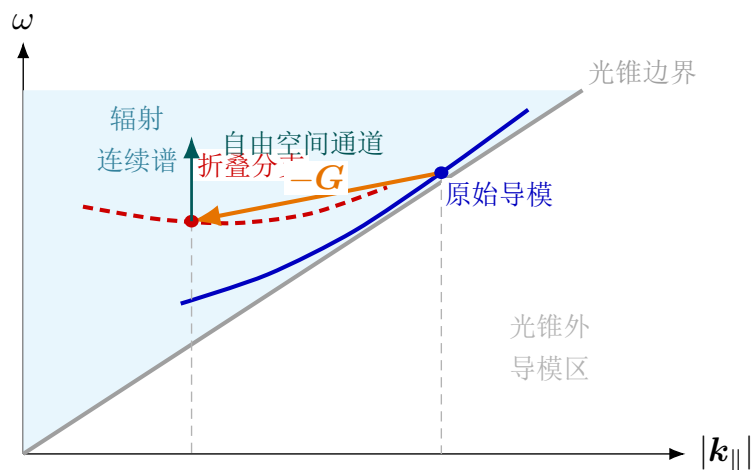


图 6.2: 平板导模色散与光锥边界示意。周期散射并不创造新色散，而是把原本位于光锥外的导模分量平移到光锥内；产生辐射的是折叠后的分支与自由空间连续谱的相交。

6.5 最小四波图像：为什么它能抓住 PCSEL 的主导物理

为了把上述图像变成可计算模型，先考虑方格晶格中最常见的四个主导面内分量：

$$R_x, S_x, R_y, S_y,$$

其中 R_x 和 S_x 分别表示沿 $+x$ 与 $-x$ 传播的波, R_y 和 S_y 类似。对这些分量做慢变包络近似, 可以写出耦合方程

$$\frac{d}{dt} \begin{bmatrix} R_x \\ S_x \\ R_y \\ S_y \end{bmatrix} = (\Omega - \Gamma + \mathbf{K}) \begin{bmatrix} R_x \\ S_x \\ R_y \\ S_y \end{bmatrix}. \quad (6.8)$$

这里 $\mathbf{K}$ 的非对角元代表不同传播方向之间的散射耦合, Γ 中则包含辐射损耗和内部损耗。图6.3直观展示了三种主要的表面发射机制及其统一描述。

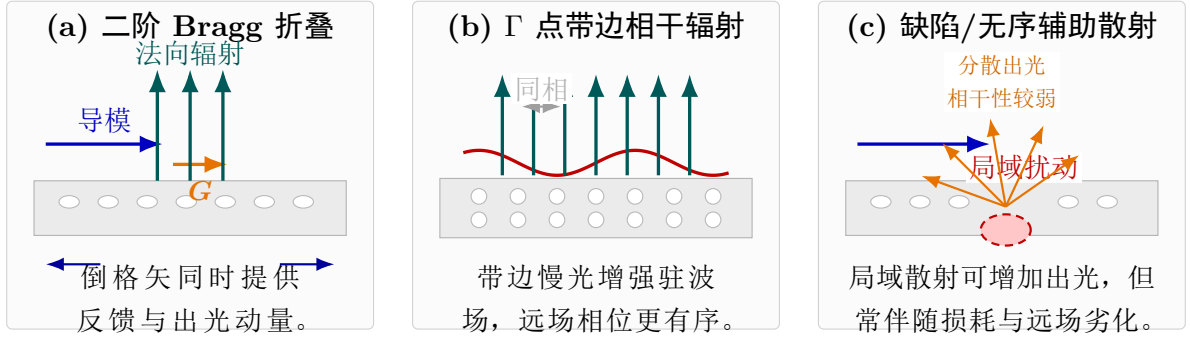


图 6.3: 光子晶体面发射激光器中三类主要面发射机制对比: 二级衍射、带边相干辐射与缺陷辅助散射。三者都可以理解为周期或局域扰动把面内导模耦合到辐射通道, 但相干性、方向性和损耗代价不同。

这里的慢变包络近似并非一句口头假设, 而有明确的尺度条件。若某支基本波写成

$$E_m(\rho) = a_m(\rho)e^{i\beta_m \cdot \rho},$$

则要求包络变化尺度远大于载波周期, 典型地可写成

$$|\nabla_{\parallel}^2 a_m| \ll |2\beta_m \cdot \nabla_{\parallel} a_m|. \quad (6.9)$$

式 (6.9) 失效时, 说明模式已经不再由少数慢变基本波主导, 此时应谨慎使用最小四波图像, 并考虑更高阶 CWT 或直接转入全波方法。

这个模型之所以有教学价值, 不在于它已经足够精确, 而在于它第一次把“模式分裂”“偏振选择”“损耗排序”“竞争关系”放进了同一个矩阵里。一旦接受这种统一表述, 后面的 CWT、RCWA 和 FEM 只是在逐层提高保真度, 而不是突然更换物理对象。

若某个结构改动主要增强了 $R_x \leftrightarrow S_x$ 这条回路, 而对 x 、 y 两组分量之间的交叉耦合作用较弱, 那么最先被改写的常常是纵横两个主偏振方向的模式分裂。反过来, 若某个改动主要打开的是通往光锥的散射分量, 那么它更可能优先改写的是辐射损耗、阈值与远场, 而不一定先改写偏振。

6.6 晶格傅里叶分量为什么如此关键

式 (6.8) 只是最小模型。它通常忽略高阶倒格矢、完整纵向矢量效应、有限尺寸边缘散射和温度依赖。它适合建立主导机制和参数方向判断, 不适合在未经校准的情况下直接替代

RCWA/FEM/FDTD 做最终定量结论。

但它已经足以说明一个需要注意的事实：孔半径、孔形状、占空比、刻蚀深度等几何参数，并不会“平均地”影响所有物理量，而是通过不同傅里叶分量选择性地改写不同耦合通道。也因此，PC layer 的几何微调不能脱离纵向外延设计单独理解；同一个几何改动，在不同外延层栈下可能对应完全不同的反馈、辐射与损耗后果。

6.7 反馈与提取之间的张力：为什么 PCSEL 设计不能只追一个指标

到这里可以把本章最重要的工程结论明确写出来：PCSEL 设计是在二者之间寻找可工作的窗口，而非“让反馈尽可能强”或“让辐射尽可能强”。

若辐射太弱，模式虽然容易积累能量，但输出提取效率差，远场可能不理想；若辐射太强，目标模还未来得及积累足够能量就被泄露掉，光学阈值反而升高。可工作的 PCSEL 设计通常不是单一指标取极值，而是多个指标同时落在可接受范围内：

- 辐射损耗足够低，使目标模可达到阈值；
- 辐射耦合又足够有序，使输出远场和方向性可控；
- 竞争模在同一结构下没有拿到更好的净增益排序。

这也解释了为什么设计讨论里会反复出现 trade-off。这里的 trade-off 来自同一组散射通道对两类过程的同时控制：腔内储能与有用输出。

6.8 模式竞争在本章图像下如何理解

现在回到上一章的模式选择问题。不同候选模之所以竞争，是因为它们对应不同的耦合网络本征态，而非因为它们名字不同。即使两个模式频率接近，如果一个模式在耦合矩阵中对应较小的辐射损耗、更好的有源区重叠和更有利的对称性，那么它在净增益排序上就可能明显占优。

换句话说，模式竞争并非“后面阈值章节才讨论的事情”，它从耦合网络建立时就已经开始了。本章给出的不是最终阈值答案，但它告诉我们：竞争的根子在哪里。

易错点

把“面发射”理解成一个独立于反馈之外的附加输出过程，是常见误读。这样一来，读者会在后面把 BIC 看成“纯辐射专题”，把阈值看成“纯增益专题”，再把动态学看成“纯速率方程专题”。对 PCSEL 而言，这三件事其实都在继承本章同一组散射通道。

6.9 有限尺寸为什么会改变本章图像

本章的大部分讨论都默认无限周期或至少大尺寸周期结构，因此耦合网络可以在 Bloch 图像下被整齐组织。真实器件一旦有限，至少会出现以下修正：

- 边缘散射新增泄露通道；
- 孔阵有限导致回路不再完美闭合；
- 注入和温度分布沿横向变化，使局部耦合条件具有空间依赖性。

因此，本章建立的是“核心反馈机制”，而不是全部器件细节。后面的 CWT、RCWA 和有限尺寸全波仿真，本质上都在做同一件事：把这里的理想耦合网络，逐步修正成真实器件中的耦合网络。

6.10 2024–2026 前沿补充：反馈网络的“调制-功率”双目标约束

近两年的器件路线给了本章一个非常具体的反向约束：PCSEL 的反馈网络不能只服务于“低阈值”，还需要同时满足高速调制与高功率输出。直接调制瓦级 PCSEL 已在自由空间链路场景中给出器件证据 [47]，随后频率调制高功率 PCSEL 与长距离相干链路也进一步验证了“高亮度 + 高速 + 系统可用性”的并行目标 [26, 4]。

对本章模型语言而言，这意味着同一组散射通道至少要同时满足三件事：

- 目标模反馈足够强，使其在竞争模中保持净增益优势；
- 法向辐射耦合足够可控，使输出方向性和远场质量可用于链路；
- 动态响应不过度受限于热漂移与寄生效应，使调制窗口可用。

因此，“把某个被动指标推到极值”通常不是最优策略。更可靠的做法是把反馈-辐射-动态三者放入同一预算：先在单胞层确认耦合方向，再在有限尺寸和器件层检查该方向是否仍保留调制与热稳定窗口。这种预算关系直接连接了本章与后续 第二十三章、第二十五章 的工作流闭环。

6.11 与后续章节的关系

本章实际上把后续几章的逻辑都搭好了。

- 第七章 将利用这里的损耗与增益图像，严格区分阈值增益和阈值电流；
- 第八章 将讨论耦合网络的对称性如何决定偏振和远场；
- 第九章 将把这里的“哪些波能向外辐射、哪些波因对称性而难以辐射”进一步翻译成 BIC / quasi-BIC 的语言；
- 第十章 则把本章的物理图像系统化耦合波理论矩阵形式。

因此，本章是从 Γ 点候选模过渡到阈值、偏振、BIC 与 CWT 的共同桥梁，而非孤立的“现象解释章”。更进一步说，本章给出的散射网络并不只决定“能否出光”，还在给后面的动力学预设边界：哪一支模式损耗更低、哪一支模式更容易耦合到法向辐射、哪一支模式更容易受结构扰动改写，都会直接影响 第二十三章 中的模式竞争、调制响应和线宽来源。

回看本章

PCSEL 的面发射机制不是“先有腔，再让光漏出来”，而是二维周期散射网络同时承担了反馈与辐射两种功能。理解这种双重功能之后，就能明白为什么晶格傅里叶分量、光锥耦合、模式竞争和提取效率会在后续章节里反复一起出现。对 PCSEL 来说，面发射并非附加效果，而是腔物理本身的一部分。也正因为如此，本章与后面的两个关键词直接相连：一条线通向第 9 章，在那里同一组辐射通道会被整理为 BIC / quasi-BIC 的辐射工程问题；另一条线通向第 7 章、第 23 章，在那里这些通道又会被重新写成阈值排序、阈上稳定性和噪声问题。若在这里把面发射理解成单纯“漏光”，后续关于 BIC、阈值和动态学的联系就会被切断。

练习题

1. [基础] 试用自己的语言解释：为什么面内反馈与垂直出光可以由同一组周期散射通道同时产生？
2. [基础] 若某结构把目标模的辐射耦合增强了一倍，你会优先检查哪三类后果？
3. [进阶] 从光锥图像出发，解释为什么某些模式虽然存在于 slab 中，却几乎不参与法向辐射。
4. [综合]（思路提示：先定层级，再列关键式）试给出一个你认为适合用最小四波模型描述的 PCSEL 场景，并说明原因。

延伸阅读

- [28, 13] 中关于周期散射、导模共振与辐射通道的讨论。
- 读完第 10 章后回看本章，会更容易把这里的图像和矩阵模型对应起来。
- 若想继续追问“哪些法向辐射是可以被对称性关闭的”，应接着读第 9 章；若想追问“这些辐射与损耗排序在阈上如何演化为模式竞争和线宽差异”，则应继续读第 23 章。

第七章 阈值、增益、损耗与约束因子

读完本章，你能

- 严格区分阈值增益、阈值电流、被动 Q 和净模增益这几个常被混淆的量；
- 从冷腔损耗和有源区重叠出发推导 PCSEL 的光学阈值条件；
- 理解为什么高 Q 不自动等于最低阈值，也不自动等于最低阈值电流；
- 明确本章结论如何接入后续的 BIC、器件电学和动态学模块。

读前准备 第 6 章、第 8 章 中关于面内耦合、辐射通道、偏振与远场的基本概念。

阅读主线 本章所处模型层级：L2

本章先问“阈值”这个词在 PCSEL 里到底指什么，再把材料增益、模增益、损耗和约束因子串成一条严格的符号链。随后我们说明被动 Q 与阈值之间的关系、为什么本章只能给出阈值增益而不能直接给出阈值电流，以及为什么阈值之上还必须继续进入动态学、注入和热反馈。

与第 6 章的连接 第 6 章分析了 PCSEL 的面内反馈机制和垂直辐射通道。现在我们有了完整的损耗图像： $\alpha_{\text{total}} = \alpha_{\text{rad}} + \alpha_{\text{loss}}$ 。本章的自然任务就是回答：需要多大的材料增益才能补偿这些损耗并开始激射？这就引出了阈值条件的严格推导。

7.1 阈值首先是净衰减被补偿的条件

若只看某一候选模式的线性小信号演化，它的复振幅可抽象写成

$$\frac{da}{dt} = (i\omega_0 - \gamma_{\text{rad}} - \gamma_{\text{int}} + \gamma_{\text{opt}}) a,$$

其中 γ_{rad} 表示辐射损耗， γ_{int} 表示内部损耗， γ_{opt} 表示有源区提供的光学增益率。于是模式的净衰减率可写为

$$\gamma_{\text{net}} = \gamma_{\text{rad}} + \gamma_{\text{int}} - \gamma_{\text{opt}}. \quad (7.1)$$

阈值的最本质定义就是：净衰减率恰好为零。

$$\gamma_{\text{opt,th}} = \gamma_{\text{rad}} + \gamma_{\text{int}}. \quad (7.2)$$

这是一个平衡条件，而非一个定义式。若把复振幅解成

$$a(t) = a(0) \exp[(i\omega_0 - \gamma_{\text{net}}) t],$$

则模强度或光子数满足

$$S(t) = |a(t)|^2 = S(0) \exp[-2\gamma_{\text{net}} t]. \quad (7.3)$$

于是只有当 $\gamma_{\text{net}} > 0$ 时，模式才会衰减；当 $\gamma_{\text{net}} < 0$ 时，线性模型会预言指数增长；而阈值点恰好是这两类行为的分界面。换句话说，式 (7.2) 的地位是“把增长和衰减的边界精确写出来”，而非“把一个符号定义成零”。

直觉框

把阈值理解成“激光刚好能自己维持住”的那一点，比把它理解成某个神秘的电流数更稳妥。因为对光场来说，最先被问到的问题始终是：每个光学周期里损失掉了多少，又补回来了多少。

7.2 从材料增益到模增益：约束因子为什么必须出现

材料增益并非直接作用在模式上的。模式只在有源区的一部分体积里有效“看到”增益，因此必须通过光学约束因子（optical confinement factor）把材料增益投影成模增益。最常用的近似写法是

$$g_{\text{mod}} = \Gamma_{\text{opt}} g_{\text{mat}}, \quad (7.4)$$

其中 g_{mat} 是材料增益， Γ_{opt} 是模式与有源区的重叠因子， g_{mod} 是模增益。在能量归一化约定下， Γ_{opt} 的标准积分定义为

$$\Gamma_{\text{opt}} = \frac{\int_{V_{\text{act}}} \varepsilon_r(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}{\int_{V_{\text{total}}} \varepsilon_r(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}, \quad (7.5)$$

其中 V_{act} 为有源区体积， V_{total} 为模式能量主要分布的全结构体积。对单量子阱（厚度约 8 nm）嵌入数百纳米波导层的典型 PCSEL 外延结构， $\Gamma_{\text{opt}} \sim 1\%$ ；三至五量子阱时约为 3–5

能量权重近似的适用条件

式 (7.5) 的能量权重形式依赖以下近似：

1. **缓变包络近似**：模式剖面在增益区域内变化缓慢，可用单一模式场 $\mathbf{E}(\mathbf{r})$ 描述
2. **标量增益近似**：材料增益在有源区内可视为局域标量，忽略各向异性和空间非均匀性
3. **小信号近似**：增益饱和效应不显著改变模式剖面形状
4. **单模近似**：忽略多模竞争和空间烧孔效应对有效 Γ_{opt} 的影响

对高功率 PCSEL 或存在强空间烧孔的情况，有效约束因子可能偏离式 (7.5) 的预测，需通过多模自洽求解获得准确的 $\Gamma_{\text{opt}}^{\text{eff}}$ 。

这里的写法刻意保持在“最小接口层”。但它不是无条件成立的定义，而依赖几条近似：材料增益在有源区内可先看作局域标量；目标模式剖面在小信号增益下不发生显著重构；极化选择和增益色散已被压缩成该模式实际看到的一条有效增益曲线。若要把 g_{mat} 写成透明载流子密度、微分增益、量子阱数目与温度失谐的函数，要等到 第十八章 才系统展开；本章先把它当作一个尚未完全展开、但已经明确属于有源区模型的输入量。

若进一步用能量速度 v_{en} 把空间增益换成时间增益率，则

$$\gamma_{\text{opt}} = v_{\text{en}} g_{\text{mod}} = v_{\text{en}} \Gamma_{\text{opt}} g_{\text{mat}}. \quad (7.6)$$

同理，若把辐射损耗和内部损耗写成沿传播方向的等效损耗系数，则

$$\gamma_{\text{rad}} = v_{\text{en}} \alpha_{\text{rad}}, \quad \gamma_{\text{int}} = v_{\text{en}} \alpha_{\text{int}}. \quad (7.7)$$

把 式 (7.7) 和 式 (7.6) 代回 式 (7.2)，才得到下面常用的阈值增益公式：

$$g_{\text{th}} = \frac{\alpha_{\text{rad}} + \alpha_{\text{int}}}{\Gamma_{\text{opt}}}. \quad (7.8)$$

量纲检查：阈值增益公式的量纲自治性

Γ_{opt} : 无量纲重叠因子; $\alpha_{\text{rad}}, \alpha_{\text{int}}$: 空间损耗系数 [m^{-1}]; g_{th} : 阈值材料增益系数 [m^{-1}]. 右边: $[\alpha]/[\Gamma] = \text{m}^{-1}/1 = \text{m}^{-1}$, 与左边 $[g_{\text{th}}] = \text{m}^{-1}$ 一致.

表 7.1: 本章阈值链中最容易混淆的量纲桥。若不先把这一层单位统一, $v_{\text{en}}g$ 与 α/Γ 的映射最容易被读错。

量	典型单位	在阈值链中的角色
$\gamma_{\text{rad}}, \gamma_{\text{int}}, \gamma_{\text{opt}}$	s^{-1}	时间增长/衰减率
τ_p	s	光子寿命
$\alpha_{\text{rad}}, \alpha_{\text{int}}$	cm^{-1} 或 m^{-1}	空间损耗系数
$g_{\text{mat}}, g_{\text{mod}}, g_{\text{th}}$	cm^{-1} 或 m^{-1}	空间增益系数
Γ_{opt}	1	重叠因子, 无量纲
v_{en}	cm/s 或 m/s	把空间增益/损耗映射成时间率的等效参数 (能量速度)

严格表述 (难度)

本书在阈值链中统一使用 v_{en} (能量速度) 表示将空间增益/损耗映射为时间增长率的参数。对弱导模 slab, v_{en} 常可用 c/n_{eff} 作数量级近似; 对 Γ 点 standing-wave 模式, 带图 Bloch 群速度为零, 但 v_{en} 仍然存在且非零, 因为两者描述的是不同层级的物理: 前者描述 Bloch 包络的传播速度, 后者描述局域光场与增益介质的相互作用速率。实际建模时, 对 GaAs 基 PCSEL 可取 $v_{\text{en}} \approx 0.9 \times 10^{10} \text{ cm/s}$ 作为参考值, 但最终应与速率方程拟合或传递矩阵法对齐。

更严格的写法可从能量速度而不是带图群速度出发。设模式总储能为 U , 有源区提供的净增益功率为 P_{gain} , 则定义时间增长率的直接方式是

$$\gamma_{\text{opt}} = \frac{P_{\text{gain}}}{U}. \quad (7.9)$$

若再把同一个模式沿传播方向的空间增益写成 g_{mod} ，并把能流定义为 P_{flow} ，则

$$P_{\text{gain}} = g_{\text{mod}} P_{\text{flow}}, \quad v_{\text{en}} \equiv \frac{P_{\text{flow}}}{U}, \quad (7.10)$$

从而得到

$$\gamma_{\text{opt}} = v_{\text{en}} g_{\text{mod}}. \quad (7.11)$$

这里出现的其实是**能量速度** v_{en} 。对弱调制 slab 波导，它常可用 c/n_{eff} 作数量级近似；因此本章写作里的 v_g ，更准确地说应理解为“把空间增益/损耗换算成时间增长/衰减率的等效能量速度参数”，而不是 exact Γ 点 band diagram 的严格 Bloch 群速度。

把式 (7.2) 和式 (7.6) 联立，就得到常见的阈值增益形式

$$g_{\text{th}} = \frac{\alpha_{\text{rad}} + \alpha_{\text{int}}}{\Gamma_{\text{opt}}}, \quad (7.12)$$

其中 α_{rad} 和 α_{int} 分别表示与辐射损耗和内部损耗对应的等效损耗系数。

这条式子需要按层级理解：它给出的首先是**光学线性阈值**，不是器件级阈值电流。

例子

[本书算例] 从损耗算到材料阈值增益。若 $\alpha_{\text{rad}} = 5 \text{ cm}^{-1}$ 、 $\alpha_{\text{int}} = 10 \text{ cm}^{-1}$ ，量子阱光学重叠 $\Gamma_{\text{opt}} = 0.02$ ，则阈值处所需模增益是 15 cm^{-1} ，对应材料增益

$$g_{\text{th}} = \frac{5 + 10}{0.02} = 750 \text{ cm}^{-1}.$$

同一组损耗在较小重叠下会被放大成很高的材料增益要求；这就是为什么只报告冷腔 Q 仍不足以判断器件阈值。

在本书的语境里， g_{th} 表示“要让某模式净衰减归零所需的模增益阈值”。它依赖该模式的辐射损耗、内部损耗和与有源区的重叠。若把这个量直接改写成 I_{th} ，等于默认注入均匀、材料增益模型已知且热漂移可忽略，而这些默认对真实 PCSEL 往往都不成立。

7.3 内部损耗的组成与标定

很多初学者把内部损耗写成一个符号后就不再追问。但对 PCSEL 而言，这往往正是器件结论出错的地方。更稳妥的写法是把内部损耗继续拆开：

$$\alpha_{\text{int}} = \alpha_{\text{FCA}} + \alpha_{\text{abs}} + \alpha_{\text{metal}} + \alpha_{\text{scatt}}. \quad (7.13)$$

这里 α_{FCA} 表示自由载流子吸收， α_{abs} 表示材料吸收或掺杂层吸收， α_{metal} 表示金属邻近引入的损耗， α_{scatt} 表示与粗糙度、侧壁和工艺缺陷相关的散射损耗。图7.1以饼图形式直观展示了典型 PCSEL 器件中各项损耗的占比情况。为避免把经验值误读为“标准常数”，本书统一采用“范围 + 条件 + 来源”写法：**[文献值]** 对近红外 GaAs 基 PCSEL/半导体激光器的代表性外延与掺杂条件，常见有 $\alpha_{\text{FCA}} \sim 1\text{--}10 \text{ cm}^{-1}$ （随掺杂与载流子密度显著变化）、 $\alpha_{\text{abs}} \sim 1\text{--}5 \text{ cm}^{-1}$ 、 $\alpha_{\text{metal}} \sim 10\text{--}100 \text{ cm}^{-1}$ （当模式尾场靠近金属层时）、 $\alpha_{\text{scatt}} \lesssim 5 \text{ cm}^{-1}$ **[10, 80]**；**[本书算例]** 辐射

光子晶体面发射激光器阈值条件下的典型损耗分解

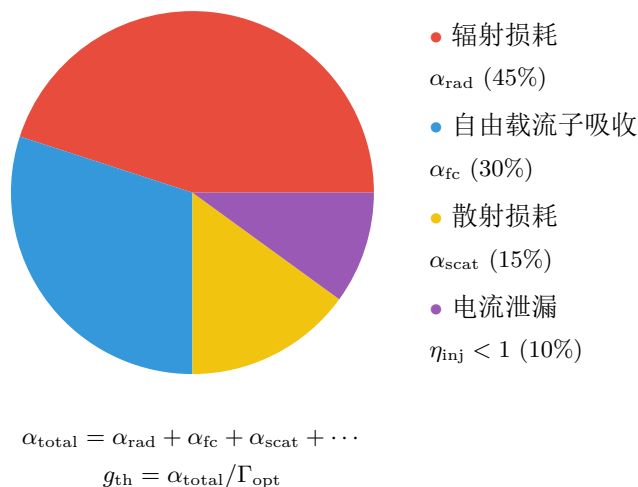


图 7.1: PCSEL 中的损耗机制分解示意图。辐射损耗通常占主导地位（尤其是对于表面发射器件），其次是自由载流子吸收和界面散射损耗。降低总损耗是实现低阈值激射的关键途径之一。注意电流泄漏虽然不直接贡献光学损耗，但会降低注入效率 η_{inj} ，从而间接提高阈值电流密度。

损耗 α_{rad} 在高- Q 候选模可低至 10^{-1} cm^{-1} 量级，在亮模可达 10 cm^{-1} 以上，具体数值必须与结构、边界与提取方法一并报告。

PCSEL 之所以不能只看冷腔模场分布，原因就在这里：只要模式在纵向上被拉向重掺杂层、金属邻近区或粗糙界面，原本“在冷腔模型中表现良好”的冷腔候选模，阈值排序就可能被内部损耗完全改写。

7.4 为什么高 Q 有利，但不等于最低阈值

从被动角度看，品质因子越高，说明能量衰减越慢。若只考虑辐射损耗，这当然通常有利于降低阈值。但对 PCSEL 来说，这个判断绝不能单独使用。高 Q 模并不必然具备以下优点：

- 与量子阱有最好的重叠；
- 对掺杂层、金属层和吸收层有最小重叠；
- 在电流注入和温升回写之后仍保持最优排序；
- 在高功率工作时仍保持合适的提取效率和远场形状。

因此，正确的说法是：高 Q 往往是低阈值的有利因素，但阈值由净模增益条件决定，而不是由 Q 单独决定。

易错点

“这个模式的被动 Q 比另一个高一倍，所以它的阈值电流一定更低”这类说法是不严谨的。它至少忽略了模式重叠、内部损耗、电流分布和热漂移四件事。

图7.2 把这个误区拆成一条更稳妥的判断链。高 Q 首先只影响辐射损耗入口；它要经过内部损耗、约束因子、注入效率和热回写之后，才可能变成低阈值电流和稳定工作窗口。

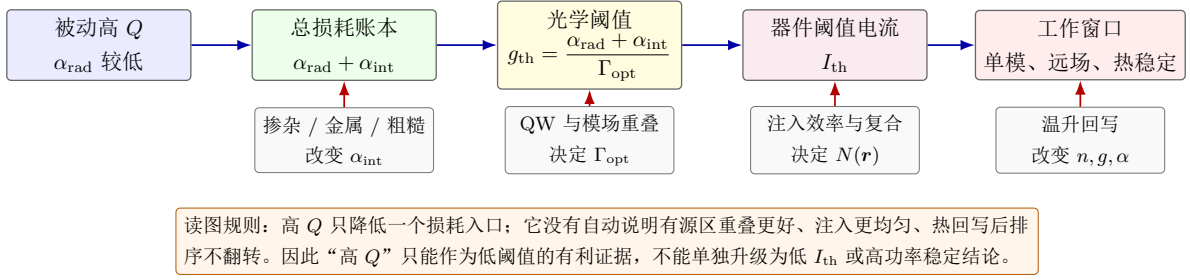


图 7.2: 从被动高 Q 到阈值和工作窗口的层级链。高 Q 有助于降低辐射损耗，但总阈值还要同时检查内部损耗、模场与有源区重叠、注入效率和热回写后的模式排序。

7.5 从速率方程再看一遍阈值

若采用单模近似，光子数 S 的最简速率方程可写为

$$\frac{dS}{dt} = v_g [\Gamma_{opt} g_{mat}(N, T) - \alpha_{rad} - \alpha_{int}] S + \beta_{sp} R_{sp} V_{mode}. \quad (7.14)$$

其中 N 为等效载流子浓度， T 为温度， $\beta_{sp} R_{sp} V_{mode}$ 表示耦合入该模式的自发辐射项， V_{mode} 为模式体积。

注记

速率方程的量纲一致性。式式 (7.14) 中各项必须具有相同量纲：

- S 是光子数（无量纲计数），故 dS/dt 的单位是 s^{-1}
- $v_g[\dots]S$ 项： v_g (m/s) $\times$ α (m^{-1}) $\times$ $S \rightarrow s^{-1}$ ✓
- $\beta_{sp} R_{sp} V_{mode}$ 项： β_{sp} (无量纲) $\times$ R_{sp} ($m^{-3}s^{-1}$) $\times$ V_{mode} (m^3) $\rightarrow s^{-1}$ ✓

若省略 V_{mode} ，则自发辐射项的量纲将是 $m^{-3}s^{-1}$ ，与光子数变化率的量纲不匹配。

在单模、小信号、给定模式剖面的近似下，阈值条件就是光子数不再指数衰减：

$$\Gamma_{opt} g_{mat}(N_{th}, T_{th}) = \alpha_{rad} + \alpha_{int}. \quad (7.15)$$

于是式 (7.2) 与式 (7.15) 只是同一件事的两种写法：前者强调时间衰减率，后者强调模增益与损耗的平衡。

若希望再从开放模语言看一遍，可把候选模的复频率写成

$$\tilde{\omega} = \omega_r - i(\gamma_{\text{rad}} + \gamma_{\text{int}} - \gamma_{\text{opt}}). \quad (7.16)$$

在本书采用的时间因子约定下， $\Im \tilde{\omega} < 0$ 对应振幅随时间衰减， $\Im \tilde{\omega} > 0$ 对应线性增长，而阈值则对应

$$\Im \tilde{\omega}_{\text{th}} = 0. \quad (7.17)$$

因此，从复频率角度看，阈值条件并非另一条新物理，而只是把“净衰减归零”重写成“复频率虚部穿过零”的开放系统表达。它和式 (7.2) 和式 (7.15) 必须给出同一个阈值。

这也正是后文 CWT 阈值条件的代数来源。若第 j 个 CWT 超模在固定模式归一化下可写成

$$\lambda_j = v_{\text{en},j} [\Gamma_{\text{opt},j} g_{\text{mat}} - \alpha_{\text{rad},j} - \alpha_{\text{int},j}], \quad (7.18)$$

那么“哪一支超模先达到阈值”就等价于寻找最先满足 $\text{Re } \lambda_j = 0$ 的那一支。也就是说，式 (7.12) 和式 (7.15) 与后文 CWT 的 ‘ $\max \text{Re}(\lambda_j) = 0$ ’ 并非两套相互竞争的判据，而是单模阈值公式与多模耦合阈值公式的同一条桥。

式 (7.2)、式 (7.15) 和式 (7.17) 给出的都只是**线性光学阈值**：它们回答“哪一个候选模最先补偿净损耗”。它们不是阈上稳态激光解，因为尚未引入饱和和有源自洽场；也不是阈值电流，因为尚未求解注入分布、复合损耗与热回写。

7.6 非均匀增益为什么会改写模式排序

真实器件里，增益往往不是均匀分布的。电流扩展、电流拥挤、热斑和载流子泄漏都会让局域增益依赖空间位置。此时更合适的模增益表达式是

$$g_{\text{mod}} = \frac{\int_{V_{\text{active}}} g_{\text{mat}}(\mathbf{r}) \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}{\int_{V_{\text{all}}} \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}. \quad (7.19)$$

这里分母中的 V_{all} 应按“模式实际占据的全部空间”理解，而不是只积到波导芯层或有源层附近一个小盒子为止。对开放 slab 共振模，它至少要覆盖主要场能量所在区域，包括向上下包层延伸的倏逝尾场；否则约束因子和模增益都会被系统性高估。实际截断时，建议与能量归一化所用积分域保持一致：取一个使分母积分值与无穷域积分值之差小于 1% 的有限截断面，并在收敛检查中验证该截断不影响最终约束因子结果。

式 (7.19) 的意义非常深刻。它说明：即便两个模式拥有相同的冷腔损耗，若它们在横向上的场分布不同，而注入分布又不均匀，那么它们得到的有效模增益也会不同。这也是后面电注入和热效应能够改写模式排序的根本原因。

7.7 为什么本章只能给出阈值增益，不能给出阈值电流

到这里，读者可能会问：既然阈值条件都写得这么明确了，为什么还不能直接算 I_{th} ？原因并不复杂。要从 g_{th} 走到 I_{th} ，至少还需要解决三件事：

1. 材料增益 g_{mat} 如何依赖载流子浓度和温度；
2. 外加电流如何通过漂移-扩散和复合过程产生该载流子分布；
3. 温升如何反过来改写折射率、损耗和增益谱。

这三步分别对应后面的有源区、电学和热耦合章节。换句话说，本章完成的是从冷腔模式到光学阈值的桥接，而不是从模式直接跳到器件电流。

7.8 为什么本章故意停在静态阈值，而不在这里直接讨论动态学

从教学顺序上看，这里有必要再把边界说清楚。阈值条件回答的是“哪一个模式最先有资格起振”，而不是“起振以后它会怎样演化”。只要器件进入阈值以上区域，载流子抽空、增益压缩、模式竞争和热回写就会一起出现；这已经超出了本章的静态线性阈值范围。

因此，本章故意把问题停在 g_{th} 与阈值电流来源的分界线上。这样做是为了确保读者先把静态层级分清，再进入后面的 第二十三章 去讨论弛豫振荡、调制响应和线宽来源，而非回避动态学。若在这里就把阈值、稳定性和线宽混写在一起，最容易造成的错误恰恰是重新把“能起振”误当成“能稳定单模工作”。

7.9 从冷腔候选模到主动激光模：本书采用的理论层级

到这里还应再补一块总说明。否则读者虽然已经知道“被动候选模、线性阈值模、阈上工作模不是一回事”，却仍可能不知道它们在更高层理论里究竟怎样排位。对本书而言，最有用的分层是：

1. **冷腔候选模**：这是前面能带、BIC、CWT、RCWA、FDTD/FEM 反复追踪的对象。它们是在无饱和、无反转动态、或只含给定损耗的开放 Maxwell 问题里得到的候选共振模或候选模族。
2. **小信号阈值模**：这是在既定冷腔候选模基础上，引入未饱和增益之后得到的线性起振判据。此时我们问的是“哪一支模最先把净损耗补偿到零”，而不是“阈上稳态场到底长什么样”。
3. **主动激光模**：这是更高层的有源自洽对象。它不再把光场视为在固定复介电常数中的被动或线性响应解，而是要求 Maxwell 方程、介质极化、反转分布和饱和非线性共同满足自洽条件。若进一步讨论阈上稳态或随机噪声，还会进入 Maxwell-Bloch 或 steady-state *ab initio* laser theory (SALT) 一类形式化 [66, 67]。

这三层之间的关系，是“前两层是被设计流程反复证明有用的工作近似，第三层则给出更完整的顶层坐标系”，而非“前两层错误、第三层才正确”。本书把冷腔模与小信号阈值模作为主线，并非否认主动激光模理论的重要性，而是基于三点考虑。

第一，对 PCSEL 入门与器件设计最先起决定作用的，通常是模式候选、损耗排序、注入分布和热反馈的主导关系，而非完整阈上自洽解。这些问题若没先理清，即使直接写出更高层形式化，也很难得到可操作的设计判断。

第二，主动激光模理论要求更完整的材料极化与反转模型。这会把教学重点迅速推向 Maxwell-Bloch 或更复杂的有源场自洽求解，而偏离本书“从 Maxwell/Bloch/CWT/数值方法走到器件建模链”的主线。

第三，真实 PCSEL 的器件不确定性常常先于顶层主动模理论成为瓶颈。外延层、注入非均匀、热回写和有限尺寸边界若尚未被清楚建模，那么把全部精力投入阈上自洽场理论，未必比先把冷腔候选模和小信号阈值链条做扎实更有效。

因此，本书采用的立场是：冷腔模理论给出候选模目录，小信号阈值理论给出起振排序，而主动激光模理论提供更高层的理论封顶。前两者构成本书的工作主线，第三者则作为读者应当知道其存在、知道其位置、但本书不把它展开成主教学骨架的上层框架。只要把这三层关系说清楚，就不会再把“冷腔共振频率”“线性阈值模”与“真实阈上激光模”混成同一个概念。

7.10 本章应输出的结果

如果把本章用在研究工作流中，最有价值的输出不应只是一个单独的 g_{th} 数字，而应至少包括：

- 各损耗分量占比：辐射损耗与内部损耗哪一项主导；
- 目标模与主要竞争模的阈值增益差；
- Γ_{opt} 与损耗重叠因子随几何、温度或外延方案的变化；
- 哪些参数一改动，就会显著改变净增益排序。

这些结果比“某模式阈值增益为某数值”更能指导后续设计。图7.3以柱状图形式直观展示了不同结构参数下阈值增益各成分的贡献对比，帮助设计者快速识别限制性能的关键因素。

7.10.1 从注入电流到阈值条件：为什么需要一张 Sankey 图

公式链条

$$I_{inj} \rightarrow N(\mathbf{r}) \rightarrow g_{mat} \rightarrow g_{mod} \rightarrow g_{th}$$

在文字里看起来很线性，但实际每一跳都会“损耗/折算”一部分可用增益。为把这种逐级折算直观化，图7.4给出一张教学型 Sankey 示意图。它不追求某个器件的精确百分比，而是强调：阈值电流上升往往是多个环节的小损失叠加而成，而非由某一项单独造成。

这张图和本章公式可一一对上：注入侧折算对应式 (7.19) 的空间非均匀项；复合侧折算对应“材料增益不是全都转化为受激增益”；模场侧折算对应式 (7.5) 中的约束因子；最终只剩能对冲式 (7.8) 右端总损耗的那一部分。对工程设计而言，这比单独报告一个 g_{th} 数字更有诊断价值，因为它直接告诉你下一步该优先优化“注入”“重叠”还是“损耗”。

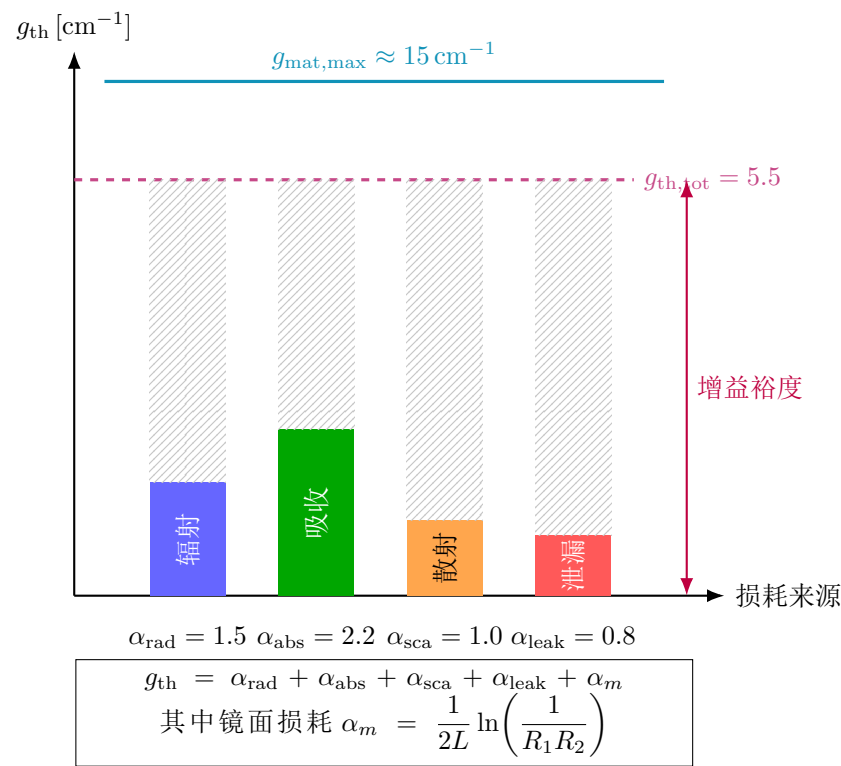
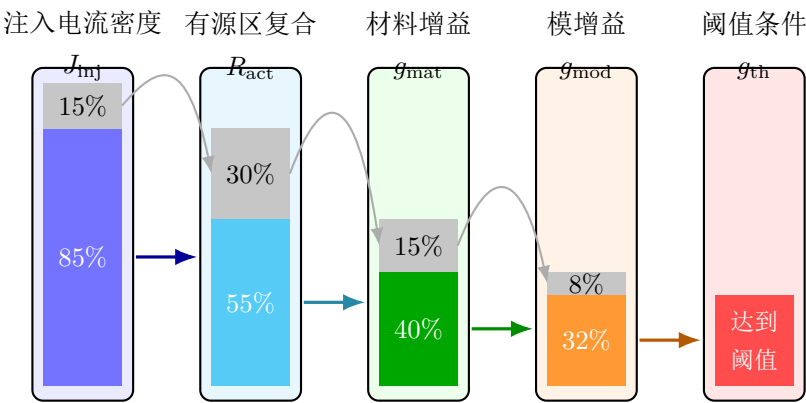


图 7.3: PCSEL 阈值增益分解示意图，展示辐射、吸收、散射与泄漏各项损耗贡献及增益裕度。



注：比例为教学型阈值链示例（用于说明“电流到阈值”的逐级折算），不同外延与注入方案会改变每一级分配。

图 7.4: 阈值电流密度链路的 Sankey 风格示意。从注入电流到阈值条件并非一步映射，而是经历注入效率、辐射复合效率、增益谱匹配和模场重叠的逐级折算。图中灰色支路表示在各环节被“损耗/失配”消耗的份额。

回看本章

本章把阈值问题严格放回了它应在的层级：阈值首先是净模增益补偿总损耗的光学条件。材料增益必须通过约束因子与模式重叠转化为模增益；总损耗则必须拆成辐射与内部各项来源。只有把这些关系讲清楚，后面讨论阈值电流、电注入和热效应才不会偷换概念。与此同时，辐射损耗这一项并非孤立参数：在第九章中，读者会看到同一类 Γ 点辐射通道如何被 BIC / quasi-BIC 语言重新组织；而在第二十三章中，读者会看到“谁先达到阈值”还必须继续延伸为“阈上谁更稳定、谁的调制更快、谁的线宽更窄”。这也是 PCSEL 中静态阈值、辐射工程与动态学不可拆开的原因。

练习题

1. [基础] 从复频率定义出发，推导被动 Q 与等效损耗系数之间的关系。
2. [基础] 解释为什么同样的材料增益谱，对两个场分布不同的模式会给出不同的模增益。
3. [进阶] 给出一个 Γ_{opt} 增大但阈值电流不降反升的物理场景。
4. [综合]（思路提示：先定层级，再列关键式）将式 (7.19) 与电注入非均匀联系起来，说明它为什么是后续器件建模的关键接口式。

延伸阅读

- [10, 61, 19] 中关于阈值条件、速率方程与线宽来源的经典章节和论文。
- 读完第十八章、第十九章后回看本章，会更清楚“阈值增益”与“阈值电流”之间相隔哪些物理环节。
- 若你想继续追问“辐射损耗本身是怎样被结构对称性重排的”，可接着读第九章；若你想追问“达到阈值以后会怎样演化”，则应直接进入第二十三章。若你还想知道这条链在更高层有源自洽理论中的位置，可把 [66, 67] 作为背景参考，而不必把它当成本书主线的前置条件。

第八章 偏振、对称性破缺与远场整形

读完本章，你能

- 理解偏振为什么是模式本征属性，而不是输出端附加标签。
- 掌握近场对称性、晶格对称性和远场角谱之间的关系。
- 建立 Γ 点小群（little group）与不可约表示（irreducible representation）的最小分类工具。
- 弄清楚对称性破缺如何同时影响偏振选择、模式分裂和远场整形。

读前准备 第**五**章、第**六**章 中的 Gamma 点模式、耦合网络、模式分裂与辐射通道概念。

阅读主线 本章所处模型层级：**L1**→**L2**

本章先说明偏振为何属于腔内本征场结构，再讨论对称性对偏振和辐射通道的约束。随后引入 Γ 点小群与不可约表示，给出方格晶格和三角晶格常用的模式分类表。最后从傅里叶角谱角度连接近场与远场，并讨论对称性破缺、外延不对称和工艺扰动对输出偏振与远场的影响。

8.1 偏振不是输出端的装饰，而是模式本身的一部分

很多初学者会下意识地把偏振理解成“光出来以后我们看到的方向”。这种想法对自由空间光束并不完全错误，但对 PCSEL 来说远远不够。PCSEL 的偏振首先是模式本征场结构的一部分，只有先在腔内场分布中形成了某种偏振组织，输出端才会看到对应的偏振态。

因此，偏振是模式解的组成部分，而非附加在模式之后的标签。模式的对称性、主要场分量、与辐射通道的耦合方式，以及与外延不对称的匹配程度，共同决定测得的线偏振、近似双偏振或工作点相关的偏振切换。

直觉框

对 PCSEL 而言，偏振分析的对象是腔内矢量模的选择，以及该模与可辐射通道的对应关系。

8.2 为什么对称性会控制偏振

上一章已经说明，PCSEL 的目标模通常位于高对称晶格的 Gamma 点附近。高对称性带来的一个直接后果，就是某些模式会形成简并或近简并模对，而这些模对往往对应不同的偏振方向或不同的场分量组合。

在理想完全对称的结构中，两个正交偏振态可能处于严格相同的地位。此时系统本身并没有偏好哪一个方向，实验上就可能表现为偏振不稳定、双偏振共存，或者对极小扰动极其敏感。一旦引入哪怕很小的对称性破缺，这种“平衡”就会被打破，模式分裂随之出现，于是系统开始偏向某一组偏振态。

因此，偏振选择问题可归结为：哪一种扰动打破原有对称平衡，以及该扰动分别如何改变模式频率和损耗。

8.3 Γ 点小群与不可约表示：最小分类工具

如果只停留在“镜面对称、旋转对称、奇偶性”这些词上，读者虽然能听懂很多物理解释，但还没有真正得到一套可操作的分类工具。本节只做三件事：先解释小群是什么、再说明 irrep 标签怎么贴、最后给出法向辐射的最小判据。

第一步：小群是“在某个 $\mathbf{k}$ 处还活着的对称性”。对任意 Bloch 波矢 $\mathbf{k}$ ，其小群定义为

$$\mathcal{G}_{\mathbf{k}} = \{g \in \mathcal{G} \mid g\mathbf{k} = \mathbf{k} + \mathbf{G}, \mathbf{G} \in \text{倒格矢集合}\}, \quad (8.1)$$

其中 $\mathcal{G}$ 是晶格的点群。直觉上，这个定义只是在说：对称操作作用到 $\mathbf{k}$ 后，如果它仍落在“同一个 Bloch 态”的等价点上（差一个倒格矢），那这个对称操作在该 $\mathbf{k}$ 处仍然有效。

对 Γ 点而言， $\mathbf{k} = \mathbf{0}$ ，因此

$$\mathcal{G}_{\Gamma} = \mathcal{G}. \quad (8.2)$$

因此， Γ 点适合进行对称性标注：此时小群等于完整点群，模式标签和选择规则最明确。

把“点群”想成整套对称工具，把“小群”想成“在某个 $\mathbf{k}$ 处还能用的那一部分工具”。 Γ 点不丢任何工具，所以最适合做标准化标签。

第二步：irrep 标签回答“对称操作作用后场怎样变化”。不可约表示（irreducible representation, irrep）的主要用途不是记忆表格，而是给出三条可操作结论：

1. **能不能混合：**同一 irrep 内的模式在保持对称性的扰动下可以混合，不同 irrep 的模式不能直接混合；
2. **有没有成对：**二维 irrep 对应受对称性保护的简并，常与双偏振、双简并候选模有关；
3. **能不能直射：**模式是否能耦合到法向辐射通道，首先取决于它的 irrep 是否与辐射通道的 irrep 匹配。

表 8.1: C_{4v} 的最小特征标表。这里的列是共轭类，而非单个操作。对 PCSEL 最常用的用途，是判断某个 singlet/doublet 在对称操作下属于哪一种 irrep。

irrep	E	$2C_4$	C_2	$2\sigma_v$	$2\sigma_d$
A_1	1	1	1	1	1
A_2	1	1	1	-1	-1
B_1	1	-1	1	1	-1
B_2	1	-1	1	-1	1
E	2	0	-2	0	0

更直观的说法是：对称操作作用到模场上，场要么保持不变（特征标 +1）、要么变号（特征标 -1）、要么与另一支模式互换（二维表示）。这三个结果已经足够把绝大多数 Γ 点模式分到 $A_1/A_2/B_1/B_2/E$ 之中。

表中的列名最好带着几何图像一起记。 E 是恒等操作； $2C_4$ 表示 $\pm 90^\circ$ 旋转； C_2 是 180° 旋转； $2\sigma_v$ 指穿过晶格主轴的两条镜面，最常见的就是 $x=0$ ($\sigma_{v,x}$) 和 $y=0$ ($\sigma_{v,y}$)； $2\sigma_d$ 则指两条对角镜面 $y=\pm x$ ($\sigma_{d,+}$ 和 $\sigma_{d,-}$)。这里的下标 v 就是主轴（vertical）镜面， d 就是对角线（diagonal）镜面。判断一个模属于哪种 irrep 的简单操作步骤：取模的近场分布图，分别对以上四条镜面做反射，观察场图是否偶对称（特征标 +1）或奇对称（特征标 -1），与表格各行对比即可确定。

更完整的生成元列表、直积表、 $C_{4v} \rightarrow C_{2v}$ 分支律和选择规则实例，已汇总为附录中的“群论工具箱”（第28.13节），需要时查阅即可，不必在此处一次性记住。

若两支近简并模在 $\pm 90^\circ$ 旋转下互相交换，而在某条镜面下分别变号/不变，那么更正确的做法是把它们合在一起作为二维 E 双重态，而非给它们各贴一个 singlet 标签。

第三步：法向辐射的最小判据。 在 Γ 点 ($k_{\parallel}=0$)，法向辐射对应的自由空间平面波是横向均匀平面波，其两支线偏振 ($\hat{x}, \hat{y}$) 在 C_{4v} 语言中张成二维表示 E 。因此：若腔模的 irrep 与辐射通道的 irrep 不匹配，则法向辐射耦合矩阵元在对称性上必为零。这也是 A_1 、 A_2 、 B_1 、 B_2 四类非简并模往往成为 dark/BIC 候选的群论原因，而 E 表示则是典型亮模候选 [57, 28, 23]。

上述判据的正式版本可写成“张量积包含全对称表示”。但在初次使用阶段，只需记住：模式 irrep 必须与法向辐射通道 irrep 相容，耦合才可能非零。

图8.1 给出这条规则的最小工作流。它把“看场图觉得亮或暗”的经验判断，改写成 three-step gate：先确定 little group，再给模式和辐射通道分别贴 irrep，最后检查张量积是否含全对称表示。

易错点

必须把一句话反复记住：**in-plane little-group 匹配只是必要的第一层筛选，不是三维 slab 中法向辐射强度的充分判据**。一个模式即便在平面内表示上允许 direct normal radiation，真实辐射强度仍会继续受纵向奇偶性、层栈不对称、上下光锥和有限厚度矢量混合的共同影响。

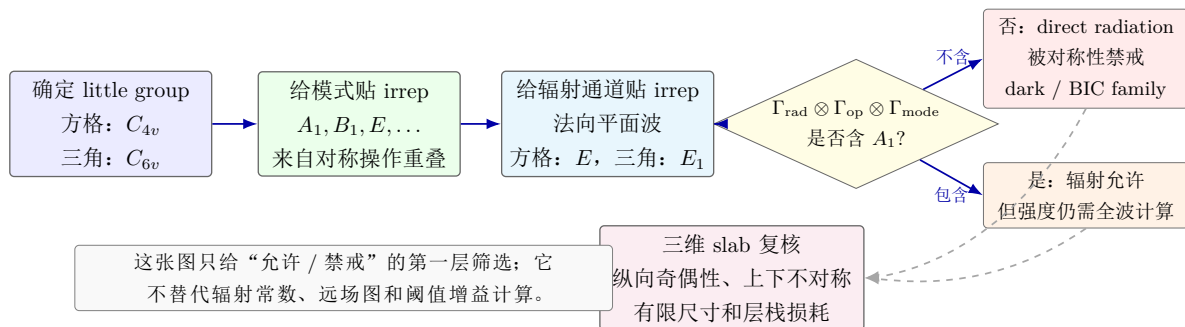


图 8.1: Gamma 点模式的 irrep-辐射通道守门流程。in-plane little-group 匹配只决定 direct normal radiation 是否被对称性允许; 真正的辐射强度仍需结合三维 slab、纵向奇偶性、层栈不对称和有限尺寸计算。

8.4 方格晶格 Γ 点: C_{4v} 分类

对理想方格晶格, Γ 点小群是

$$\mathcal{G}_{\Gamma} = C_{4v}.$$

它有四个一维不可约表示 A_1, A_2, B_1, B_2 和一个二维不可约表示 E 。对 PCSEL 读者来说, 最重要的是知道这些 irrep 分别意味着什么, 而非把角色表背下来。一个可反复使用的最小表见表 8.2。

表 8.2: C_{4v} 小群在 Γ 点的最小模式分类工具。这里的“典型基函数”是帮助读者建立 irrep 直觉的代表性多项式或向量, 而非场的完整表达。

irrep	维数	典型基函数 / 图像	对 PCSEL 的最小含义
A_1	1	$1, x^2 + y^2$, 全对称	标量型 singlet; 可作为“全对称模”标签, 但自动不等于可法向辐射模
A_2	1	R_z , 旋转型	标量型 singlet; 常与“旋涡式”或伪标量变换性质相关, 法向辐射通常受限
B_1	1	$x^2 - y^2$	轴向四瓣型 singlet; 常与沿晶格轴的场型差异有关
B_2	1	xy	对角四瓣型 singlet; 常与沿对角方向的场型差异有关
E	2	(x, y)	二维双重态; 与两支正交线偏振、法向平面波辐射通道和 Γ 点双简并最直接相关

对 PCSEL 最常用的结论是: 在 C_{4v} 下, 法向自由空间的两支横向平面波偏振 $(\hat{x}, \hat{y})$ 张成二维表示 E 。因此, 若只看平面内 little group 选择规则, 直接允许法向辐射的模式族首先应与 E 表示匹配; 属于 A_1, A_2, B_1, B_2 的模式, 在 exact Γ 点更容易成为 symmetry-protected dark family。

8.5 三角晶格 Γ 点: C_{6v} 分类

对理想三角晶格或等价的六重旋转晶格, Γ 点小群为

$$\mathcal{G}_\Gamma = C_{6v}.$$

它的常用不可约表示包括四个一维表示 A_1, A_2, B_1, B_2 和两个二维表示 E_1, E_2 。对 PCSEL 来说, 表 8.3 是比完整角色表更有用的最小摘要。

表 8.3: C_{6v} 小群在 Γ 点的最小模式分类工具。

irrep	维数	典型基函数 / 图像	对 PCSEL 的最小含义
A_1	1	1, 全对称	singlet; 常作“全对称模”标签
A_2	1	R_z	singlet; 与旋转变换性质相关
B_1, B_2	1	更高阶标量型图样	在最小模式分类中常作为一维 dark family 处理
E_1	2	(x, y)	二维双重态; 与法向自由空间横向偏振直接匹配, 是最直接允许法向辐射的 irrep
E_2	2	$(x^2 - y^2, 2xy)$	二维双重态; 常对应 quadrupole-like family, 在 exact Γ 点更容易成为 symmetry-protected BIC 候选

与方格晶格完全平行, 三角晶格的法向自由空间横向偏振 $(\hat{x}, \hat{y})$ 在 C_{6v} 下张成 E_1 。因此, E_1 是最直接的法向辐射表示, 而 E_2 则经常出现在“高对称、双简并、但 exact Γ 点不直接向法向辐射”的候选模讨论中。

8.6 法向辐射通道本身也有表示: 这是 BIC 判断的最小规则

前面直觉性地说“某模与某辐射通道是否匹配”, 现在可以把这句话写得更正式。设 Γ 点某个模式属于 irrep α , 法向辐射通道属于 irrep β , 则对称性允许的直接耦合矩阵元可记作

$$M_{\beta\alpha} = \left\langle \text{radiation}, \beta \left| \hat{\mathcal{H}}_{\text{rad}} \right| \text{mode}, \alpha \right\rangle. \quad (8.3)$$

若 α 与 β 不匹配, 则在保持该对称性的前提下 $M_{\beta\alpha} = 0$ 。这就是 symmetry-protected BIC 最小判断规则的群论版本。

8.6.1 把选择定则写成真正可复用的张量积条件

若要把上面的口头规则升级成可重复使用的教材工具, 就应把“允许耦合”写成群表示的张量积条件。设模式属于 irrep Γ_{mode} , 辐射通道属于 irrep Γ_{rad} , 而耦合算符属于 irrep Γ_{op} , 则矩阵元

$$\langle \text{rad} | \hat{\mathcal{O}} | \text{mode} \rangle$$

只有在

$$\Gamma_{\text{rad}} \otimes \Gamma_{\text{op}} \otimes \Gamma_{\text{mode}} \supset A_1 \quad (8.4)$$

时才可能非零, 其中 A_1 表示单位表示。式 (8.4) 才是“选择定则”的正式版本。

这里的“ $\supset A_1$ ”可表述为：将辐射通道、耦合算符和模式三个表示做张量积后，若其分解中包含全对称表示 A_1 ，则单胞积分可能非零；若不包含 A_1 ，则该耦合矩阵元在保持对称性时为零。 A_1 在这里表示单胞平均后可留下的标量分量。

对 exact Γ 点法向直接辐射这一最常见场景，若耦合算符本身保持单胞的点群对称性，则可将 Γ_{op} 近似视为单位表示 A_1 。于是式 (8.4) 简化为

$$\Gamma_{\text{rad}} \otimes \Gamma_{\text{mode}} \supset A_1. \quad (8.5)$$

在本书使用的最小实表示框架中，这通常进一步退化为一最便于操作的判据：**模式 irrep 与法向辐射通道 irrep 相同或含有与之兼容的分量时，direct radiation 在对称性上才是允许的**。因此前文会把“方格晶格看 E ，三角晶格看 E_1 ”说成第一性筛选规则，而不只是经验口诀 [57, 23]。

式 (8.4) 给出的只是对称性允许性，不是耦合强度大小。也就是说，满足该条件只说明矩阵元可以非零，并不保证它一定很大；而一旦不满足，则在保持对应对称性的前提下矩阵元必须为零。对 PCSEL 而言，这也是“允许法向辐射”和“实际辐射有多强”必须分两步判断的原因。

于是，对 PCSEL 最有用的两句话是：

- 对方格晶格，法向辐射通道最小地属于 E ；
- 对三角晶格，法向辐射通道最小地属于 E_1 。

因此，拿到一个新结构后，读者应先问的是“它在 Γ 点属于哪个 irrep，这个 irrep 是否与法向辐射通道匹配”，而非“这模看起来亮不亮”。若不匹配，则 exact Γ 点直接法向辐射被对称性禁止；若匹配，则 direct radiation 至少在对称性上是允许的，后面再讨论耦合强弱、损耗大小和器件级阈值。

8.6.2 如何从数值模场中真正给出 irrep 标签

教材若只告诉读者“先标 irrep”，但不说数值上怎么标，这套工具就还不够可操作。最小可执行流程如下。

对单个近似非简并模式 u ，先选取 little group 的生成元，例如方格晶格常取 C_4 与一条镜面 σ_v 。把数值模场作用一次对称变换后，与原场做归一化重叠：

$$\eta_g = \frac{\langle u, \hat{T}(g)u \rangle}{\langle u, u \rangle}, \quad g \in \mathcal{G}_\Gamma. \quad (8.6)$$

若 u 是 singlet，且数值误差足够小，则 η_g 会接近该 irrep 在生成元上的角色值，例如 ± 1 。这就能把它初步标到 A_1, A_2, B_1, B_2 一类。

若模式是近简并的双重态，则不应分别给两支模单独贴标签，而应先在该近简并子空间中构造表示矩阵

$$D_{ij}(g) = \frac{\langle u_i, \hat{T}(g)u_j \rangle}{\sqrt{\langle u_i, u_i \rangle \langle u_j, u_j \rangle}}, \quad (8.7)$$

再通过 $\text{tr } D(g)$ 在生成元上的取值，判断它属于 E 、 E_1 还是 E_2 一类二维 irrep。随后若再引入小扰动，就能直接追踪这个双重态如何分裂成 singlet，以及哪一支与目标偏振、法向辐射和 BIC 条件更接近。

这一步的意义是：**irrep** 不应仅凭场图外观判定，而应由对称操作在数值模场子空间中的表示给出。采用这一流程后， Γ 点模判别、偏振分裂和 BIC 初筛可使用同一套标注方法。

8.7 从对称操作看模式：为什么有些辐射能出，有些出不来

设结构具有某个对称操作，例如镜面对称、旋转对称或反演对称。若一个模式在该对称操作下是偶的，另一个模式是奇的，那么它们与外部辐射通道的耦合规则往往完全不同。最简单的例子是镜面对称平面：

- 若近场在某方向上是偶对称，则其傅里叶变换在零角附近通常有非零贡献；
- 若近场在某方向上是奇对称，则其角谱在对应方向上往往在中心发生抵消。

这解释了为什么有些模式虽然存在，却在法向方向上几乎不辐射；也解释了为什么小小的几何扰动可能会显著改变远场主瓣。因为扰动不仅改变了“场有多强”，还改变了“不同位置上的场如何相干相加或相互抵消”。

8.8 从近场到远场：为什么傅里叶角谱是理解 PCSEL 远场的核心

在辐射近似下，远场图样可由出光面附近的横向电场分布通过傅里叶变换得到。记横向近场为 $\mathbf{E}_t(x, y)$ ，则远场角谱幅度可写为

$$\tilde{\mathbf{E}}(k_x, k_y) = \iint \mathbf{E}_t(x, y) e^{-i(k_x x + k_y y)} dx dy. \quad (8.8)$$

式 (8.8) 表明，远场由近场中不同位置的相位和振幅在角域中叠加得到。因此，近场的奇偶性、旋转对称性和局部相位关系会直接影响远场图样。

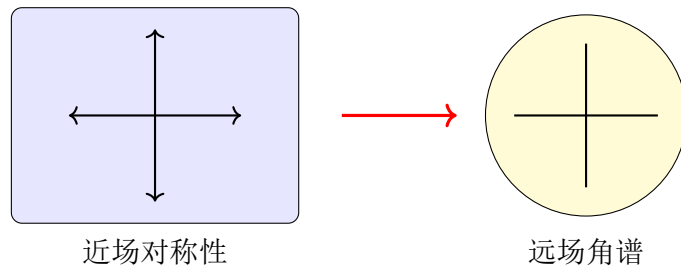
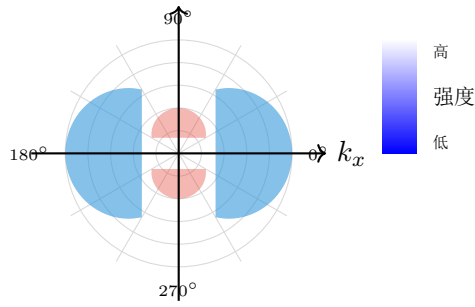


图 8.2: 近场对称性通过傅里叶变换映射为远场角谱结构的概念图。

例如，若近场在 x 方向关于中心是奇对称的，那么在 $k_x = 0$ 附近，来自左右两侧的贡献会相互抵消，远场中心就可能出现零点或主瓣分裂。反之，若近场是偶对称的，则中心角附近更容易形成主瓣。这种“对称性 $\rightarrow$ 远场中心结构”的映射，是 PCSEL 远场整形的基本判据。图 8.3 展示了两种典型模式的远场辐射方向图对比。

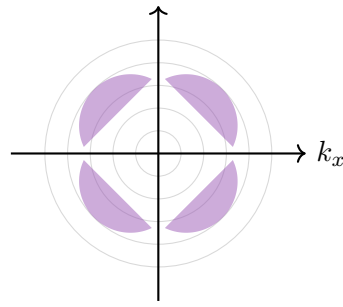
光子晶体面发射激光器典型远场辐射方向图



(a) 基模偶极辐射

主瓣半角宽度 $\sim 10^\circ\text{--}20^\circ$

高阶模四极辐射



(b) 四极矩模式

特征四叶草形状

图 8.3: 光子晶体面发射激光器的远场辐射方向图。(a) 基模（通常是偶极模）显示出双瓣特性，主辐射方向沿晶格主轴；(b) 高阶四极模呈现特征的四叶草形状，表明存在方位角依赖的辐射干涉。这些图案可以直接从傅里叶变换的近场分布计算得到，也可通过角度分辨电致发光/光致发光测量获取。注意中心 Γ 点的辐射强度与模式的对称性密切相关，某些对称性保护的连续谱束缚态模式在完美结构中完全不辐射（暗模），只有在对称性破缺后才泄漏出有限辐射。

8.9 偏振选择为什么常常和模式分裂绑在一起

回到上一章的模式竞争图像。若两个正交偏振候选模在理想结构中简并，那么真实器件需要某种机制让其中一个在净增益排序上占优。常见的路径有两条：

8.9.1 频率分裂路径

几何或外延扰动使两偏振模频率分开，从而让其中一个更靠近增益谱峰值。

8.9.2 损耗分裂路径

即使频率几乎不变，扰动也可能让某一偏振态与辐射通道或吸收区耦合更强，从而改变阈值。

对 PCSEL 而言，第二条路径尤其常见，因为偏振与辐射选择规则紧密相关。许多偏振工程并非单纯改变偏振方向，而是在重新分配不同偏振态与辐射/损耗通道的相对权重。

图 8.4 给出了四模三重态图像的完整可视化：在弱耦合极限下，正向行波 $(1,0)$ 与反向行波 $(-1,0)$ 形成驻波配对，而侧向行波 $(0,\pm 1)$ 由于对称性保持简并，最终形成三支非简并模——两支 singlet 加一支 doublet。

图 8.4 显示了 PCSEL 模式结构的层级组织：先由行波重组为驻波，再按 irrep 进行对称性分类，最后由微扰解除剩余简并。这一结构用于后续偏振选择和模式分裂分析。

8.10 对称性破缺如何在 irrep 语言下导致分裂

现在可以把“对称性破缺导致分裂”这句话说得更正式。若高对称结构在 Γ 点具有二维 irrep，例如方格晶格的 E 或三角晶格的 E_1, E_2 ，那么该双重态的简并是受 little group 保护

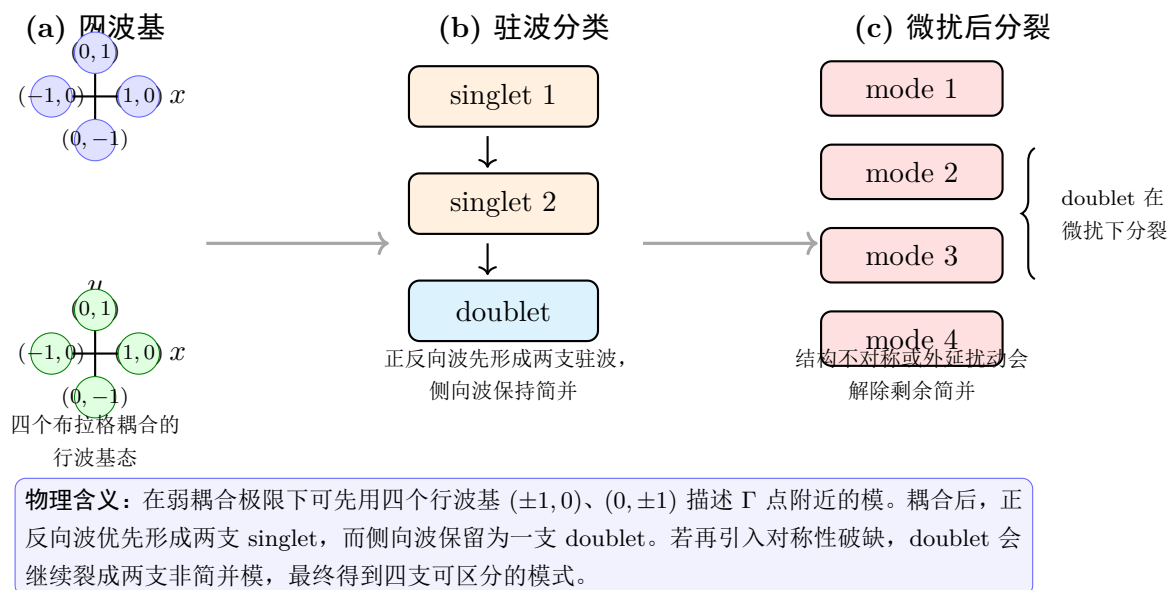


图 8.4: 四模三重态图像。(a) 弱耦合极限下的四波 ansatz: 正向行波 $(1,0)$ 、反向行波 $(-1,0)$ 和一对侧向行波 $(0, \pm 1)$; (b) 随着耦合增强，正反向行波混合形成驻波，按对称性分类为两支 singlet (偶宇称/奇宇称) 和一支 doublet (侧向简并对); (c) 当存在对称性破缺微扰时，doublet 进一步分裂，最终得到四支完全分离的非简并模。

的。一旦几何扰动把对称性从高群降到其子群，二维 irrep 通常会分裂成两个一维 irrep，于是：

- 原本双简并的模式分成两支非简并模；
- 两支模往往分别锁定到扰动的主轴方向，因此偏振被钉扎；
- 若这两支模与法向辐射通道的表示匹配程度不同，损耗分裂就会随之出现。

最典型的例子是方格晶格圆孔的 C_{4v} 对称被轻微椭圆孔降低到只保留两条镜面的子群。此时原先的 E 双重态会分裂成两支沿椭圆长轴和短轴定向的 singlet，于是“偏振选择”在群论上不再是一句经验话，而是“二维 irrep 被降群后发生分裂”的直接结果。

若一个理想方格晶格在 Γ 点有一对属于 E 的双简并候选模，则在圆孔情况下两者地位完全等价。把圆孔轻微拉成椭圆后，这对 E 双重态会沿扰动主轴方向分裂成两支非简并模。若其中一支仍与法向辐射通道更匹配，另一支则更接近 symmetry-protected dark family，那么实验上就会同时看到偏振钉扎、损耗分裂和远场主瓣方向选择。

例子

实测数量级锚点。Vafadar 与 Zhao 在 367 nm GaN 纳米线光子晶体面发射激光器中，用检偏器测得主轴方向强度约为正交方向的 10 倍。按

$$\rho = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$$

换算得到偏振度 $\rho \approx 0.8$ [68]。这个例子说明“偏振被锁定”应落到带测量协议的强度比或偏振度上；它也不意味着所有材料平台、孔形和工作点都应达到同一数值。

8.11 拿到一个新结构后，如何自己做 Γ 点模标注

至此，可以把 Γ 点模标注整理为以下最小流程：

1. 先判断平面晶格类型，从而确定 Γ 点 little group：方格晶格取 C_{4v} ，三角晶格取 C_{6v} ；
2. 用生成元操作去看目标模如何变换，给它标上 irrep；
3. 再判断法向辐射通道在该 little group 下属于哪个 irrep：方格取 E ，三角取 E_1 ；
4. 若 mode irrep 与 radiation irrep 不匹配，则 exact Γ 点 direct radiation 被对称性禁止，优先检查它是否属于 symmetry-protected BIC family；
5. 若随后加入对称性破缺，再问原 irrep 在子群下怎样分裂，以及哪一支更容易获得目标偏振和远场。

这套流程不能替代全波仿真，但足以在拿到一个新单胞时，先完成“模标注、偏振判断、BIC 初筛”三项前置工作，而不是一开始就进行无约束参数扫描。

若某一偏振态在二维平面模型中显示出明显优势，但该优势依赖的实际上是上下完全对称的假设，那么一旦放到真实外延层栈中，这种结论可能翻转。偏振设计必须把纵向不对称也纳入模型。

8.12 哪些对称性破缺最常见

真实 PCSEL 中，对称性破缺几乎不可避免。常见来源包括：

- 孔形由圆变椭圆，或双晶格扰动；
- 外延层上下不对称，导致上下辐射通道不同；
- 电极窗口、接触层和刻蚀区域打破原始旋转对称；
- 工艺偏差引入局部尺寸不均匀。

这些破缺并不都意味着性能退化。许多 PCSEL 偏振设计会主动引入可控的对称性破缺，使原本简并的偏振态分离。有效的对称性破缺应具有可预期、可复现且方向明确的特征，而不应依赖随机工艺偏差来选择偏振。

8.13 外延不对称为什么会影响偏振和远场

对初学者来说，一个容易忽视的事实是：偏振和远场不只由平面内晶格决定，纵向层栈同样会参与。原因有两点。

第一，真实模式是全矢量三维模式。上下包层不同、掺杂不同、接触层不同，会改变不同场分量在纵向上的分布，从而间接影响偏振选择。

第二，上下辐射通道不对称会改变远场的上下功率分布，进而影响总辐射损耗和远场图样。某一平面内看似很“对称”的模式，在真实层栈中可能已经因为上下不对称而发生偏置。

因此，后文讨论外延层和热效应时仍需回到偏振问题。偏振是贯穿光学与器件设计的变量，而非在本章即可封闭处理的独立主题。

将圆孔改成轻微椭圆，往往会使两个正交偏振态的耦合系数分离，进而在远场上表现为主瓣沿某一晶格方向更加集中。若此时外延上下不对称又偏向其中一支辐射通道，那么最终实验上看到的偏振纯度会比二维平面对称模型预期的更强，或者方向发生额外旋转。图8.5总结了三种典型的偏振调控策略及其物理机制。

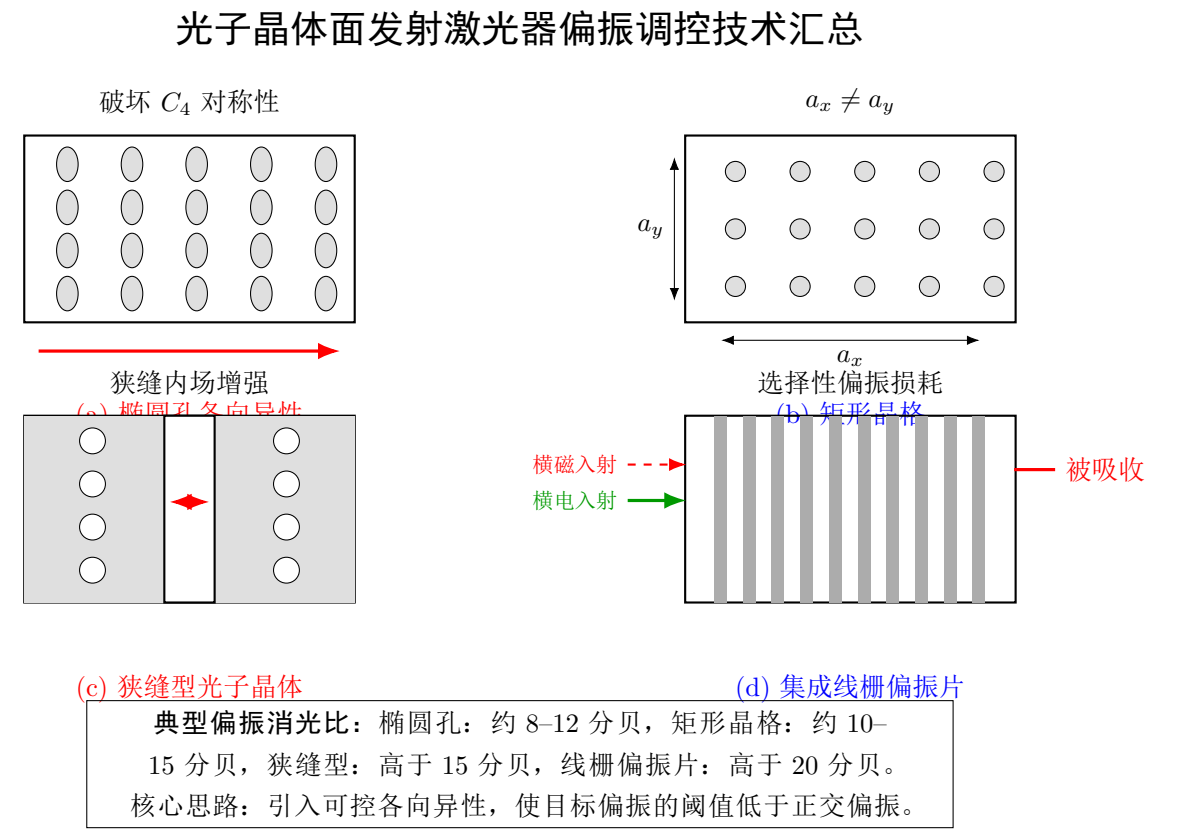


图 8.5: 光子晶体面发射激光器中四类常用偏振调控方法对比。

8.14 远场整形不是“把图样画漂亮”，而是组织角谱能量

远场整形不只是改变输出图样的外观；更准确地说，它是在重新组织角谱能量分布。

根据式 (8.8)，若想得到单主瓣、环形、双瓣或某种方向性增强的远场，就必须在近场里安排好对应的相位和振幅结构。因此：

- 改变晶格对称性，常常是在改变允许的角谱分量组合；
- 改变孔形扰动，常常是在改变不同角谱分量的相对强度；
- 改变外延和工作点，则可能在不改几何的情况下，通过模式重分布改写远场。

远场设计还需要检验该图样在工作电流和工作温度范围内的稳定性，而不能只依据冷腔仿真图样。

8.15 为什么偏振和远场必须放在同一章讨论

到这里，读者应该已经能体会：偏振和远场虽然在实验中常被当作两个指标测量，但在 PCSEL 理论里它们共享同一组决定因素：

- 模式的矢量近场结构；
- 晶格和器件的对称性；
- 模式与辐射通道的耦合方式；
- 工作点引起的模式重分布。

因此，本章合并讨论偏振和远场，是因为二者对应同一模式问题的两个投影：偏振强调矢量方向，远场强调角域分布。

只在二维平面模型里判定偏振纯度，或者只在冷腔近场上判断远场稳定性，都会系统性高估真实器件性能。只要外延不对称、电流拥挤或热漂移显著，这种高估几乎必然出现。

8.16 通向后续章节：偏振问题为何最终会回到器件问题

本章讨论的仍然主要是光学层。但请注意，偏振稳定和远场稳定并不只由几何决定。随着注入升高和温升出现：

- 折射率分布变化会引起模式重分布；
- 有源区增益谱漂移会改变两个偏振模的相对净增益；
- 局部热热点和电流非均匀会使原本对称的模式结构偏斜。

因此，本章给出的是偏振与远场的光学框架，而不是器件级稳定性判据。某种偏振与远场能否在工作区间内保持稳定，还需在热-电-光耦合章节中判断。

回看本章

PCSEL 的偏振与远场是模式本征结构、对称性和辐射选择规则的结果，而非输出端的附加现象。可操作的起点是在 Γ 点按 little group 的不可约表示分类模式：方格晶格用 C_{4v} ，三角晶格用 C_{6v} ；法向辐射通道分别最小地属于 E 和 E_1 。建立这一标注后，偏振选择、symmetry-protected BIC 判断、双简并分裂和远场工程可归入同一分类工具。对称性破缺既可作为设计手段，也可能成为不可控扰动。偏振与远场工程必须同时考虑平面内晶格、纵向外延、有限尺寸以及工作点变化。

练习题

1. [基础] 对理想方格晶格，说明为什么 Γ 点 little group 等于 C_{4v} ，并解释二维 irrep E 为什么天然与双偏振候选模有关。
2. [基础] 试解释为什么一个近场奇对称模式往往在远场中心更容易出现零点或弱化。
3. [进阶] 对三角晶格，说明为什么法向自由空间通道最小地属于 E_1 ，并据此解释 E_2 family 为什么常出现在 symmetry-protected BIC 讨论中。
4. [进阶] 给出两个“偏振选择由损耗分裂主导而非频率分裂主导”的可能场景。
5. [进阶] 讨论为什么上下层栈不对称会影响远场主瓣和偏振，而不仅仅影响总输出功率。
6. [综合]（思路提示：先定层级，再列关键式）设某器件在低电流下单偏振稳定、高电流下出现偏振混合，试给出一条可能的热-电-光机理链。

延伸阅读

- 下一章将引入 BIC / quasi-BIC 语言，把本章的 little group、irrep、法向辐射选择规则进一步写成开放系统中的辐射通道问题。
- [57, 28] 中关于对称性、表示与选择规则的章节，可作为本章表示论语言的背景材料。
- 配合 第十七章、第二十章 阅读本章，可检查二维平面模型、纵向层栈和热-电-光回写之间的接口。

第九章 BIC / quasi-BIC、 Γ 点辐射与远场工程

读完本章，你能

- 理解连续谱中的束缚态 (bound state in the continuum, BIC) 与准 BIC 在 PCSEL 中为什么重要。
- 学会把 Γ 点、辐射损耗、偏振、远场工程和 BIC 语言明确接起来，而不是把它们当成彼此独立的主题。
- 学会把 第八章 中的 little group / 不可约表示分类直接用来判断 exact Γ 点的法向辐射允许性。
- 明确 BIC / quasi-BIC 在 PCSEL 中是候选模与辐射工程语言，不是绕过器件物理的捷径。

读前准备 建议已掌握 第三章、第五章和第八章 中关于开放系统、 Γ 点模、辐射损耗和偏振/远场的内容。

阅读主线 本章所处模型层级：L1→L2

PCSEL 的目标模通常位于 Γ 点附近；设计上既需要较低辐射损耗以降低阈值增益，也需要保留可控的法向出射以形成可用远场。本章用 BIC / quasi-BIC 语言统一描述这两个要求。先定义开放周期系统中的 BIC，再把 第八章 的 little group / 不可约表示工具用于法向辐射通道，说明 exact Γ 点某些 irrep 的 direct radiation 为何可被对称性禁止。随后讨论小对称性破缺如何打开 quasi-BIC 辐射，并说明真实 PCSEL 通常工作在 quasi-BIC 设计区，而不是无限大、完美无损的理想 BIC 点。

9.1 为什么 BIC 不是旁枝话题，而是 PCSEL 主线中的辐射语言

PCSEL 是开放系统，因此需要同时描述本征模及其向连续辐射通道的泄漏。BIC 描述的是一种频率落在辐射连续谱内、却因干涉或对称性而不向外泄漏的极限状态。也就是说，BIC 是在辐射区内使净辐射振幅为零，而非频率离开辐射区。

这里说的“辐射连续谱”是 slab 外自由空间中、位于光锥以内的一整族可传播平面波模，而非抽象的谱论术语。对 PCSEL 读者来说，把它先理解成“所有允许向上、向下逃逸的远场平面波通道的连续集合”就足够了。

这一区分对 PCSEL 很重要：设计目标是判断哪些 Γ 点候选模因对称性或干涉而辐射受限，哪些模更容易向法向出光，而非单纯寻找高 Q 模。BIC / quasi-BIC 语言把高 Q 、偏振、法向远场和对称性归入同一辐射机制。

直觉框

对 PCSEL 来说，BIC 是描述辐射抑制及其受控打开的极限参照。实际设计通常不工作在理想 BIC 点本身，而是在其附近的 quasi-BIC 区域选择辐射损耗与输出耦合的折中。

例子

历史数量级锚点。Kodigala 等人在 2017 年用悬空圆柱纳米谐振器阵列实现了室温光泵 BIC 激光，激射位于约 $1.55\ \mu\text{m}$ ，阵列缩小到 8×8 单元后仍可起振 [31]。在理想无限周期 BIC 点，辐射衰减率趋于零，因而辐射 Q_{rad} 在数学上趋于无穷；实验有限阵列的总 Q 必然有限。把波长与这两个不同的 Q 口径同时写出，可以避免把“理想 BIC 的发散 Q ”误当成实测腔品质因数。

9.2 从辐射振幅到 BIC：什么叫“连续谱中的束缚态”

在进入公式之前，先把 BIC、普通束缚态和准 BIC 的边界一次说清。普通束缚态的“不辐射”来自频率本身不在允许传播的连续谱里；BIC 的“不辐射”则更微妙：它的频率已经落在连续谱内，按频率和动量本应可以向外辐射，但所有可用辐射通道的净耦合振幅恰好为零。准 BIC 则是理想 BIC 在有限尺寸、轻微几何扰动、上下层栈不对称或工艺误差作用下打开一条很弱辐射通道后的现实版本。

把某个 Γ 点附近候选模与自由空间辐射通道之间的耦合写成一个最小时间耦合模模型。设腔内模振幅为 a ，上方和下方辐射通道的出射振幅记为 $s_{\uparrow,-}$ 与 $s_{\downarrow,-}$ ，则可写

$$\dot{a} = (i\omega_0 - \gamma_{\text{nr}} - \gamma_{\text{rad}}) a, \quad \begin{bmatrix} s_{\uparrow,-} \\ s_{\downarrow,-} \end{bmatrix} = \begin{bmatrix} d_{\uparrow} \\ d_{\downarrow} \end{bmatrix} a, \quad (9.1)$$

其中 γ_{nr} 表示非辐射损耗， γ_{rad} 表示辐射损耗， $d_{\uparrow}, d_{\downarrow}$ 是模到辐射通道的耦合振幅。若在某个对称性或干涉条件下，

$$d_{\uparrow} = d_{\downarrow} = 0, \quad (9.2)$$

则该模即使频率位于辐射连续谱中，也不会向外辐射；理想无损情况下， $\gamma_{\text{rad}} = 0$ ，这就是 BIC 的最小数学图像。

需要特别强调的是，式 (9.2) 表示辐射振幅为零，而不是模式根本不与外界有任何关系。只要结构轻微破坏对称性、有限尺寸边缘打开旁路通道、表面粗糙引入散射，或者工作点变化导致场型重分布，这个零就会被抬起，理想 BIC 便退化成具有有限辐射线宽的 quasi-BIC。

BIC 的定义依赖开放边界与连续辐射通道。若把问题错误地近似成封闭本征值问题，就既不会看到真正的辐射损耗，也不会正确看到 BIC 的成因。

用干涉理解 BIC。 BIC 并不要求辐射通道不存在，而是要求通道中的净辐射振幅相消为零。对称性保护型 BIC 中，这种相消由结构对称性强制给出，而不是数值上的偶然抵消。

例子

BIC 的严格口径是什么？

普通束缚态可概括为”频率不在连续谱内，所以无辐射”；BIC 则应表述为”频率位于连续谱内，但辐射耦合矩阵元为零，所以无辐射”。对 PCSEL 来说，这个区分非常关键，因为 BIC / quasi-BIC 讨论的是开放辐射问题本身，而不是简单靠频率躲开辐射区的模式。

9.3 对称性保护 BIC 的物理图像：从偶极子到点群表示

第 8 章已经从点群不可约表示角度给出了分类规则。本节用一个偶极子辐射例子说明，对称性为何能够使辐射矩阵元为零。

考虑局域源的多极展开，最低阶项为电偶极辐射。电偶极矩定义为

$$\mathbf{d} = \int_V \mathbf{P}(\mathbf{r}) dV, \quad (9.3)$$

其中 $\mathbf{P}$ 是极化强度分布。关键在于： $\mathbf{d}$ 在空间反演操作 $\mathcal{I}: \mathbf{r} \rightarrow -\mathbf{r}$ 下如何变换？

若模式电场满足奇宇称条件 $\mathbf{E}(-\mathbf{r}) = -\mathbf{E}(\mathbf{r})$ ，则极化分布也具有相同宇称 $\mathbf{P}(-\mathbf{r}) = -\mathbf{P}(\mathbf{r})$ 。此时对整个单胞积分：

$$\mathbf{d} = \int_{\text{cell}} \mathbf{P}(\mathbf{r}) dV = \int_{\text{cell}} \mathbf{P}(-\mathbf{r}') dV' = - \int_{\text{cell}} \mathbf{P}(\mathbf{r}') dV' = -\mathbf{d}, \quad (9.4)$$

其中用了变量代换 $\mathbf{r}' = -\mathbf{r}$ 且积分区域不变（周期单胞关于原点对称）。于是得到 $\mathbf{d} = -\mathbf{d}$ ，唯一可能是 $\mathbf{d} = 0$ 。电偶极矩为零意味着最低阶辐射通道被关闭。

从偶极子禁戒到点群 irrep

上面的论证用的是最简单的空间反演对称性（宇称）。对 PCSEL 常用的正方晶格 C_{4v} 点群，情况更丰富：

- A_1, A_2 表示是一维的，场分布在某些操作下保持不变或变号
- B_1, B_2 也是一维，但变换规则不同
- E 表示是二维的，对应简并对

法向辐射的自由空间平面波构成 C_{4v} 的某个表示。若模式的 irrep 与辐射场的 irrep 不匹配（直积不含恒等表示），则耦合矩阵元必为零。这就是第 8 章中表格的物理来源。

将该论证推广到 C_{4v} 点群：exact Γ 点模式可按 C_{4v} 不可约表示分类，法向辐射通道（沿 z 轴传播的平面波）也有确定的变换行为。若模式 irrep 与辐射通道 irrep 不相容，则由群论正交性可知其重叠积分为零。因此，某些 Γ 点模式可成为 BIC；其原因是对称性选择定则，而非数值巧合。

9.4 Γ 点与 BIC 语言的对应关系

在二维周期 slab 中， Γ 点对应零平面波矢，恰好也是法向出射和法向辐射最直接相关的位置。对 PCSEL 来说，这种联系尤为自然：目标模式大多围绕 Γ 点组织，因为它最容易同时连接面内布拉格反馈与法向面发射。

而在 Γ 点，晶格对称性和场分量奇偶性经常会直接决定某一模是否能够与法向辐射通道耦合。若一个模式的近场对称性与外部法向平面波的表示不匹配，那么即使频率处于辐射连续谱中，它也可以因对称性而成为 symmetry-protected BIC。反过来，若通过轻微单胞扰动、双晶格设计或椭圆孔设计打破这类严格匹配条件，就会把原本封闭的辐射通道重新打开，得到 quasi-BIC。

因此， Γ 点、模简并、偏振选择和法向远场不是彼此独立的主题；在 Γ 点，它们由同一组对称性与辐射通道规则共同约束。

9.5 把法向辐射通道写成 irrep：BIC 判断才真正可操作

第八章已给出最小分类工具：方格晶格的 Γ 点 little group 为 C_{4v} ，三角晶格为 C_{6v} 。BIC 初筛还需要补充一条规则：法向自由空间辐射通道也必须按同一 little group 分类。

对 exact normal incidence，外部自由空间只提供两支横向平面波偏振，因此：

- 在 C_{4v} 下，法向辐射通道最小地张成二维表示 E ；
- 在 C_{6v} 下，法向辐射通道最小地张成二维表示 E_1 。

于是，若某个 Γ 点模的 irrep 与这些辐射通道表示不匹配，那么 direct normal-radiation matrix element 在保持该对称性的前提下就必须为零。这就是 symmetry-protected BIC 最简洁、也最适合教学复用的判断规则。

表 9.1: exact Γ 点法向辐射允许性的一阶 irrep 判断规则。这里给出的是最小 in-plane little-group 规则；更严格的三维 slab 判断还需再结合纵向奇偶性与层栈对称。

晶格	Γ 点 little group	法向辐射通道 irrep	最小结论
方格晶格	C_{4v}	E	属于 E 的模式在对称性上允许 direct normal radiation；属于 A_1, A_2, B_1, B_2 的模式可形成 symmetry-protected dark family
三角晶格	C_{6v}	E_1	属于 E_1 的模式在对称性上允许 direct normal radiation；属于 A_1, A_2, B_1, B_2, E_2 的模式在 exact Γ 点可成为 symmetry-protected dark family

这张表把“看场图判断是否辐射”的经验判断改写为可执行流程：先确定 little group，再标注 irrep，最后与辐射通道 irrep 比较。若不匹配，可优先把该模视为 symmetry-protected BIC family 候选；若匹配，再计算耦合强度。

易错点

in-plane little-group 匹配只是必要的第一层筛选，不是三维 slab 中法向辐射强度的充分判据。也就是说，irrep 匹配说明 direct radiation 在对称性上“可以非零”，并不保证它一定很强；而真正的法向辐射强度还会继续受纵向奇偶性、层栈不对称、有限厚度矢量混合和有限尺寸边缘共同影响。

若想把“为什么不匹配就一定不辐射”再往前推进半步，可以用最小偶极矩积分来理解。对 exact Γ 点法向直接辐射，主导耦合常可压缩为单胞内某个零级辐射矩

$$\mathbf{d}_0 \propto \int_{V_{\text{cell}}} \mathbf{P}_{\parallel}(\mathbf{r}) dV, \quad (9.5)$$

其中 $\mathbf{P}_{\parallel}$ 是与法向辐射通道耦合的平面内极化或等效辐射源分量。若某个对称操作 g 保持单胞与法向辐射通道不变，而模式在该操作下属于特征值 $\chi_{\alpha}(g) \neq 1$ 的 irrep，则

$$\mathbf{d}_0 = \int \mathbf{P}_{\parallel} dV = \int \hat{T}(g) \mathbf{P}_{\parallel} dV = \chi_{\alpha}(g) \mathbf{d}_0.$$

于是只有 $\mathbf{d}_0 = \mathbf{0}$ 才可能满足上式。这给出 symmetry-protected BIC 的最小推导：该模并非只是场图较暗，而是与法向均匀辐射通道耦合的零级积分在对称性上为零。

这里还应把“辐射通道的表示”说得更正式一点。完整的法向辐射二维偏振子空间在方格晶格下按 E 表示变换；但当我们把耦合振幅进一步投影到某一个固定的法向线偏振端口，或等价地研究某个固定的零级辐射矩分量时，对应的标量重叠泛函在保持该端口不变的对称子群下就是平凡表示。此时若模式属于非平凡 irrep，群平均投影

$$\Pi_{\text{rad}} = \frac{1}{|G|} \sum_{g \in G} \chi_{\text{rad}}^*(g) \hat{T}(g) \quad (9.6)$$

便会把该重叠压成零。上文用单个群元写出的 $\chi_{\alpha}(g) \neq 1 \Rightarrow \mathbf{d}_0 = \mathbf{0}$ ，只是这个更一般投影结论的最小判据：当存在某个保持辐射端口不变的群元，而模式在其下取得不同字符时，禁耦合就已经成立。

若某个方格晶格 slab 在理想上下对称、单胞高对称的条件下支持一个关于镜面对称为奇的 Γ 点模，而法向自由空间平面波关于同一镜面为偶，那么该模对法向辐射的耦合振幅便可严格为零。此时它在无限大、理想结构中表现为 symmetry-protected BIC。一旦将圆孔轻微拉成椭圆，或引入双晶格扰动，这个严格零耦合被打破，模式会转化为 quasi-BIC，并出现有限但可控的法向辐射。

9.6 从 BIC 到 quasi-BIC：为什么真实 PCSEL 更关心后者

若只追求无穷大的被动 Q ，理想 BIC 可以作为极限点；但 PCSEL 作为激光器还需要可控的垂直辐射，用于输出和远场整形。因此，真实器件通常选择靠近 BIC 的 quasi-BIC 区域，在低辐射损耗和有限输出耦合之间折中。

图9.1给出了 BIC 形成机制的完整可视化：(a) 对称性保护 BIC，当模式的不可约表示与辐射通道的表示不匹配时，耦合矩阵元严格为零；(b) Friedrich-Wintgen 型 BIC，两个共振模通过共享辐射连续谱发生干涉相消；(c) 准 BIC 区域，轻微对称性破缺将理想 BIC 转化为具有超高但有限 Q 值的共振态。

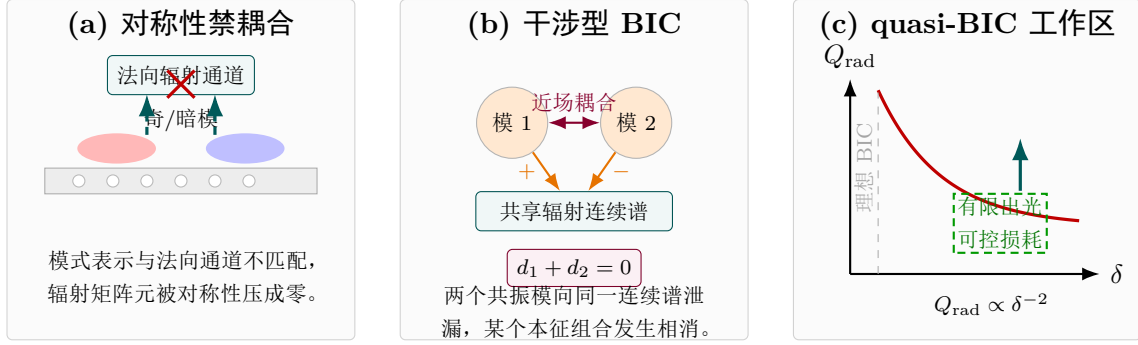


图 9.1: BIC 形成机制示意图。(a) 对称保护 BIC；(b) 干涉型 BIC；(c) quasi-BIC 区域及其 $Q \propto \delta^{-2}$ 标度。

图9.1概括了 BIC 的两类形成机制：对称性禁耦合和共用连续谱中的干涉相消。对 PCSEL 设计而言，问题是选择 quasi-BIC 区域，使辐射 Q 、增益、阈值和远场需求相互匹配，而非简单追求 BIC。

在最简单的对称破缺参数 δ 下，辐射振幅常可近似线性展开为

$$d \propto \delta, \quad \gamma_{\text{rad}} \propto |d|^2 \propto \delta^2, \quad Q_{\text{rad}} \propto \delta^{-2}. \quad (9.7)$$

$Q \propto \delta^{-2}$ 的量纲分析与物理论证

设对称性破缺参数为 δ ，辐射振幅 d 是最小的线性响应量，则 Taylor 展开给出 $d = d_0 + d_1\delta + O(\delta^2)$ 。由于 $\delta = 0$ 时对应理想 BIC ($d_0 = 0$)，故主导项为 $d \approx d_1\delta$ 。辐射功率正比于 $|d|^2$ ，因而 $\gamma_{\text{rad}} \propto \delta^2$ ，再由 $Q = \omega_0/2\gamma_{\text{rad}}$ 即得 $Q \propto \delta^{-2}$ 。

从量纲角度看， δ 是无量纲几何扰动（如位移除以晶格常数）， Q 也是无量纲量，因此幂次关系只能是纯数。二次方的根源在于：辐射是偶极矩（或更高阶多极矩）的平方，而小扰动下多极矩的变化是线性的。

式 (9.7) 往往还可进一步写成更适合工程估算的形式

$$Q_{\text{rad}} \approx \frac{Q_0}{\delta^2}, \quad \delta = \frac{d}{a}, \quad (9.8)$$

其中 d 是破缺位移或等效几何扰动幅度， a 是晶格常数， Q_0 吸收了结构 family、工作带支和归一化定义相关的常数。式 (9.8) 的直接后果是：若其余条件近似不变，把 δ 从 0.02 降到 0.01，辐射 Q 通常会提升约四倍。该二次标度在对称破缺 quasi-BIC 文献中有明确支撑 [33]。它说明小几何扰动通常以连续方式打开辐射通道。设计问题因此转化为：在给定输出功率、阈值增益预算、偏振纯度和远场要求下选择 δ 的工作区间，而不是机械追求无限接近 BIC。

这也构成 quasi-BIC 工作区与“最优单点设计”的差别。若把 δ 压得过小，也许被动 Q 极高，但法向提取太弱、有限尺寸边缘和工艺扰动又会变得更敏感；若把 δ 放得过大，则目标

模虽然更易出光，却可能因辐射损耗和竞争模重新排序而抬高阈值。PCSEL 设计更关心器件约束下可容差的 quasi-BIC 工作区，而不是数学上的极限点。

9.7 前沿补充：quasi-BFIC 与宽角高 Q 设计

除“单点附近的 quasi-BIC”外，近年还出现了 quasi-bound flat bands in the continuum (quasi-BFIC) 路线：其目标是在更宽的角度/动量区间维持较高 Q 与可用辐射耦合，而非只在单一动量点获得高 Q 。对 PCSEL 设计的意义是两点：第一，工艺与角度容差可提高；第二，有限尺寸阵列中的角谱展宽不再必然迅速拉低整体辐射品质。该结论在 2025 年直接工作中已有证据支撑 [54]。但 quasi-BFIC 是补充：前者更强调宽角鲁棒，后者更强调中心点高选择性，而非“替代”传统 Γ 点 quasi-BIC。工程上应根据目标（窄主瓣极限、容差、功率提取）选择路线；BIC/quasi-BIC 综述与标度工作更适合作为背景理论锚点 [23, 33]。进一步说，quasi-BFIC 的“宽角高 Q ”结论本身仍是开放周期或大尺寸近似下的结论：一旦转入有限阵列，边缘散射与角谱离散化会显著改写可用窗口。因此，任何“宽角鲁棒”结论都必须补交有限尺寸证据（阵列尺寸、边界模型、收敛曲线与误差样本），否则只能算机理提示而非器件级判据 [52]。

报告规范

跨平台外推三元注释

- 外推来源：BIC/quasi-BIC 理论锚点与 quasi-BFIC 直接证据联合使用 [23, 33, 54]
- 边界条件：需要在同一器件目标下同时给出角谱范围、有限尺寸边界、远场与模排序收敛
- 不可迁移项：不可把“宽角高 Q ”直接等价为“器件阈值更低或更鲁棒”；需经过有限阵列与电热回写验证

9.8 进阶：far-field polarization vortex 与拓扑电荷语言

前文已经建立 BIC 的最小对称性与干涉描述。现代 BIC 理论还常使用动量空间远场偏振矢量场的奇点来描述 BIC，即 far-field polarization vortex。

为说明这种偏振绕转，把某一支 slab 共振模在上方辐射半空间的远场振幅写成二维复向量

$$\mathbf{c}(\mathbf{k}_{\parallel}) = \begin{bmatrix} c_x(\mathbf{k}_{\parallel}) \\ c_y(\mathbf{k}_{\parallel}) \end{bmatrix}, \quad (9.9)$$

其中 $\mathbf{k}_{\parallel} = (k_x, k_y)$ 是平面内波矢， c_x, c_y 表示该模向两支法向线偏振通道的辐射振幅。BIC 点的最小判据就是

$$\mathbf{c}(\mathbf{k}_{\parallel, \text{BIC}}) = \mathbf{0}. \quad (9.10)$$

也就是说，BIC 是远场矢量场中的零点。

一旦离开这个零点，远场会重新出现，并携带某个主偏振方向。对近线偏振远场，主轴角可先写成

$$\phi(\mathbf{k}_{\parallel}) = \frac{1}{2} \arg[c_x^2(\mathbf{k}_{\parallel}) + c_y^2(\mathbf{k}_{\parallel})], \quad (9.11)$$

它把式 (9.9) 中的复振幅 c_x, c_y 直接转成偏振椭圆长轴偏角。于是，若沿着包围 BIC 的闭合回路 C 绕一圈，主偏振方向角 $\phi(\mathbf{k}_{\parallel})$ 发生整数倍 2π 的绕转，则可定义拓扑电荷

$$q = \frac{1}{2\pi} \oint_C \nabla_{\mathbf{k}_{\parallel}} \phi \cdot d\mathbf{k}_{\parallel}. \quad (9.12)$$

这个整数 q 就是 far-field polarization vortex 的 winding number。它把“这个 BIC 为什么不是数值偶然出现的零点”转化成了一个拓扑语言：只要相关带支与远场矢量场连续变化，这个零点不能凭空消失，而必须通过带着总电荷守恒地移动、成对产生或成对湮灭的方式演化 [79, 23, 73]。

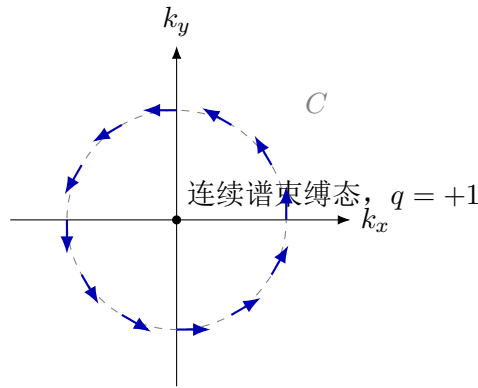


图 9.2: k 空间中远场偏振涡旋的最小示意。闭合回路 C 围绕连续谱束缚态奇点一周时，远场主偏振方向同步绕转一周，对应拓扑电荷 $q = +1$ 。

拓扑电荷保护的对象，是动量空间远场矢量场中的奇点结构，不是器件性能的一切性质。也就是说，电荷守恒可以帮助我们追踪 BIC 在 $\mathbf{k}_{\parallel}$ 空间和参数空间中的迁移、生成与湮灭，但它并不自动保证最低阈值、最佳热稳定性或最佳注入均匀性。对 PCSEL 而言，拓扑语言是用来追踪候选模辐射家族的，不是用来跳过器件模型的。

这套语言对 Γ 点附近模式尤其有用。对称性保护型 Γ 点 BIC 可以理解成被晶格对称性钉扎在区中心的偏振涡旋中心。当单胞扰动、双晶格参数或外延不对称连续变化时，读者不必只盯着“ Q 有没有升高”，还可以追问：

- 这个偏振涡旋中心是否仍停留在 Γ 点，还是已经移到 off- Γ ；
- 总拓扑电荷是否守恒，是否出现了成对移动的 BIC 奇点；
- 目标模式是仍属于原来的 quasi-BIC family，还是已经越过了亮模/暗模家族的分界。

现代 BIC 拓扑语言在这里的作用，是提供一种比单独比较谱线宽度更稳定的模追踪方法。特别在 Γ 点附近模式接近、偏振与远场同时变化时，far-field polarization vortex 与拓扑电荷可辅助判断模族是否连续演化。

设某个方格晶格 PCSEL 单胞在高对称参数处支持一个位于 Γ 点的 symmetry-protected BIC。随着双晶格扰动参数缓慢增大，直接法向辐射被逐步打开，目标模退化为 quasi-BIC。此时若只看单点 Q 值，读者可能只会得到“ Q 在下降”这一条信息；而若同时追踪 $c(\mathbf{k}_{\parallel})$ 的零点与绕转方向，就能看出该模究竟是在沿同一拓扑家族连续偏离 BIC，还是已经与邻近亮模发生重组。前者更适合解释为“可控打开辐射”，后者则意味着模式排序、偏振和远场可能一起改写。

9.9 对称性破缺在 irrep 语言下如何打开 quasi-BIC

使用 little group 工具后，可以更正式地表述“对称性破缺打开 quasi-BIC”。设高对称结构在 Γ 点支持某个 dark family，其 irrep 与法向辐射通道 irrep 不匹配，因此 exact Γ 点 direct radiation 为零。若小扰动把点群降低到某个子群，则原 irrep 会在子群下重新分解。常见结果有两类：

1. 原本的二维 irrep 分裂成两个 singlet，于是偏振被钉扎、频率与损耗同时分裂；
2. 原本与辐射通道不匹配的 dark family，在子群中与辐射通道落入同一表示，于是零耦合被抬起，BIC 退化为 quasi-BIC。

这一步把第八章的偏振选择/模式分裂与本章的 BIC 打开/关闭联系起来。二者都是降群问题的结果：前者关注本征频率与偏振，后者关注辐射矩阵元与远场。

9.10 BIC 如何连接偏振选择与远场工程

第八章已经说明，偏振与远场由近场矢量结构和辐射通道共同决定。用 BIC 语言表述，某一偏振能否辐射、沿何方向辐射以及远场中心亮暗，取决于该模与相应 BIC 条件的距离。

对称保护型 BIC 常与特定偏振表示绑定。更严格地说，在方格晶格里这常表现为某一支 E 双重态在降群后更强地获得与法向辐射通道的匹配，而另一支仍更接近 dark family；在三角晶格里则常表现为 E_1 和 E_2 family 在小扰动下偏离 BIC 条件的速度不同。于是，当设计者通过单胞形变打破旋转对称性时，实际上做的并不只是“让偏振分裂”，更是在改变不同偏振支路偏离 BIC 条件的速度。某一偏振支路若仍更接近辐射抑制条件，其辐射损耗会更低、阈值增益更小；另一偏振支路若更远离该条件，则更容易辐射、更容易在远场中形成主瓣，但被动损耗也更高。

从远场角度看，若某模严格处于 symmetry-protected BIC 点，它在法向中心常表现为零辐射；小扰动打破对称性后，法向中心会逐步亮起，远场主瓣和偏振纯度随之变化。因此，许多远场整形方法可表述为控制目标模式偏离 BIC 条件的方式。

把 BIC 理解成“理想 PCSEL 就应该完全不辐射”，是概念错误。激光器若完全不向目标通道辐射，就无法作为可用面发射器件工作。BIC 在 PCSEL 中的作用是帮助设计者理解辐射如何被抑制与重开，而不是让器件退化成只会储能的封闭态。

9.11 为什么 BIC 仍然不能替代阈值与器件模型

到这里必须再次把层级边界说清楚。BIC / quasi-BIC 语言擅长回答的问题是：在开放周期光学层面，辐射损耗为什么高或低、怎样通过对称性与干涉调控它、以及这如何影响偏振与远场。但它不能直接回答阈值电流、热稳定性和真实工作线宽。

原因很直接。首先，理想 BIC 讨论的是无限大、精确对称、无工艺误差的被动结构；真实器件里存在有限阵列边缘、粗糙、刻蚀误差、金属和注入非均匀。其次，即使某个模式具有非常小的辐射损耗，也未必与量子阱重叠最好，未必避开重掺杂损耗层，也未必在热回写后继续保持优势。最后，线宽和调制稳定性还涉及载流子、增益压缩和相位噪声，这已经超出了 BIC 自身的描述范围。

常见错误：无限周期到有限器件的外推

错误理解：“无限周期仿真中该模接近 BIC，高 Q ，所以有限器件必然高 Q 且低阈值。”

正确做法：把无限周期结果视为候选筛选，随后必须报告有限阵列尺寸、边界截断、粗糙模型，以及 Q 与远场随尺寸收敛曲线。**最小报告项：**至少给出阵列尺寸、边界模型、收敛曲线、误差样本四要素。

报告规范

无限周期结果转有限阵列时的最小报告四要素

1. **阵列尺寸：**至少报告 3 个横向尺寸样本，并明确主模/竞争模排序是否翻转。
2. **边界模型：**明确边缘截断方案、PML/端口设置与粗糙/无序代理模型。
3. **收敛曲线：**至少给出 1 组关键量（如 Q 、主瓣角、 ΔG_{*j} ）随尺寸或网格的收敛趋势。
4. **误差样本：**给出最坏样本或通过率统计，而不只给名义参数点。

因此，本章在全书中的位置是：把前文关于 Γ 点与偏振/远场的描述改写为开放系统辐射语言，并为后续 CWT、全波和器件章节提供候选模筛选工具，而不在本章给出器件级结论。

回看本章

BIC / quasi-BIC 为 PCSEL 提供了连接 Γ 点、辐射损耗、偏振和远场工程的开放系统语言。BIC 的关键是在连续谱中通过对称性或干涉使辐射振幅为零，而非频率离开辐射连续谱；在 Γ 点，这可通过比较模式 irrep 与法向辐射通道 irrep 来判断。若二者不匹配，则 direct radiation 在对称性上被禁止。quasi-BIC 则是在真实设计中通过小而可控的对称破缺重新打开辐射通道。PCSEL 通常关注理想 BIC 附近的工作区，以同时兼顾低辐射损耗、可用面发射、偏振选择和远场工程；BIC 是阈值与器件模型的前端语言，但不能替代后续器件分析。

练习题

1. [基础] 用自己的话解释为什么 BIC 的定义必须建立在开放边界与辐射连续谱上，而不能在封闭腔语言中直接照搬。
2. [基础] 结合式 (9.2) 说明“耦合到连续谱”与“向连续谱辐射”为何不是同一句话。
3. [进阶] 对方格晶格，说明为什么 exact Γ 点法向辐射通道属于 E ，并据此解释为什么一维 irrep family 更容易成为 symmetry-protected BIC 候选。
4. [进阶] 利用式 (9.7) 讨论为什么 quasi-BIC 设计常出现“越靠近理想 BIC，越需要警惕工艺敏感性”的权衡。
5. [综合]（思路提示：先定层级，再列关键式）选择一种你熟悉的单胞扰动方式，说明它更可能首先改写的是偏振分裂、法向远场，还是辐射损耗，并解释原因。

延伸阅读

- 下一章开始进入 CWT 语言。本章的 irrep 与辐射选择规则将在那里对应到耦合矩阵中的损耗项和模式分裂。
- BIC 的一般背景可参考 [13, 22, 23, 31]；far-field polarization vortex 与拓扑电荷可参考 [79, 73]。
- 阅读后续器件章节时，应检查 quasi-BIC 设计在有限尺寸、注入非均匀和热回写后是否仍保持目标模优势。

第十章 PCSEL 的耦合波理论 (CWT): 最小模型

读完本章，你能

- 理解为什么在能带图之后必须引入耦合波理论 (Coupled-Wave Theory, CWT)，以及它在 PCSEL 研究链条中的准确位置。
- 从 Maxwell 方程与场展开出发，建立最小 CWT 模型，明确每一步近似究竟丢掉了什么物理。
- 学会把四波 CWT 基底改写成 Γ 点 little-group 的 irrep 基底，从而把“耦合矩阵”直接连到“模标注、偏振与 BIC 判断”。
- 学会把耦合系数、辐射损耗、材料增益和模竞争放进同一组低维方程中阅读，而不是把它们拆成互不相干的名词。

读前准备 建议已掌握 第三章和五至七 中关于开放 Maxwell 问题、Gamma 点候选模、面内耦合与阈值条件的内容。

阅读主线 本章所处模型层级: L1→L2

PCSEL 的面发射来自周期散射将面内导模耦合到垂直辐射通道；阈值问题要求把增益和损耗写入同一动力学框架。CWT 提供介于能带图和全波仿真之间的低维模型。本章先说明能带图的不足，再从场展开和投影出发建立最小模型，解释矩阵中各类项的物理来源，并把四波基底改写成 Γ 点 irrep 基底，使模式标签、偏振选择和辐射选择规则进入同一矩阵语言。最后说明其在设计流程中的用法与边界。

记号约定 (避免符号歧义): 本章与导波相关的 β 默认都指 β_{slab} (slab 面内传播常数)；自发辐射耦合因子统一写作 β_{sp} ，仅在动态与噪声章节使用。时间因子约定与全书一致，均采用 $e^{-i\omega t}$ ，后文所有等效损耗、增益和复频率符号应与 第三章、第二十八章 的频域约定联读。

10.1 为什么在能带图之后还需要 CWT

对初学者来说，能带图往往已经很有吸引力。它告诉我们哪些 Gamma 点模值得关注，也告诉我们模之间可能存在简并、分裂与光锥耦合。但是，能带图主要回答的是“有哪些候选

模”，而不是“这些候选模为什么会互相竞争、又为什么某种几何扰动会特别有效”。后一类问题需要把注意力从“单个本征模的频率”转移到“若干主导平面波分量之间如何相互耦合”。

CWT 的作用正是在此处引入。它不替代 Maxwell 方程，也不否定全波仿真的必要性；它是在已知目标模主要面内波分量后，将这些分量投影为低维耦合系统。这样，几何参数的作用可通过耦合系数、损耗项和增益项逐项解释。

10.2 最小 CWT 要解决的物理问题是什么

在方格晶格 PCSEL 的最典型工作区间里，Gamma 点附近的目标模常常主要由四个等价的面内导模分量组成：沿 $+x$ 、 $-x$ 、 $+y$ 、 $-y$ 方向传播的基本波。上一章已经解释过，周期折射率调制通过倒格矢把这些分量连接起来，并进一步把其中的某些组合折叠进光锥，形成垂直辐射。因而，若只关心 Gamma 点附近最核心的反馈网络，一个自然的最小基底就是这四个波分量。

于是我们想做的事就变得很明确：

第一，把三维矢量 Maxwell 场写成“纵向模剖面”乘以“少数几个面内平面波”的组合；

第二，把周期结构导致的相互散射压缩为若干耦合系数；

第三，把辐射损耗、内部损耗和增益作为等效非厄米项写进同一套方程；

第四，通过求这个低维系统的本征值和本征向量，判断哪一类线性组合更可能成为实际候选激射模。

若这四步成立，所得矩阵是由 Maxwell 问题投影得到的物理压缩模型，而非任意经验矩阵。

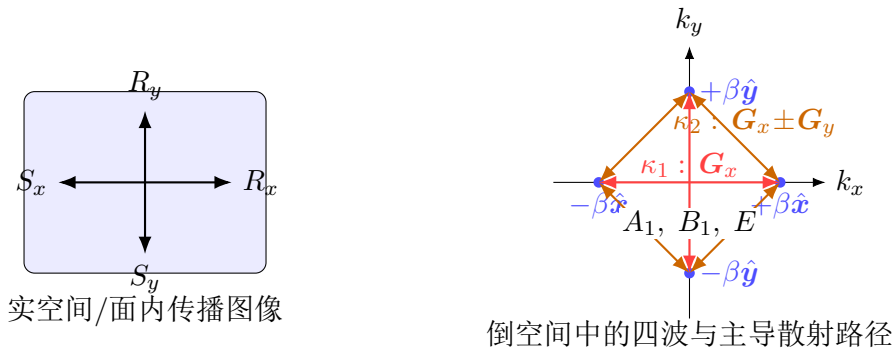


图 10.1: 最小四波 CWT 图像。左图强调四支基本波在面内传播；右图强调它们在倒空间中如何通过主导倒格矢散射彼此连接。对 PCSEL 而言，后续的偏振、BIC 与阈值排序都建立在这类最小闭合散射网络之上。

10.3 从 Maxwell 方程到场展开：四波近似从哪里来

为了不让推导显得像突然写下一个猜测矩阵，我们从背景波出发。设器件在没有横向周期调制时，已经支持一个主要的纵向导模剖面 $e_0(z)$ 。当加入二维周期折射率扰动后，我们

在 Gamma 点附近只保留四个最主要的面内波分量, 写成

$$\mathbf{E}(\rho, z, t) \approx e^{-i\omega t} \mathbf{e}_0(z) \sum_{m=1}^4 a_m(\rho) e^{i\beta_m \cdot \rho}, \quad (10.1)$$

其中 $\rho = (x, y)$, 四个传播向量取为

$$\beta_1 = (\beta, 0), \quad \beta_2 = (-\beta, 0), \quad \beta_3 = (0, \beta), \quad \beta_4 = (0, -\beta).$$

这里的 $a_m(\rho)$ 是缓慢变化的包络, 表示每个基本波分量在大尺度上的权重; $\mathbf{e}_0(z)$ 则把纵向层栈的信息压缩进一个主要导模剖面。之所以允许这种写法, 是因为我们关注的是 Gamma 点附近、频率接近且相互强耦合的一小族波, 而不是整个倒格空间中无穷多个分量。

接下来, 把介电常数写成背景项与周期扰动之和:

$$\varepsilon(\rho, z) = \varepsilon_0(z) + \Delta\varepsilon(\rho, z), \quad \Delta\varepsilon(\rho + \mathbf{R}, z) = \Delta\varepsilon(\rho, z). \quad (10.2)$$

再将 $\Delta\varepsilon$ 作二维傅里叶展开,

$$\Delta\varepsilon(\rho, z) = \sum_{\mathbf{G}} \Delta\varepsilon_{\mathbf{G}}(z) e^{i\mathbf{G} \cdot \rho}, \quad (10.3)$$

便得到不同 β_m 之间的散射通道。把式 (10.1) 和式 (10.3) 相乘时, 每一个基本波 $a_n e^{i\beta_n \cdot \rho}$ 都会生成一组

$$\Delta\varepsilon_{\mathbf{G}}(z) a_n(\rho) e^{i(\beta_n + \mathbf{G}) \cdot \rho}$$

项。于是, 投影到第 m 个基底后, 真正留下来的耦合必须满足

$$\beta_m \approx \beta_n + \mathbf{G} \quad \Longleftrightarrow \quad \mathbf{G} \approx \beta_m - \beta_n. \quad (10.4)$$

这就是 CWT 里“耦合矩阵元来自哪一个 Fourier 分量”的最小选择定则。只保留“近共振”的倒格矢耦合后, 四个包络之间便不再独立, 而是由少数几个主导傅里叶分量彼此连接。

10.4 投影步骤到底在做什么

形式上, 将式 (10.1) 代回频域 Maxwell 方程, 再用背景导模剖面 $\mathbf{e}_0(z)$ 和四个平面波基函数作为测试函数做加权投影。这个步骤等价于限定系统只在四个主导波分量张成的子空间中演化, 并求出 Maxwell 算符在该子空间内的矩阵表示。

进行投影时, 通常还要用到两个额外近似。

第一是慢变包络近似 (slowly varying envelope approximation)。它要求 $a_m(\rho)$ 相对载波相位因子变化得慢, 从而二阶横向导数和远离共振的高频混频项可以忽略。这一步使方程从“完整波动问题”降到“包络动力学问题”。若想把“变化得慢”说得更定量一些, 可把包络变化长度记为 L_{env} , 则最小要求是

$$L_{\text{env}} \gg a, \quad |\nabla_{\rho}^2 a_m| \ll \beta^2 |a_m|.$$

它的物理含义是: 一个晶格常数尺度上, 主要由快速 Bloch 相位承载振荡; 包络只在更大的器件尺度上记录边缘、增益和热斑导致的慢变化。对典型 PCSEL 而言, 若晶格常数 $a \approx 280 \text{ nm}$ 、

有限阵列直径 $D \approx 100 \mu\text{m}$ to $300 \mu\text{m}$, 则 $D/a \approx 360\text{--}1100$, 这也是最小 CWT 在器件中央区通常仍有教学价值的原因。也因此, 一旦进入边缘急剧截断、局部热斑或强非均匀注入区, 最小 CWT 的慢变包络假设就会先变得可疑。

第二是主导耦合截断。周期结构的傅里叶分量原则上可以把一个波分量耦合到许多其他分量, 但真正主导 Gamma 点物理的往往是少数几个满足近布拉格条件的通道。最小 CWT 正是主动舍弃那些次要通道, 以换取可解释性。

把这个“截断”说得更公式化一点: 若本征频率附近存在许多可能的平面波分量, 则只有同时满足下列两类条件的通道才值得进入最小模型:

1. 动量近匹配: 满足式 (10.4);

2. 失谐足够小: 对应未耦合导模频率差 $|\omega_m^{(0)} - \omega_n^{(0)}|$ 不大于主导耦合系数数量级。

对方格晶格、Gamma 点附近、二阶 Bragg 工作的最小教材模型来说, $\pm x$ 与 $\pm y$ 这四支波恰好同时满足这两类要求, 因此构成自然的四波闭合子空间; 更远的倒格矢通道则因频率失谐或纵向重叠太弱而被先行舍弃。这一截断的定量充分条件可粗略写成: 主导傅里叶分量幅度满足 $|\Delta\epsilon_{\mathbf{G}}|/\bar{\epsilon} \ll 1$ (折射率对比度足够小), 同时次近邻布拉格通道的失谐远大于主导耦合系数 $|\kappa|$ 。若这两个条件不满足, 则应升级到更多波分量的展开或直接转入 RCWA/FDTD 全波方法。

完成这些步骤后, 就得到线性耦合方程

$$\frac{d}{dt}\mathbf{a} = (\mathbf{i}\Omega - \mathbf{\Gamma} + \mathbf{G})\mathbf{a}, \quad \mathbf{a} = \begin{bmatrix} R_x & S_x & R_y & S_y \end{bmatrix}^T, \quad (10.5)$$

其中 R_x, S_x, R_y, S_y 分别表示四个基本波分量的包络振幅。式 (10.5) 看起来像一个普通矩阵方程, 但每个矩阵块背后都有清楚的来源和职责, 下面要逐项解释。

若进一步假设包络按单一复频率偏移演化,

$$\mathbf{a}(t) = \mathbf{u} e^{-i\delta\omega t}, \quad (10.6)$$

则在本书统一采用的时间因子 $e^{-i\omega t}$ 约定下, 时间导数满足

$$\frac{d}{dt} \rightarrow -i\delta\omega.$$

把式 (10.6) 代回式 (10.5), 便得到

$$[-i\delta\omega \mathbf{I} - (\mathbf{i}\Omega - \mathbf{\Gamma} + \mathbf{G})]\mathbf{u} = 0. \quad (10.7)$$

式 (10.7) 给出了从时间演化方程到频域本征值问题的最小转换。后文高阶或三维 CWT 中写成 $(\delta + i\alpha)\mathbf{V} = \mathbf{C}\mathbf{V}$, 只是对这一步进行不同记号重排与归一化, 并非另一套独立理论。

若把投影步骤写得再直白一点, 它其实就是以下三步。

1. 先选定低维基底 $\{\mathbf{e}_0(z)e^{i\beta_m \cdot \rho}\}$;

2. 再把完整 Maxwell 算符夹在这组基底左右, 形成矩阵元

$$M_{mn} \sim \langle \mathbf{e}_0 e^{i\beta_m \cdot \rho}, \hat{\mathcal{L}} \mathbf{e}_0 e^{i\beta_n \cdot \rho} \rangle;$$

3. 最后只保留满足式 (10.4) 的主导矩阵元, 把其余远离共振的通道当作高阶修正。

因此, CWT 是把完整 Maxwell 问题投影到预先选定且物理上主导的有限维共振子空间, 而非猜测一个 4×4 矩阵。

严格表述 (难度)

严格地说, 式 (10.5) 是从完整问题向低维共振子空间做投影之后的近似闭合系统, 而非从完整 Maxwell 方程“精确等价变形”而来。它的可信度取决于两件事: 一是被丢弃的基底是否真的次要, 二是工作点是否确实靠近所选共振子空间。脱离这两个条件, CWT 只能作为启发图像, 而不能作为定量工具。

图10.2给出了 CWT 信号流的完整可视化: 四个行波基底通过倒格矢散射形成闭合耦合网络, 其中 κ_1 描述反向散射 (1D DFB 型), κ_2 描述 90° 转角散射 (2D PCSEL 特有), 辐射损耗通过光锥耦合项体现, 材料增益作为对角项作用于所有基底。

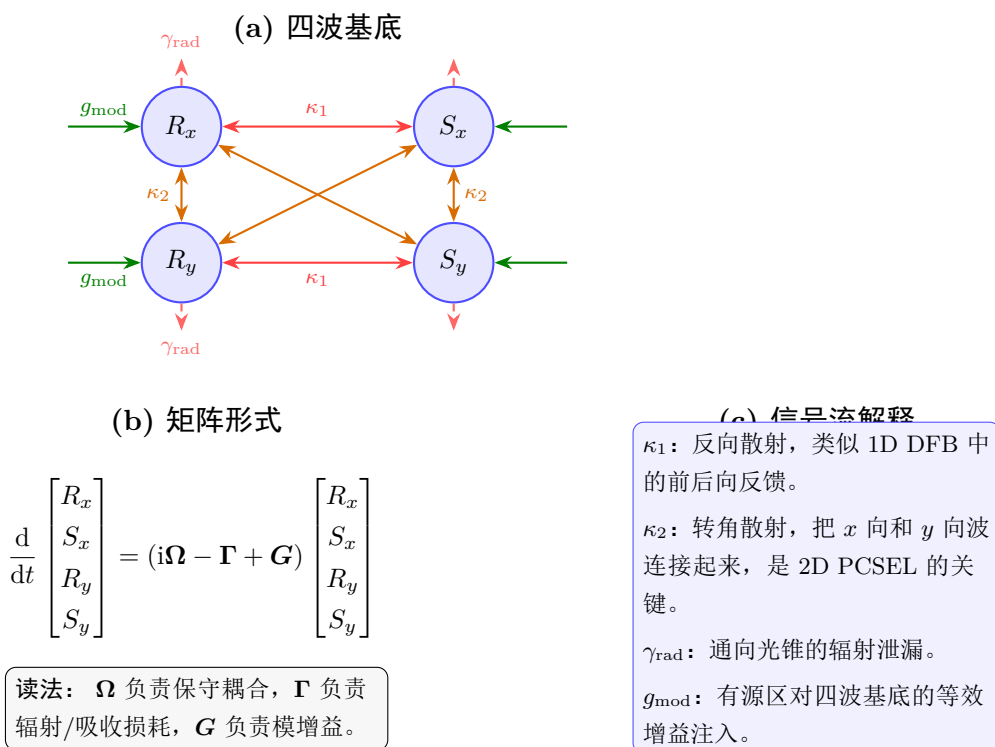


图 10.2: CWT 信号流与耦合网络。(a) 四波 ansatz 的实空间图像: 正向行波 R_x 、反向行波 S_x 、侧向行波 R_y 和 S_y 在面内传播; (b) 倒空间中的耦合路径: 红色双向箭头表示由倒格矢 G_x 驱动的反向散射 (耦合强度 κ_1), 橙色双向箭头表示由 $G_x \pm G_y$ 驱动的 90° 转角散射 (耦合强度 κ_2); (c) 完整的信号流图: 每个节点代表一个行波基底的复振幅, 边上的标记表示对应的耦合机制 (DFB 反馈、转角散射、辐射泄漏、材料增益), 闭环增益条件要求绕任意闭合回路一周的总增益等于总损耗。

图10.2 将耦合矩阵元素对应到散射网络。 κ_1 对应传统 1D DFB 激光器的反向散射机制, κ_2 对应 2D PCSEL 中 x 与 y 方向波分量之间的转角散射; 后者使面内反馈具有二维特征, 并

参与表面发射通道的形成。

10.5 三类矩阵项各自代表什么

先看 Ω 。它负责相位演化和相互散射，也就是通常所说的“保守耦合”部分。若只讨论被动、无损、完全封闭的近似系统， Ω 的本征值主要决定模的频率位置和简并如何被打破。几何参数如孔半径、填充比和晶格常数，首先往往就是通过改变 Ω 中的耦合系数来改写模的排列。

再看 Γ 。它并不只是一项“把数值减掉”的阻尼常数，而是所有损耗通道在该低维基底中的压缩表示。对 PCSEL 来说，其中至少要区分两类来源：一类是垂直辐射损耗，它来自模与自由空间通道的耦合；另一类是内部损耗，包括自由载流子吸收、掺杂层吸收、金属邻近损耗和粗糙散射等。也就是说， Γ 本身就已经把“开放系统”信息带进了模型。

最后是 G 。如果只讨论冷腔，这一项可以先不写；但一旦要谈阈值， G 就不可回避。它描述的是有源区增益在这组波分量基底中的投影，而非抽象的“有源”二字。最简单的教学写法是

$$G = g_{\text{mod}} I,$$

即先假设四个主波共享同一个均匀模增益。若增益分布近似均匀，这确实是一个自然起点；但若注入明显非均匀，或不同波分量对应的场在横向上有不同重叠，那么 G 的对角元就会分裂，进一步表现出模式选择性。

严格表述（难度）

g_{mod} 的具体形式取决于量子阱材料、注入条件和工作温度。在第十八章中，材料增益被展开为 $g_{\text{mat}}(N, T, \hbar\omega)$ ；经由约束因子 Γ_{opt} 投影后得到 g_{mod} ，再代入 $G = g_{\text{mod}} I$ （或更一般的对角分裂形式）。若跳过这一步而把 g_{mod} 当作自由拟合参数，则 CWT 的可解释性会大幅下降，因为它把“有源区物理”变成了纯粹的数字调参。

于是，式 (10.5) 的本质就很清楚了：

- Ω 决定“哪些组合容易形成共振”；
- Γ 决定“这些组合会以多快的速度泄露或耗散”；
- G 决定“有源区是否优先把增益送给其中某些组合”。

真正的激射候选模，是这三者共同作用下的本征态，而不是只由其中某一项决定。

10.5.1 先说清归一化：重叠积分和耦合系数的数值可比性

在解析推导和数值仿真中，读者会频繁看到“场归一化”“模归一化”“包络振幅”这些说法。它们背后对应的并非唯一标准。

在被动本征问题中，本征场常常只确定到一个乘法常数，因此场幅度本身并不直接对应实际功率；可比较的是频率、损耗率、归一化后的重叠积分等量。在有源问题中，若把模展开

代入耦合模方程, 包络振幅的定义又取决于采用的是能量归一化、功率归一化还是某种方便解析的尺度。

因此, 本书对归一化采取一个务实原则: 只要某个公式中出现重叠积分、模增益或耦合系数, 就必须在附近说明归一化方式属于哪一类; 如果没有说明, 就不能把不同章节或不同软件的数值直接拿来比较。

本书后续会频繁使用两种最常见的归一化方式, 先在此给出可计算的标准写法:

$$\text{能量归一化 (被动本征模常用):} \quad \int_V \varepsilon_r(\mathbf{r}) |\mathbf{E}_n(\mathbf{r})|^2 dV = 1, \quad (10.8)$$

$$\text{功率归一化 (导波/端口问题常用):} \quad \frac{1}{2} \int_S \text{Re}(\mathbf{E}_n \times \mathbf{H}_n^*) \cdot \hat{\mathbf{z}} dA = 1. \quad (10.9)$$

其中 S 为与传播方向正交的截面; 对开放 slab 结构若需统计上下两个辐射通道, 应明确 S 是单侧截面还是上下两侧合并截面, 以免功率计数出现两倍差异。两种归一化得到的模幅 a_m 单位不同, 耦合系数 κ 的量纲也随之不同。跨章节或跨软件比较 κ 数值时, 必须先确认双方使用同一种归一化。

10.6 耦合系数为什么是几何和场重叠的压缩表达

最关键的非对角元来自周期扰动 $\Delta\varepsilon$ 对不同平面波分量的散射。把投影过程写开, 典型的耦合系数具有如下结构:

$$\kappa_{mn} \propto \int_V \Delta\varepsilon(\rho, z) [\mathbf{e}_0^*(z) \cdot \mathbf{e}_0(z)] e^{-i(\beta_m - \beta_n) \cdot \rho} dV. \quad (10.10)$$

式 (10.10) 表明耦合系数可由扰动 Fourier 分量和纵向场重叠估计, 而不只是拟合参数。它给出三点结论。

第一, 耦合并非由几何尺寸单独决定的, 而是由“几何的傅里叶分量”和“纵向模场重叠”共同决定。因此, 改变外延层厚度或波导层折射率, 也会通过 $\mathbf{e}_0(z)$ 改写面内耦合强度。

第二, 不同的几何扰动并不会平均地影响所有耦合通道。某种孔形改变可能主要放大 x 向和 y 向之间的耦合, 另一种扰动则可能主要增强进入光锥的散射。也正因为如此, PCSEL 设计不能只说“增加扰动”或“减小扰动”, 而必须问清楚“增强的是哪一类傅里叶分量”。

第三, 如果某个耦合系数同时参与面内反馈和垂直辐射, 那么它天然就对应设计中的权衡。因为增强它, 可能一边在提升反馈, 另一边也在提升辐射损耗。

把式 (10.10) 和前面的选择定则结合起来, 还能直接读出 κ_1, κ_2 为什么恰好由少数 Fourier 分量给定。对最小四波模型而言, 沿 $\pm x$ 的前后向耦合主要由

$$\beta_1 - \beta_2 \approx \left(\frac{2\pi}{a}, 0 \right)$$

对应的一族 Fourier 分量控制; 沿 x/y 两组波之间的交叉耦合则主要由

$$\beta_1 - \beta_3, \quad \beta_1 - \beta_4$$

所对应的 Fourier 分量控制。于是, κ_1, κ_2 并非两个凭经验命名的常数, 而是“沿不同倒格矢方向的几何 Fourier 分量, 经过同一纵向模场剖面重加权后得到的压缩参数”。

为了避免“从 Maxwell 方程突然跳到一个耦合矩阵”显得过快, 这里至少把一个矩阵元显式写出来。设方格晶格最主要的倒格矢为

$$\mathbf{G}_x = \left(\frac{2\pi}{a}, 0 \right),$$

并关注沿 $\pm x$ 传播的两支基本波之间的耦合, 即 $\beta_1 = (\beta, 0)$ 与 $\beta_2 = (-\beta, 0)$ 。把周期扰动展开为

$$\Delta\varepsilon(\rho, z) = \sum_{\mathbf{G}} \Delta\varepsilon_{\mathbf{G}}(z) e^{i\mathbf{G}\cdot\rho},$$

则由式 (10.10) 得

$$\begin{aligned} \kappa_{12} &\propto \int_V \sum_{\mathbf{G}} \Delta\varepsilon_{\mathbf{G}}(z) e^{i\mathbf{G}\cdot\rho} [\mathbf{e}_0^*(z) \cdot \mathbf{e}_0(z)] e^{-i(\beta_1 - \beta_2)\cdot\rho} dV \\ &= \sum_{\mathbf{G}} \int \Delta\varepsilon_{\mathbf{G}}(z) |\mathbf{e}_0(z)|^2 dz \int_{A_c} e^{i[\mathbf{G} - (2\beta, 0)]\cdot\rho} d^2\rho. \end{aligned} \quad (10.11)$$

在近布拉格条件 $2\beta \approx 2\pi/a$ 下, 面积积分只保留 $\mathbf{G} = \mathbf{G}_x$ 这一项, 于是

$$\kappa_{12} \equiv \kappa_1 \propto A_c \int \Delta\varepsilon_{\mathbf{G}_x}(z) |\mathbf{e}_0(z)|^2 dz. \quad (10.12)$$

式 (10.12) 说明矩阵元来自特定倒格矢 **Fourier** 分量与纵向模场重叠的乘积, 而不是抽象常数。其他矩阵元类似, 只是对应不同的 $\beta_m - \beta_n$ 与不同的主导倒格矢。

例子

对方格晶格圆形空气孔 (半径 r 、晶格常数 a 、背景折射率 $n_b \approx 3.4$), 主导倒格矢 $\mathbf{G}_x = (2\pi/a, 0)$ 对应的截面 Fourier 分量可解析估算为

$$\Delta\varepsilon_{\mathbf{G}_x} \approx 2 \left(\frac{r}{a} \right) J_1 \left(\frac{2\pi r}{a} \right) \cdot (n_{\text{air}}^2 - n_b^2) \approx -2 \left(\frac{r}{a} \right) J_1 \left(\frac{2\pi r}{a} \right) \cdot n_b^2,$$

其中 J_1 是一阶 Bessel 函数。对典型占空比 $r/a \approx 0.3$, $J_1(2\pi \times 0.3) \approx 0.40$, 因此

$$|\Delta\varepsilon_{\mathbf{G}_x}| \approx 2 \times 0.3 \times 0.40 \times 11.6 \approx 2.8.$$

再乘以纵向重叠积分 ([本书算例] 中光子晶体层常贡献约 10–20% 纵向模能量), 有效耦合系数 κ_1 的典型量级约为 $\kappa_1/(\omega/c) \sim 10^{-2} - 10^{-1}$, 对应波矢修正 $\delta k \sim |\kappa_1|/v_g \sim 1 - 10 \text{ cm}^{-1}$ 。这应读作“数量级一致性”而非通用拟合律; 与已报道 CWT/实验量级对照可见于 [41, 42, 74]。

10.7 四个基本波如何重新组合成实际候选模

四波基底真正有价值的地方, 不在于记住 R_x, S_x, R_y, S_y 四个名字, 而在于理解它们会重新组合成更适合讨论对称性的超模。以方格晶格为例, 常见的线性组合包括“全同相”的亮模、“沿某方向相消”的暗模, 以及在 x 、 y 两个方向上具有不同奇偶性的组合。不同线性组合与辐射通道的重叠不同, 因而辐射损耗也不同。

这一步正好把前面几章连起来: 第5章 告诉我们 Gamma 点为什么容易产生简并和模式分裂, 第6章 解释了哪些散射通道打开了垂直辐射, 而在 CWT 里, 这些叙述第一次统一成了一个低维本征问题。求解式 (10.5) 的本征向量, 我们得到的就是“由四个基本波相干叠加形成的候选超模”; 求其本征值, 则能读出频率位置和净增长或净衰减趋势。

如果把式 (10.5) 写成本征值形式

$$(\mathbf{i}\Omega - \mathbf{\Gamma} + \mathbf{G}) \mathbf{u}_j = \lambda_j \mathbf{u}_j, \quad (10.13)$$

那么 $\mathbf{u}_j$ 就给出第 j 个候选超模在四个基本波基底上的权重, 而 λ_j 的实部则表征该超模在小信号近似下是净增长还是净衰减。

10.8 把四波基底改写成 C_{4v} 的 irrep 基底

为了使 CWT 成为可复用的教材工具, 不能只说明四个基本波可形成若干组合, 还应把这些组合写成 第8章 little-group 框架中的不可约表示。对方格晶格, Γ 点 little group 为 C_{4v} 。在最小四波子空间

$$\mathbf{a} = \begin{bmatrix} R_x & S_x & R_y & S_y \end{bmatrix}^T$$

中, 一个特别有用的 irrep 基底可取为

$$\mathbf{u}_{A_1} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad (10.14)$$

$$\mathbf{u}_{B_1} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \quad (10.15)$$

$$\mathbf{u}_{E,x} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{u}_{E,y} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}. \quad (10.16)$$

其中 $\mathbf{u}_{A_1}$ 是全同相 singlet, $\mathbf{u}_{B_1}$ 描述 x 向与 y 向波对之间的相对反相, 而 $\mathbf{u}_{E,x}, \mathbf{u}_{E,y}$ 则张成二维 E 双重态, 对应最直接的 x -like 与 y -like 偏振支路。

把这四个列向量拼成变换矩阵

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 1 & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & -1 & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & -1 \end{bmatrix}, \quad (10.17)$$

则对理想方格晶格 (κ_1, κ_2 均为实数), CWT 矩阵 $\tilde{H}$ 在 irrep 基底 $\mathbf{b} = U^\dagger \mathbf{a}$ 下化为块对角形式: A_1 与 B_1 各对应一个 1×1 块, E 双重态对应一个 2×2 块 (在 E 子空间中仍可能存在内部耦合, 但与 A_1, B_1 块解耦)。

这一写法说明, 四波 CWT 中应关注的是这些分量在 little-group 意义下如何组织成 A_1 、 B_1 和 E family, 而非单个 R_x 或 S_y 分量。于是:

- 模式标注不再依赖“看图感觉像哪种场型”, 而可以直接由 irrep 给出;
- E 双重态天然与 第八章 中的双偏振/法向辐射通道对应;
- A_1 和 B_1 则天然提示“这是 symmetry-distinct singlet”, 其辐射与混合规则必须和 E 分开看。

式 (10.14) 至式 (10.16) 是最小四波子空间中的 irrep 分类, 而不是整个 Maxwell 问题所有 irrep 的完整列表。在这个最小子空间里, A_2 和 B_2 没有显式出现; 若加入更多 Fourier 分量、更多纵向模或更完整矢量分量, 它们可以在更高维投影中出现。

这件事最好再说得更物理一些。最小四波基底只保留

$$(\pm\beta, 0)\hat{\mathbf{y}}, \quad (0, \pm\beta)\hat{\mathbf{x}}$$

这四支 TE-like 基本波, 因此它张成的是一个四维表示。对该表示做 character reduction, 恰好只分解出 $A_1 \oplus B_1 \oplus E$ 。要得到 A_2/B_2 , 至少还需要再引入额外的场自由度, 例如更高阶 Fourier 分量、不同纵向场型, 或使镜面对称下的奇偶结构足以支撑新的 singlet。因此 A_2/B_2 的缺席是最小四波子空间本身就不承载它们, 而非“被我们忘了写”。

10.9 这套 irrep 基底怎样直接连接偏振、辐射与 BIC 判断

现在可以把 第八章 和上一章的分类工具直接落到 CWT 矩阵里。对方格晶格, 法向自由空间辐射通道在最小 in-plane 分类下属于 E 。因此, 在保持 C_{4v} 对称的前提下:

- E 双重态最直接地与法向辐射通道匹配, 因此往往承担“可法向辐射偏振支路”的角色;
- A_1 与 B_1 这类 singlet 即使存在, 也不会与 E 通道发生对称性允许的直接混合;
- 若某一 singlet family 在 exact Γ 点呈现 direct normal-radiation suppression, 就应优先把它放进 symmetry-protected dark family / BIC 候选的语境中理解。

于是, CWT 的矩阵语言第一次真正和“irrep 标注”“偏振选择”“BIC 初筛”闭合起来: 求式 (10.13) 的本征向量之后, 读者不仅能说“哪个模本征值最大”, 还能说“这个本征模在 irrep 意义下属于哪一类, 它在对称性上是否允许法向辐射, 它与哪一支偏振通道最直接相关”。

10.10 对称性破缺怎样在 CWT 矩阵里体现为 irrep 分裂

当结构保持理想方格对称时, CWT 矩阵在 irrep 基底中往往接近分块形式: A_1 、 B_1 和 E family 彼此分离。引入小对称性破缺后, 例如圆孔变椭圆、双晶格位移或外延/电极造成的有效各向异性, 这种分块结构会被重新组织。

对常见偏振工程问题, 关键步骤是: 二维 E 双重态在降群后分裂成两支非简并 singlet, 使 x -like 与 y -like 偏振不再等价。同时, 某些原本 direct radiation 受限的 family 可能因降群允许混合而获得有限辐射损耗。因此, 在 CWT 中, “偏振钉扎” “模式分裂” “quasi-BIC 打开” 可视为同一耦合矩阵在不同对称性下的不同投影结果。

若方格晶格圆孔的 C_{4v} 对称被轻微椭圆孔降到只保留两条镜面对称的子群, 则原先的 E 双重态会沿椭圆主轴分裂成两支非简并偏振模。在 CWT 矩阵里, 这表现为 $u_{E,x}$ 与 $u_{E,y}$ 所在子块不再共享同一本征值。若其中一支同时更强地接近法向辐射通道, 另一支则更接近 dark family, 就会同时观察到偏振选择、损耗分裂与远场主瓣定向。

10.11 为什么二维八波 CWT 仍然不够: 它不是错误, 而是压缩得太狠

已有的二维八波 CWT 对 PCSEL 入门非常重要, 因为它第一次把 “四个基本波 + 若干高阶中继波” 这件事说清楚了。但教材里应进一步说明: 二维模型之所以会在真实器件中失效, 是因为它压缩了三个在真实 PCSEL 中会重新变重要的自由度, 而非因为它方向错了。

1. **纵向场自由度被压掉了。** 二维模型默认不同波矢分量共享相同的垂直场型, 而真实器件中的基本波、辐射波和高阶波在 z 方向上往往看见的是不同的波导环境;
2. **高阶 Fourier 分量被截断得过狠。** 对圆孔、弱调制、近对称结构, 这种截断常可接受; 但对深刻蚀、非对称孔形、多孔单胞和三晶格结构, 高阶 Fourier 分量对二维耦合本身就可能是主导项;
3. **垂直辐射被过度经验化。** 若只用一个拟合辐射常数去代表所有法向耦合, 就很难可靠处理上下出光不对称、远场整形与结构微扰引起的辐射重排。

因此, “从二维 CWT 升级到三维 CWT” 的含义是显式纳入真实 PCSEL 中被二维模型压缩、但会重新影响结果的通道, 而非单纯增加模型复杂度。

10.12 三维 CWT 能算什么: 一页速览

上述三点压缩, 在第十一章中会被逐项恢复: Liang-Noda 三维 CWT 从矢量波动方程出发, 让辐射波与高阶波各自携带真实的纵向场型并用 Green 函数解析求解, 最终仍归结为一个 4×4 本征问题, 其中每个矩阵元都由孔形、填充比、晶格常数和外延层栈解析确定, 原则上不需要经验拟合参数。

在进入完整推导之前, 先记住几组数字 (均来自 [41]), 它们解释了为什么这套模型在设计流程中不可替代:

- **速度:** 一组参数下求解 4×4 本征问题约需 1 秒, 而同等精度的 3D FDTD 约需 4 小时; 整条能带 (约 100 个 $\mathbf{k}$ 点) 三维 CWT 约 2 分钟, 3D FDTD 需要数百小时。
- **精度:** 模式频率预测与 3D FDTD 吻合到 $< 1\%$ 。
- **物理输出:** 对方格晶格圆孔, 模式 A/B 在精确 Γ 点辐射常数为零 (即 第 九章 的 symmetry-protected BIC); 换成直角等腰三角形孔后 α_{rad} 约为 $50\text{--}200\text{ cm}^{-1}$, Q 从无穷大变为有限可设计。
- **收敛代价:** 辐射常数收敛需要把平面波截断取到 $D \geq 10$ (即 441 个分量), 而频率收敛 $D \geq 5$ 已经足够——两类输出的收敛门槛并不相同。

完整推导、有限尺寸包络方程和非 Γ 点能带扩展见 第 十一章; 从 CWT 本征矢量提取远场偏振 (方位角、椭圆率与 Stokes 参量) 的实用公式已收入 第 二十八章。本章余下部分回到最小模型能够独立回答的问题: 阈值条件、模式竞争与设计语言。

10.13 CWT 怎样表达阈值条件

现在把 第 七章 的阈值条件接进来。如果仅讨论光学阈值而不求阈值电流, 那么在 CWT 里最直接的说法是: 当某个本征值的实部恰好穿过零时, 该超模达到线性阈值。也就是说, 阈值条件可写成

$$\max_j \text{Re}(\lambda_j) = 0. \quad (10.18)$$

这个表达比 “ $g_{\text{th}} = \alpha_{\text{rad}} + \alpha_{\text{int}}$ ” 更适合描述 PCSEL 的模竞争, 因为它允许不同超模具有不同的辐射损耗、内部损耗投影和增益重叠。CWT 并不推翻 第 七章 的阈值公式, 而是把单模阈值条件推广为多模耦合后的超模本征值问题。

若把第 j 支超模的本征值显式拆成

$$\lambda_j = v_{\text{en},j} [\Gamma_{\text{opt},j} g_{\text{mat}} - \alpha_{\text{rad},j} - \alpha_{\text{int},j}] + i \Delta\omega_j, \quad (10.19)$$

那么 式 (10.18) 就立刻退化回

$$\Gamma_{\text{opt},j} g_{\text{mat}} = \alpha_{\text{rad},j} + \alpha_{\text{int},j} \quad (\text{对最先起振的 } j). \quad (10.20)$$

这也是 第 七章 中单模阈值平衡式在 CWT 超模语言下的写法。区别只在于: Ch07 把这些量写成单模参数, CWT 则允许它们随超模 family 改变。

但也必须保持严格: 式 (10.18) 给出的仍然只是光学线性阈值。若要把它变成阈值电流, 仍然必须知道 $\mathbf{G}$ 如何随注入电流、载流子分布和温度变化。这一部分必须由 第 十八章 至 第二十章 提供。

10.14 从线性阈值到模式竞争动力学

阈值条件 $\max_j \text{Re} \lambda_j = 0$ 给出的只是 “哪一支超模最先达到光学阈值”。但真实器件一旦跨过阈值, 就不再是单一阈值问题, 而是一个多模竞争的动力学问题。这是因为: 各超模

family 通常拥有不同的净增益和不同的辐射/重叠特性, 一旦某支模率先达到阈值, 其光场就会通过增益饱和反过来压制其他模的净增益, 而饱和本身又与注入、空间烧孔和非均匀电流分布交织在一起。

CWT 的本征值方程在这一层级已经不能直接使用, 因为式 (10.5) 中的 α 默认为固定背景场中的线性包络, 而阈值以上的激光场要求增益、载流子密度和光场三者满足自洽条件。更准确的动力学语言是耦合到多模速率方程:

$$\frac{dS_j}{dt} = [\Gamma_{\text{opt},j} g_{\text{mat}}(N, T) - \alpha_{\text{rad},j} - \alpha_{\text{int},j}] S_j + \beta_{\text{sp}} R_{\text{sp},j}, \quad j = 1, \dots, 4, \quad (10.21)$$

其中 S_j 是第 j 支超模的光子数, $\Gamma_{\text{opt},j} g_{\text{mat}}$ 中的 g_{mat} 不再是常数, 而是通过载流子速率方程与所有光子数共同决定:

$$\frac{dN}{dt} = \frac{I}{eV_{\text{act}}} - [R_{\text{rad}} + R_{\text{nr}}] - \sum_j \frac{\Gamma_{\text{opt},j}}{v_{\text{en}}} g_{\text{mat}}(N, T) S_j. \quad (10.22)$$

在这条方程链中, 每个超模的 $\Gamma_{\text{opt},j}$ 、 $\alpha_{\text{rad},j}$ 和 $\alpha_{\text{int},j}$ 来自 CWT 本征值分析; 注入电流 I 、载流子密度 N 和温度 T 由电注入、热反馈和增益模型给出。因此, CWT 提供模式竞争的线性输入, 阈值以上哪一支模占优还需由速率方程或自洽动力学模型决定。

从教学编排上说, 本章之所以把动力学问题留到这里才提, 是因为读者必须先知道“各超模的净损耗和重叠因子从哪来”(CWT 的职责), 才能追问“一旦跨过阈值, 这些参数如何被速率方程重新分配”(第二十三章的职责)。若跳过这一步, 直接把 $\max_j \text{Re } \lambda_j = 0$ 当成最终器件结论, 就会在阈值和动态稳定性之间留下一个不该有的概念空白。

严格表述 (难度)

式 (10.21) 中的 $\Gamma_{\text{opt},j}$ 、 $\alpha_{\text{rad},j}$ 和 $\alpha_{\text{int},j}$ 均来自 CWT 本征值分解, 它们已经是给定结构几何和外延层栈下的“冷冻”参数。真正的动态竞争发生在阈值以上: 先起振的模通过增益饱和改变 $g_{\text{mat}}(N, T)$, 而 g_{mat} 的改变又反过来重新定义各模的净增益——这个反馈回路只能用时域或稳态自洽方程描述, 而不能用 CWT 的线性本征值直接读出。

10.15 一个最典型的设计问题: 几何扰动到底在帮你什么

很多设计讨论会说“加一个小扰动以获得单模”或者“调孔形以改善偏振”, 但这类说法只有落到 CWT 的语言里才真正可检验。设某个小扰动主要改变了 x 向与 y 向基本波之间的耦合, 而几乎不改变与光锥的直接耦合, 那么它更可能首先体现在模式分裂和偏振选择上, 而未必显著降低阈值。相反, 如果某个扰动主要增强了与辐射通道相关的矩阵元, 那么它可能更直接改变辐射损耗和提取效率。

考虑两种孔形修改方案。方案 A 主要改变晶格的二阶傅里叶分量, 使 x 向与 y 向波分量的交叉耦合显著增强; 方案 B 主要增强与零面内动量辐射通道相关的傅里叶分量。用 CWT 语言看, 方案 A 更容易引起模式分裂和偏振选择, 方案 B 更容易改写辐射损耗与输出耦合。若目标是降低阈值并维持单峰远场, 两者的优先级可能完全不同。

这个例子说明, CWT 的价值不只在计算速度, 还在于把设计动作分解为可逐项检查的物理效果。

10.15.1 三晶格如何直接进入 CWT: Fourier 放大因子与 180° 耦合增强

若要把多孔单胞或三晶格结构真正纳入教材, 而不是只把它当成“经验上更强反馈”的例子, 最好的方式是直接在 Fourier 系数层面写清楚。设单晶格光子晶体的介电扰动写成

$$\Delta\epsilon_1(x, y) = \sum_{m,n} F_{m,n} \exp\left[i\frac{2\pi}{a}(mx + ny)\right]. \quad (10.23)$$

若在同一晶胞中再叠加两个相对位移分别为 (d_{1x}, d_{1y}) 和 (d_{2x}, d_{2y}) 的嵌套晶格, 则三晶格总扰动满足

$$\Delta\epsilon_{3L}(x, y) = \sum_{m,n} A_{m,n} F_{m,n} \exp\left[i\frac{2\pi}{a}(mx + ny)\right], \quad (10.24)$$

其中放大因子为

$$A_{m,n} = 1 + \exp\left[-i\frac{2\pi}{a}(md_{1x} + nd_{1y})\right] + \exp\left[-i\frac{2\pi}{a}(md_{2x} + nd_{2y})\right]. \quad (10.25)$$

对正方晶格 PCSEL 而言, 四个基本波之间最关键的是 180° 反向耦合, 因此首先应关注 $(m, n) = (2, 0)$ 与 $(0, 2)$:

$$A_{2,0} = 1 + \exp\left(-i\frac{4\pi}{a}d_{1x}\right) + \exp\left(-i\frac{4\pi}{a}d_{2x}\right), \quad (10.26)$$

$$A_{0,2} = 1 + \exp\left(-i\frac{4\pi}{a}d_{1y}\right) + \exp\left(-i\frac{4\pi}{a}d_{2y}\right), \quad (10.27)$$

因此有

$$0 \leq |A_{2,0}|, |A_{0,2}| \leq 3. \quad (10.28)$$

这条结果表明, 三晶格的优势不在于单纯增加孔数, 而在于可将关键二阶 Fourier 分量从相消调到三倍相干增强, 从而定向增强 180° 反馈通道 [42, 69]。

对设计而言, 这还有两层重要含义。第一, 若把 $A_{2,0}$ 或 $A_{0,2}$ 调到接近零, 就会系统性削弱反向反馈; 第二, 若把二者同时调到接近 3, 则可显著增强面内光反馈、降低水平泄露, 并为更小尺寸谐振腔创造条件。但边界仍需保留: 增强 180° 耦合首先改善的是冷腔反馈与模式损耗排序, 并不自动等于器件阈值电流最低。真实阈值仍要继续经过 C_{rad} 、内部吸收、量子阱重叠、注入与热回写的再筛选。

10.16 CWT 到底能算什么, 不能算什么

如果用得恰当, 最小 CWT 至少能做五件非常有价值的事。

第一, 它能给出候选超模的构成, 告诉我们目标模主要由哪些基本波叠加而成。

第二, 它能预测某些几何参数变化会沿哪条物理通道起作用, 是更多地改频率、改损耗, 还是改模式分裂。

第三, 它能用很低的计算代价做灵敏度扫描, 为 RCWA、FDTD、FEM 缩小参数搜索空间。

第四, 它能在冷腔或简化增益模型下, 快速比较几个候选模的相对阈值趋势。

第五, 它能把“偏振工程”“远场工程”“模式抑制”放进同一种耦合语言下理解, 而不是把这些主题拆成彼此独立的经验技巧。

但它同样有明确的不能之处。

它不能严格替代全波方法去计算开放三维结构的精确辐射损耗; 不能在没有外延、电学和热学模型的前提下直接给出阈值电流; 不能可靠处理远离所选共振子空间的大尺度模重组; 也不能自动处理有限尺寸边缘、注入不均匀和强热漂移引起的复杂模式变形。

易错点

把 CWT 输出的“目标模本征值最大”直接当成“器件一定单模工作”的证据, 是非常常见的误判。CWT 最多告诉你在所选基底、所选近似和所选增益模型下, 某个超模线性起振更有优势。真实器件还要接受有限尺寸、热漂移、注入不均匀和非线性增益压缩的再筛选。

10.17 CWT 在整本书 workflow 中的准确位置

把本书的建模链条回顾一遍, 就能看清 CWT 的职责边界。

在第 [第四章](#) 至第 [第六章](#) 中, 我们建立的是模式分类和面发射的物理图像; 在第 [第七章](#) 至第 [第九章](#) 中, 我们把阈值、偏振和远场放进开放系统视角; 从本章开始, 则进入“半解析模型”层。这里的目标是建立一套足够清楚的中间语言, 使后续 RCWA、FDTD、FEM 以及器件级电热模型之间能互相对话, 而非追求一次性得出最终器件结论。

因此, 最合理的研究流程通常不是“CWT 或全波二选一”, 而是:

1. 用 PWE 或经验直觉找到 Γ 点附近的候选模族;
2. 用 CWT 理解主导耦合通道和参数灵敏度;
3. 用 RCWA 精修开放周期单胞的辐射与谱线行为;
4. 用 FEM/FDTD 验证有限尺寸器件模式;
5. 再把这些结果送入外延、电注入和热耦合模型, 判断器件级阈值与稳定工作区。

因此, CWT 在 PCSEL 研究中的作用不在于最高精度, 而在于充当跨模型的中间描述。

能带章节回答的是“哪些 Γ 点或近 Γ 点带边值得进入候选名单”; CWT 章节回答的是“这些候选模为何会因为 180° 反馈、二维高阶中继耦合与法向辐射通道而被重新排序”。二维 PWEM、三维 CWT 与后续 RCWA/FDTD/FEM 是同一个 PCSEL 问题在不同层级上的投影, 而非彼此替代关系。

严格表述 (难度)

第十一章 中的三维 CWT 推导 (第 [11.4](#) 节) 基于 Liang 等人 [[41](#), [42](#)] 针对方格和三角晶格的工作。该框架已被应用于三晶格 PCSEL 设计 [[69](#)], 其中 [文献值] 耦合系数 κ_{1D} 的计算直接指导了三个子晶格间距的优化, 将 180° 反馈耦合强度相比单晶格提升约三

倍, 从而显著降低阈值电流。这说明三维 CWT 可作为孔形、子晶格位移等工程参数优化的半解析工具。

常见错误: CWT 无限周期结果外推到有限阵列

错误理解: “CWT 给出某候选超模最优, 有限阵列器件一定保持同排序。” **正确做法:** 把 CWT 作为候选筛选与灵敏度解释层, 再用有限阵列全波验证尺寸效应。**最小验收:** 至少报告 3 个阵列尺寸下的主模/竞争模排序、远场主瓣角与旁瓣变化, 确认不存在尺寸引发的排序翻转。

10.18 2024–2026 前沿补充: CWT 的极化与晶格泛化正在变成可复用工具链

传统教学常把 CWT 先绑定到“方格晶格 + 近 Γ + 主偏振”的最小场景。近年的进展表明, 这个边界正在被系统扩展: 例如三角晶格、TM 偏振下的二维耦合波理论已经给出可复核的解析系数与模频预测, 并与有限器件计算形成了同向证据 [56]。这对本章最直接的意义是: CWT 不再只是“解释器”, 而是在更广晶格族上可用于前端参数筛选的半解析求解器。

与此同时, 快速设计工具的系统比较工作也指出: CWT 及其迭代扩展在 2D 设计空间中仍然具备很高性价比, 但其输出必须经由更高保真方法做边界校准 [36]。因此, 本章建议把 CWT 的工程使用口径固定为:

- 用于快速识别“哪类耦合通道最可能改写模式排序”;
- 用于给出参数扫描的方向性而非最终器件结论;
- 在进入发布级结论前, 必须与 RCWA/FDTD/FEM 的有限尺寸结果闭环互证。

回看本章

最小 CWT 的核心思想是: 在 Gamma 点附近选取少数主导平面波分量, 把完整 Maxwell 问题投影到一个低维耦合子空间中, 再把保守耦合、开放损耗和增益同时写进同一组矩阵方程。这样做的收益是把几何、外延和有源效应压缩成可解释的耦合参数; 代价则是模型只在所选子空间和近似条件内可信。本章进一步说明, 这个低维子空间不应只按“哪四个波分量”来读, 还应按 little-group 的 irrep 来读。对方格晶格最小四波模型, A_1 、 B_1 与 E family 的分解, 正是把模式标注、偏振选择和 BIC 初筛接到同一个矩阵语言上的关键一步。对 PCSEL 来说, CWT 既不是能带图的重复, 也不是全波求解器的廉价替身, 而是连接物理图像、对称性分类与定量设计的中间层。

练习题

1. [基础] 从式 (10.1) 出发, 说明为什么 Gamma 点附近最自然的最小基底是四个基本波, 而不是任意选取四个平面波。
2. [基础] 验证式 (10.14)至式 (10.16) 给出的四个向量在最小四波子空间中构成一组正交基, 并说明为什么它们适合作为 C_{4v} 的最小 irrep 基底。
3. [进阶] 结合式 (10.10), 讨论改变孔半径、改变刻蚀深度和改变纵向层栈厚度三种动作分别会怎样影响耦合系数。
4. [进阶] 说明为什么在保持 C_{4v} 的前提下, E family 与 A_1/B_1 family 不能直接混合; 再说明一旦降群, 这种结论为什么可能被改写。
5. [进阶] 解释为什么式 (10.18) 给出的是光学线性阈值, 而不是器件级阈值电流。
6. [综合] (思路提示: 先定层级, 再列关键式) 设计一个你认为“最小 CWT 会失效”的场景, 并指出失效是因为缺少哪些基底或物理通道。

延伸阅读

- 下一章讨论三维修正后, 可回到本章核对二维有效模型在何种 PCSEL 结构上会系统性失效。
- 如需与标准光子晶体教材对照, 可参考 [28, 57] 中关于平面波耦合、对称性和模式投影的相关章节。
- PCSEL 专用 CWT 的代表性建立过程可参考 [41, 42]。

第十一章 三维与更高阶 CWT：完整推导与器件接口

读完本章，你能

- 理解最小二维 CWT 为什么会在真实 PCSEL 中失效，以及失效通常先表现在哪里。
- 跟随 Liang-Noda 三维 CWT 的完整推导路线，理解辐射波与高阶波如何被 Green 函数解析求解，使辐射耦合从经验参数变为可算量。
- 掌握把纵向层栈、辐射通道和有限尺寸信息系统地并入 CWT 的基本思路。
- 学会判断何时值得继续升级 CWT，何时应当直接转向 RCWA、FEM 或 FDTD。

读前准备 建议已掌握 第**六**章、第**十**章和第**十七**章 中关于最小 CWT、纵向层栈与面发射机制的内容。

本章定位

本章是在其基础上有选择地恢复被压缩的自由度，而非对 第**十**章 的否定。具体而言，从 Ch09 继承到本章仍然成立的内容包括：（1）四波近似的基底选取思想（ $\pm x, \pm y$ 四个主导平面波分量）；（2）把保守耦合、开放辐射与增益/损耗写进同一组低维矩阵方程的物理框架；（3）irrep 基底变换与 C_{4v} 不可约表示的分类语言；（4）阈值条件 $\max_j \operatorname{Re} \lambda_j = 0$ 的形式。

本章**补全或升级**的内容包括：（1）给出 Liang-Noda 三维 CWT 的完整推导，把耦合矩阵显式分解为 $\mathbf{C}_{1D} + \mathbf{C}_{\text{rad}} + \mathbf{C}_{2D}$ ，其中每个矩阵元都解析可算；（2）辐射波与高阶波的纵向场型不再共用单一 $\phi_0(z)$ ，而是分别求解；（3）有限尺寸从常数振幅升级为空间包络方程；（4）非对称孔形、三晶格等高阶 Fourier 分量的处理方式。

因此，若把 Ch09 视为最小教学模型，本章是在保持其物理骨架的前提下，逐项显式纳入真实 PCSEL 中被压缩、但可能重新影响结果的通道。

阅读主线 本章所处模型层级：**L1**→**L2**

上图把本章“为何升级、升级什么、输出什么”放在同一视角下，便于在后文每一节定位当前恢复的是哪一类自由度。

上一章建立了最小 CWT，可解释许多 Gamma 点附近的主导耦合现象。真实 PCSEL 还包含纵向层栈、上下不对称辐射、金属和重掺杂层损耗，以及有限孔阵边界和横向注入不均

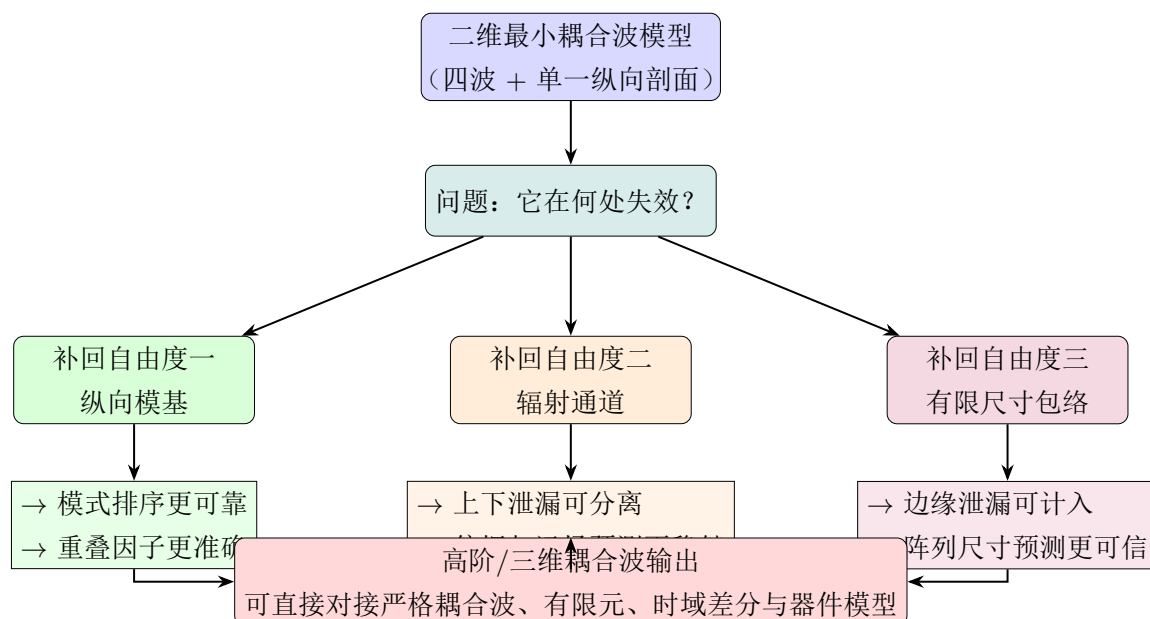


图 11.1: 本章耦合波升级路径：从二维最小模型出发，显式补回三个关键自由度（纵向模基、辐射通道、有限尺寸包络），得到可用于工程接口的高阶/三维模型。

匀。本章讨论四个问题：二维最小模型何时失真；升级时应显式纳入哪些自由度；Liang-Noda 三维 CWT 如何从矢量 Maxwell 方程一步步推导出闭合的 4×4 本征问题；升级后的 CWT 如何与全波和器件模型接口。

本章延续 第三章、第十章 的 $e^{-i\omega t}$ 约定。若后文把辐射、损耗和增益写成非厄米矩阵项，其符号含义默认继承这一全书约定。

11.1 为什么二维最小 CWT 迟早会不够用

最小 CWT 建立在有意压缩之上：将纵向复杂层栈压缩为一个主要导模剖面，将倒格空间压缩为少数主导平面波，将开放系统辐射损耗压缩为少数非厄米项。只要被压缩的自由度不是主导因素，该模型就很有效；一旦这些自由度开始影响结果，模型便会失真。

最常见的失真来源有四类。

第一类是纵向模剖面的改变不能再忽略。比如刻蚀深度较深、光子晶体层不再只是“轻微侧向扰动”，或者外延层中存在强烈的折射率跳变和吸收层，使得不同候选模的 z 向分布明显不同。这时，用单一的 $\mathbf{e}_0(z)$ 去代表所有模式，就可能把本来由纵向分布差异决定的损耗差别抹平。

第二类是上下辐射不对称开始主导。真实器件常常上方为空气、下方是衬底和多层半导体，甚至还叠加了金属、电极和封装。这样一来，“向上辐射”和“向下辐射”不再是同一个通道的镜像，而是两套不同的开放边界。二维有效模型若只给一个总辐射损耗，往往不足以讨论远场主瓣、提取效率和偏振稳定性。

第三类是高阶倒格矢和额外面内波分量变得重要。当几何扰动增强、孔形复杂或目标工作点远离最简单的四波闭合网络时，只保留四个基本波分量会遗漏参与竞争的通道。此时模

型不仅可能出现数值偏差，还可能给出错误的模式排序。

第四类是有限尺寸与横向非均匀性开始进入主导地位。实际器件有边缘、有限注入窗口、热斑和电流拥挤。这些效应会使包络在横向上强烈变化，而不再只是一个几乎均匀的整体振幅。最小 CWT 若没有横向包络方程，就很难描述这类器件级现象。

直觉框

若 CWT 对总体趋势仍有效，但在偏振、上下出光比、边缘损耗或外延微调问题上与全波结果反复不一致，通常说明模型基底不足。此时应增加相应自由度，而不是继续用旧模型拟合参数。

11.2 升级思路：恢复哪三类自由度

从方法上看，更高阶或三维 CWT 不只是扩大矩阵维数，而是恢复至少三类自由度。

第一类是纵向模自由度。也就是不再假定所有候选面内波分量共用一个固定的 z 向剖面，而是允许它们在若干背景纵模基底上展开。

第二类是更丰富的倒格空间分量。也就是除了四个基本波外，再引入那些虽然权重较小，但确实会通过散射、辐射或对称性选择规则影响结果的高阶分量。

第三类是横向包络与边缘通道。也就是让振幅不再只是一个全局常数，而成为横向坐标的函数，从而显式描述有限器件的模式包络、边缘吸收和横向非均匀注入。

若只把升级理解为增加自由度，会忽略其建模逻辑。更准确地说，升级是把原本合并进有效参数的物理过程，重新拆分为显式可追踪的通道。

11.3 从三维 Maxwell 方程重新投影：为什么会出现“平面波乘纵模”的基底

要把纵向层栈纳入模型，可先重新定义背景问题，而不必放弃 CWT。设背景结构为只沿 z 变化的层状波导，其介电常数记为 $\varepsilon_0(z)$ ；光子晶体层造成的横向周期扰动为 $\Delta\varepsilon(\rho, z)$ 。频域 Maxwell 方程写成

$$\nabla \times (\mu^{-1} \nabla \times \mathbf{E}) - \omega^2 \varepsilon(\rho, z) \mathbf{E} = 0, \quad \varepsilon(\rho, z) = \varepsilon_0(z) + \Delta\varepsilon(\rho, z). \quad (11.1)$$

时域到频域的过渡 CWT 的标准特征值形式式 (11.1) 隐含了一个关键的数学步骤：从含时的波动方程出发，假设场具有谐波依赖 $\mathbf{E}(t) = \mathbf{E}e^{-i\omega t}$ ，代入后消去时间因子得到频域本征值问题。这一过渡在封闭无源系统中是严格的，但在开放系统中需要注意两点：其一，辐射损耗会使本征频率变为复数 $\tilde{\omega} = \omega - i\gamma$ ，对应准模的衰减；其二，当考虑非线性或有源增益时，介电常数本身依赖于场强或载流子密度，此时频域本征值问题需要通过迭代或线性化处理。本章采用频域表述是为了突出模式结构和耦合机制，而在第二十三章中将回到时域框架讨论动态响应。对背景层状结构，先求出一组纵向本征基底 $\{\mathbf{e}_n(z)\}$ 。然后不再像上一章那样只用一

个 $\mathbf{e}_0(z)$ ，而是写成

$$\mathbf{E}(\rho, z) = \sum_{m,n} A_{mn}(\rho) \mathbf{e}_n(z) e^{i(\mathbf{k} + \mathbf{G}_m) \cdot \rho}. \quad (11.2)$$

这里 m 表示倒格空间中的面内分量， n 表示纵向模基底。式 (11.2) 表明，每个面内波分量不再共用同一个纵向场剖面，而是可在若干纵向模之间分配权重。

注记

开放系统中的非正交模。在封闭无损系统中，纵向模满足标准正交关系 $\int \mathbf{e}_m^*(z) \cdot \mathbf{e}_n(z) dz = \delta_{mn}$ 。但在开放系统中，辐射边界条件使本征模成为准模（quasi-modes），其本征值为复数 $\tilde{\omega}_n = \omega_n - i\gamma_n$ 。此时左、右本征模不再互为厄米共轭，而需要使用双正交（biorthogonal）归一化：

$$\int \tilde{\mathbf{e}}_m^L(z) \cdot \tilde{\mathbf{e}}_n^R(z) dz = \delta_{mn},$$

其中 $\tilde{\mathbf{e}}^L$ 和 $\tilde{\mathbf{e}}^R$ 分别是左、右本征模。若仍使用封闭系统的正交归一化，会导致模式重叠因子和辐射损耗的计算出现系统性偏差，尤其是对高 Q 准 BIC 模。

把式 (11.2) 投影回式 (11.1)，便得到

$$\sum_{m',n'} \left[\mathcal{L}_{mn,m'n'}(\omega) + \mathcal{C}_{mn,m'n'} \right] A_{m'n'} = 0. \quad (11.3)$$

其中 $\mathcal{L}$ 主要来自背景层状结构的传播算符， $\mathcal{C}$ 则来自周期扰动对不同倒格波分量和不同纵模分量的耦合。与上一章相比，这里的矩阵不只告诉我们“哪个面内波耦合到哪个面内波”，还告诉我们“这种耦合会不会顺便把场从一种纵向剖面推到另一种纵向剖面”。

二维有效折射率近似相当于把式 (11.2) 中的纵模索引 n 截断为一个主基底。若单基底近似足够好，这样做是有效的；但当不同候选模在纵向重叠、损耗层接触或辐射出口上差异显著时，只保留一个 n 会把模式选择机制过度平均化。

对第一次真正实现三维 CWT 的读者，还需要一个更可执行的截断判据。最直接的做法，是定义主基底占比

$$\eta_{N_z} = \frac{\sum_m \sum_{n=1}^{N_z} |A_{mn}|^2}{\sum_m \sum_{n=1}^{N_{\max}} |A_{mn}|^2}, \quad (11.4)$$

并要求在把纵向基底数从 N_z 增加到 $N_z + 1$ 或 $N_z + 2$ 时，目标模频率、辐射损耗和关键重叠因子的变化都已经进入平台区。式 (11.4) 不是唯一标准，但它至少给出一个工程上可执行的问题：到底是“主基底已足够”，还是“你只是还没检查第二、第三纵模对损耗与偏振的改写”。教学上更实用的建议是：若 $\eta_{N_z} < 10^{-2}$ 且增广一个到两个纵向基底后关键结论不翻转，可先把该高阶纵模视为次要修正；若问题涉及偏振翻转、上下出光比或近简并模排序，则最好把门槛再收紧到 $\eta_{N_z} < 10^{-3}$ 。

上一节的多纵模投影给出的是一般框架：它告诉我们升级的方向，但没有给出可直接编程实现的闭合公式。本节转入一个可完全解析闭合的具体实现——Liang-Noda 三维 CWT [41]。它在基本波层面仍采用单一纵向剖面 $\phi_0(z)$ （对应上一节框架中只保留主纵模基底），但把辐

表 11.1: 纵向模截断的教学级判据。满足任一“可丢弃”条件时，可暂不继续增加该纵模。

判据类别	可丢弃条件	物理含义
频率失谐	候选纵模与目标模失谐超过目标频率的约 5%–10% ([工程经验] 筛选阈值)	该纵模已远离目标频段，通常只剩高阶修正作用
能量占比	$\eta_{N_z} < 10^{-2}$ ；若近简并或强不对称问题取 10^{-3}	该纵模对目标模的权重已低到只会提供次要修正
有源区重叠	$\Gamma_{n_z} < 0.01 \Gamma_1$	与有源区或关键损耗层重叠太弱，不足以显著改写阈值链
收敛检验	纵模数从 N_z 增至 $N_z + 1$ 或 $N_z + 2$ 后，目标模频率变化 $< 0.5\%$ ，辐射损耗变化 $< 5\%$ ，关键耦合系数变化 $< 1\%$	当前截断已足以稳定主导结论

射波与高阶波分别用各自的 Green 函数显式求解，从而把辐射耦合从经验参数变成解析可算量。

11.4 从二维到三维：Liang–Noda 3D CWT 的推导框架

上一章建立的最小四波 CWT 是一个二维模型：它假设结构沿 z 方向均匀，因此无法描述法向辐射（面发射）这一 PCSEL 最核心的特征。Liang 等人 [41] 发展了三维 CWT，克服了这一根本限制。本节按照该文献的推导路线，给出从三维矢量 Maxwell 方程到耦合波本征方程的完整物理链条。

11.4.1 为什么必须从矢量波动方程出发

二维 CWT 以 H_z 标量方程为起点。但法向辐射波的电场分量为 (E_x, E_y) ，不含 H_z ，因此 H_z 标量方程从根本上无法描述面发射。三维 CWT 的第一步改进，就是回到矢量波动方程

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) = k_0^2 \tilde{n}^2(\mathbf{r}) \mathbf{E}(\mathbf{r}), \quad (11.5)$$

其中 $k_0 = \omega/c$ ， $\tilde{n}^2(\mathbf{r}) = n^2(\mathbf{r}) + 2in(\mathbf{r})\tilde{\alpha}(\mathbf{r})/k_0$ 包含折射率实部和增益/损耗虚部。TE 极化场 $\mathbf{E} = (E_x, E_y, 0)$ 按 Bloch 定理展开为

$$E_i(\mathbf{r}) = \sum_{m,n} E_{i,mn}(z) e^{-i(m\beta_0 x + n\beta_0 y)}, \quad i = x, y, \quad (11.6)$$

折射率平方展开为

$$n^2(\mathbf{r}) = n_0^2(z) + \sum_{(m,n) \neq (0,0)} \xi_{m,n}(z) e^{-i(m\beta_0 x + n\beta_0 y)}, \quad (11.7)$$

其中 $\beta_0 = 2\pi/a$ ， $n_0^2(z)$ 是各层平均介电常数， $\xi_{m,n}(z)$ 是光子晶体层内的 Fourier 系数（光子晶体层外为零）。

将式 (11.6) 和式 (11.7) 代入式 (11.5)，收集所有乘以 $e^{-i(m\beta_0 x + n\beta_0 y)}$ 的项，得到关于每个 (m, n) 分量的纵向方程组 [41]：

$$\left[\frac{d^2}{dz^2} + k_0^2 n_0^2 + 2i\tilde{\alpha} k_0 n_0 - n^2 \beta_0^2 \right] E_{x,mn} + mn\beta_0^2 E_{y,mn} = -k_0^2 \sum_{\substack{m',n' \\ (m',n') \neq (m,n)}} \xi_{m-m',n-n'} E_{x,m'n'}, \quad (11.8)$$

$$\left[\frac{d^2}{dz^2} + k_0^2 n_0^2 + 2i\tilde{\alpha} k_0 n_0 - m^2 \beta_0^2 \right] E_{y,mn} + mn\beta_0^2 E_{x,mn} = -k_0^2 \sum_{\substack{m',n' \\ (m',n') \neq (m,n)}} \xi_{m-m',n-n'} E_{y,m'n'}, \quad (11.9)$$

$$\frac{d}{dz} [m\beta_0 E_{x,mn} + n\beta_0 E_{y,mn}] = 0. \quad (11.10)$$

式 (11.10) 是横向性约束 (transversality constraint)，来自 $\nabla \cdot (\varepsilon \mathbf{E}) = 0$ 。这三个方程是三维 CWT 的完整出发点，后续所有推导都是对它们做分类求解。

11.4.2 基本波的场型假设 (separation of variables)

三维 CWT 的关键技术步骤是变量分离：假设每个波分量的场可以写成面内振幅与纵向剖面的乘积。对二阶 Γ 点共振，四个基本波的 TE 极化场分量为 [41]

$$(m, n) = (1, 0) : \quad E_{x,1,0} = 0, \quad E_{y,1,0} = R_x \phi_0(z), \quad (11.11)$$

$$(m, n) = (-1, 0) : \quad E_{x,-1,0} = 0, \quad E_{y,-1,0} = S_x \phi_0(z), \quad (11.12)$$

$$(m, n) = (0, 1) : \quad E_{x,0,1} = R_y \phi_0(z), \quad E_{y,0,1} = 0, \quad (11.13)$$

$$(m, n) = (0, -1) : \quad E_{x,0,-1} = S_y \phi_0(z), \quad E_{y,0,-1} = 0. \quad (11.14)$$

这里有两个要点。第一，沿 $\pm x$ 传播的基本波只有 E_y 分量，沿 $\pm y$ 传播的只有 E_x 分量——这是 TE 极化的直接要求（电场垂直于传播方向）。第二，四个基本波共享同一个纵向剖面 $\phi_0(z)$ ，即无周期调制时的基模场，满足

$$\frac{d^2 \phi_0}{dz^2} + [k_0^2 n_0^2(z) - \beta^2] \phi_0 = 0, \quad \int_{-\infty}^{\infty} |\phi_0(z)|^2 dz = 1. \quad (11.15)$$

β 和 $\phi_0(z)$ 可由传递矩阵法 (TMM) 对无周期调制的平板波导求解得到。

11.4.3 从分量方程到耦合方程的推导

以 $(m, n) = (1, 0)$ 为例，将式 (11.11) 代入式 (11.9)，左侧变为

$$\left[\frac{d^2 \phi_0}{dz^2} + (k_0^2 n_0^2 + 2i\tilde{\alpha} k_0 n_0 - \beta_0^2) \phi_0 \right] R_x.$$

利用式 (11.15) 消去 $d^2 \phi_0 / dz^2$ ，得到

$$(\beta^2 - \beta_0^2) R_x \phi_0 + 2i\tilde{\alpha} k_0 n_0 R_x \phi_0 = -k_0^2 \sum_{\substack{m',n' \\ (m',n') \neq (1,0)}} \xi_{1-m',-n'} E_{y,m'n'}. \quad (11.16)$$

在 $\beta \approx \beta_0$ 的近 Bragg 条件下, $\beta^2 - \beta_0^2 \approx 2\beta_0\delta$, 其中 $\delta = \beta - \beta_0$ 。再对式 (11.16) 两侧乘以 $\phi_0(z)$ 并沿 z 积分 (利用归一化条件), 右侧的求和按波矢类型分成三组:

1. 另一个基本波 $(m', n') = (-1, 0)$: 贡献 $\xi_{2,0}$ 的重叠积分 $\rightarrow$ 一维反馈耦合;
2. 辐射波 $(m', n') = (0, 0)$: 贡献 $\xi_{1,0}$ 与辐射波场的重叠 $\rightarrow$ 辐射耦合;
3. 高阶波 $m'^2 + n'^2 > 1$: 贡献各阶 ξ 与高阶波场的重叠 $\rightarrow$ 二维耦合。

对其余三个基本波做完全类似的推导, 最终得到四波耦合方程组

$$(\delta + i\alpha)\mathbf{V} = \mathbf{C}\mathbf{V}, \quad \mathbf{V} = \begin{bmatrix} R_x & S_x & R_y & S_y \end{bmatrix}^T, \quad (11.17)$$

其中 $\delta = \beta - \beta_0$ 是偏离 Bragg 条件的传播常数偏移, $\alpha = \text{Im}(\lambda)$ 表示该本征支的空间增益/损耗 (被动结构中 $\alpha > 0$ 对应沿传播方向衰减)。关键在于耦合矩阵 $\mathbf{C}$ 自然分解为三个物理来源明确的子矩阵:

$$\mathbf{C} = \mathbf{C}_{1D} + \mathbf{C}_{\text{rad}} + \mathbf{C}_{2D}. \quad (11.18)$$

11.4.4 辐射波的解析求解

在得到四波耦合方程式 (11.16) 之后, 右侧仍然包含未知的辐射波场和高阶波场。要把耦合方程变成闭合的本征问题, 必须先把这些未知场用基本波振幅表示出来。本小节和下一小节分别处理辐射波和高阶波。

辐射波对应 $(m, n) = (0, 0)$, 即零面内波矢分量。将 $(m, n) = (0, 0)$ 代入式 (11.8) 和式 (11.9), 并只保留基本波作为源项 (这是三维 CWT 的核心近似之一), 得到

$$\left[\frac{d^2}{dz^2} + k_0^2 n_0^2(z) \right] E_x(z) = -k_0^2 (\xi_{0,-1} R_y + \xi_{0,1} S_y) \phi_0(z), \quad (11.19)$$

$$\left[\frac{d^2}{dz^2} + k_0^2 n_0^2(z) \right] E_y(z) = -k_0^2 (\xi_{-1,0} R_x + \xi_{1,0} S_x) \phi_0(z), \quad (11.20)$$

其中 $E_x(z) \equiv E_{x,0,0}(z)$, $E_y(z) \equiv E_{y,0,0}(z)$ 是辐射波的两个电场分量。注意左侧算符 $[\partial_z^2 + k_0^2 n_0^2(z)]$ 恰好是纵向 Helmholtz 方程, 其 Green 函数满足

$$\left[\frac{d^2}{dz^2} + k_0^2 n_0^2(z) \right] G(z, z') = -\delta(z - z'), \quad (11.21)$$

解为

$$G(z, z') = -\frac{i}{2k_z} e^{-ik_z|z-z'|}, \quad k_z = k_0 n_0(z), \quad (11.22)$$

其中 k_z 是辐射波在纵向的波数。这里使用了忽略层状波导界面反射的近似形式; 更精确的表达式可通过传递矩阵法构造 [41]。

严格表述 (难度)

式 (11.22) 不是层状波导的严格 Green 函数, 而是把辐射波在光子晶体层附近近似看成“局域均匀背景中的向外行波”。它的适用条件至少包括: 层栈界面反射不是主导效应;

辐射波纵向传播不接近明显的 Fabry-Pérot 共振；以及我们关心的是 $\mathbf{C}_{\text{rad}}$ 的主导数量级与参数趋势，而不是每个矩阵元的百分之一级精度。若界面 Fresnel 反射已达不可忽略量级，或上下层栈形成强驻波/腔增强，则必须把 $G(z, z')$ 升级为由传递矩阵法或层状 Green 张量构造的表达式，再计算 $\mathbf{C}_{\text{rad}}$ 。

从误差角度做更直接的估计：设光子晶体层厚度为 d ，层内折射率为 n_{PC} ，相邻层折射率为 n_{adj} ，则界面 Fresnel 反射系数约为 $r \sim (n_{\text{PC}} - n_{\text{adj}})/(n_{\text{PC}} + n_{\text{adj}})$ 。若 $|r| \lesssim 0.05$ （即 $|\Delta n|d/\lambda \lesssim 0.1$ ，其中 λ 为工作波长），则界面反射引入的修正通常在 $\mathbf{C}_{\text{rad}}$ 的 5% 以内（[工程经验] 阈值），该近似可接受；当 $|r| \gtrsim 0.2$ （如深刻蚀空气孔与高折射率包层的界面），则必须采用层状 Green 张量或 TMM 构造的严格 Green 函数，否则辐射损耗常数可能出现系统性偏差。

利用 Green 函数，辐射波场可直接用基本波振幅表示：

$$E_x(z) = k_0^2 (\xi_{0,-1} R_y + \xi_{0,1} S_y) \int_{\text{PC}} G(z, z') \phi_0(z') dz', \quad (11.23)$$

$$E_y(z) = k_0^2 (\xi_{-1,0} R_x + \xi_{1,0} S_x) \int_{\text{PC}} G(z, z') \phi_0(z') dz'. \quad (11.24)$$

式 (11.23) 和式 (11.24) 的物理含义非常清楚：辐射波的 E_x 分量由沿 $\pm y$ 传播的基本波 R_y, S_y 通过一阶衍射产生， E_y 分量由沿 $\pm x$ 传播的基本波 R_x, S_x 产生。Green 函数积分则把光子晶体层内的源分布转化为整个纵向空间中的辐射场。

将式 (11.24) 代入 R_x 的耦合方程式 (11.16) 中的辐射波项，再乘以 $\phi_0(z)$ 并沿 z 积分，辐射耦合矩阵元为

$$\mu^{(r,s)} = -\frac{k_0^4}{2\beta_0} \xi_{p,q} \xi_{-r,-s} \int_{\text{PC}} \int_{\text{PC}} G(z, z') \phi_0(z') \phi_0(z) dz' dz, \quad (11.25)$$

其中 (p, q) 和 (r, s) 分别对应源波和目标波的 Fourier 分量指标。

为避免索引歧义，更稳妥的写法是把一般矩阵元写成

$$[\mathbf{C}_{\text{rad}}]_{(r,s) \leftarrow (p,q)} = -\frac{k_0^4}{2\beta_0} \xi_{p,q} \xi_{-r,-s} I_{\text{rad}}, \quad I_{\text{rad}} = \int_{\text{PC}} \int_{\text{PC}} G(z, z') \phi_0(z') \phi_0(z) dz' dz. \quad (11.26)$$

这里列指标 (p, q) 表示源基本波，行指标 (r, s) 表示目标基本波；两者都按

$$(1, 0), (-1, 0), (0, 1), (0, -1)$$

这一顺序与 R_x, S_x, R_y, S_y 一一对应。于是， $\xi_{p,q}$ 负责“源波 $\rightarrow$ 辐射通道”的耦合， $\xi_{-r,-s}$ 负责“辐射通道 $\rightarrow$ 目标波”的回馈。前面写成 $\mu^{(r,s)}$ 的简记，只是把在方格晶格对称下会重复出现的条目压缩成较短记号。

11.4.5 高阶波的解析求解：线性组合技巧

高阶波对应 $m^2 + n^2 > 1$ 的所有分量。直接求解式 (11.8) 和式 (11.9) 比较困难，因为 $E_{x,mn}$ 和 $E_{y,mn}$ 通过 $mn\beta_0^2$ 项耦合在一起。Liang 等人 [41] 引入了一个关键的线性组合技巧：

定义

$$E_+^{(m,n)}(z) \equiv mE_{x,mn}(z) + nE_{y,mn}(z), \quad E_-^{(m,n)}(z) \equiv nE_{x,mn}(z) - mE_{y,mn}(z). \quad (11.27)$$

这一变换把原来的耦合方程组解耦为两个独立方程:

$$\left[\frac{d^2}{dz^2} + k_0^2 n_0^2 \right] E_+^{(m,n)} = -k_0^2 \sum_{\substack{m',n' \\ (m',n') \neq (m,n)}} \xi_{m-m',n-n'} E_+^{(m',n')}, \quad (11.28)$$

$$\left[\frac{d^2}{dz^2} + k_0^2 n_0^2 - (m^2 + n^2) \beta_0^2 \right] E_-^{(m,n)} = -k_0^2 \sum_{\substack{m',n' \\ (m',n') \neq (m,n)}} \xi_{m-m',n-n'} E_-^{(m',n')}. \quad (11.29)$$

同时, 横向性约束 式 (11.10) 变为

$$\frac{d}{dz} E_+^{(m,n)} = 0, \quad (11.30)$$

这意味着 $E_+^{(m,n)}$ 在 z 方向上是常数。将 式 (11.30) 代入 式 (11.28), 可得到一个纯代数关系

$$n_0^2(z) E_+^{(m,n)} = - \sum_{\substack{m',n' \\ (m',n') \neq (m,n)}} \xi_{m-m',n-n'} E_+^{(m',n')}, \quad (11.31)$$

这等价于横向性约束 $\nabla \cdot (\varepsilon \mathbf{E}) = 0$ 的 Fourier 空间表达。

真正需要用 Green 函数求解的只有 $E_-^{(m,n)}$ 。由于 $m^2 + n^2 > 1$, 式 (11.29) 左侧的有效纵向波数为

$$k_{z,mn} = i\sqrt{(m^2 + n^2)\beta_0^2 - k_0^2 n_0^2} \equiv i\kappa_{z,mn}, \quad (11.32)$$

其中 $\kappa_{z,mn} > 0$, 因此高阶波在纵向为倏逝波。对应的 Green 函数满足

$$\left[\frac{d^2}{dz^2} + k_0^2 n_0^2 - (m^2 + n^2) \beta_0^2 \right] G_{m,n}(z, z') = -\delta(z - z'), \quad (11.33)$$

解为

$$G_{m,n}(z, z') = \frac{1}{2\kappa_{z,mn}} e^{-\kappa_{z,mn}|z-z'|}. \quad (11.34)$$

与辐射波 Green 函数 式 (11.22) 的振荡行为不同, $G_{m,n}$ 是实数且指数衰减的, 这直接反映了高阶波被束缚在光子晶体层附近的物理事实。

在只保留基本波作为源的近似下, $E_-^{(m,n)}$ 可用 Green 函数表示为

$$E_-^{(m,n)}(z) = k_0^2 (-m\xi_{m-1,n}R_x - m\xi_{m+1,n}S_x + n\xi_{m,n-1}R_y + n\xi_{m,n+1}S_y) \int_{\text{PC}} G_{m,n}(z, z') \phi_0(z') dz'. \quad (11.35)$$

结合 式 (11.31) 给出的 $E_+^{(m,n)}$, 可通过逆变换恢复原始分量:

$$\begin{bmatrix} E_{x,mn}(z) \\ E_{y,mn}(z) \end{bmatrix} = \frac{1}{m^2 + n^2} \begin{bmatrix} m & n \\ n & -m \end{bmatrix} \begin{bmatrix} E_+^{(m,n)}(z) \\ E_-^{(m,n)}(z) \end{bmatrix}. \quad (11.36)$$

至此, 所有高阶波场都已用基本波振幅 R_x, S_x, R_y, S_y 解析表示。

11.4.6 三个子矩阵的显式矩阵形式

将辐射波和高阶波的解析解代回耦合方程式 (11.16)，并对其余三个基本波做完全类似的处理，最终得到闭合的 4×4 本征问题。三个子矩阵的显式形式为 [41]：

C_{1D} ：一维反馈耦合矩阵。

$$C_{1D} = \begin{bmatrix} 0 & \kappa_{2,0} & 0 & 0 \\ \kappa_{-2,0} & 0 & 0 & 0 \\ 0 & 0 & 0 & \kappa_{0,2} \\ 0 & 0 & \kappa_{0,-2} & 0 \end{bmatrix}, \quad (11.37)$$

其中一维耦合系数为

$$\kappa_{\pm 2,0} = -\frac{k_0^2}{2\beta} \int_{PC} \xi_{\pm 2,0}(z) |\phi_0(z)|^2 dz, \quad \kappa_{0,\pm 2} = -\frac{k_0^2}{2\beta} \int_{PC} \xi_{0,\pm 2}(z) |\phi_0(z)|^2 dz. \quad (11.38)$$

C_{1D} 的结构非常直观：它只连接沿同一坐标轴的前后向波对 ($R_x \leftrightarrow S_x$, $R_y \leftrightarrow S_y$)，不涉及 x - y 之间的交叉耦合。这与一维 DFB 激光器的反馈耦合完全类比，只是 Fourier 分量取的是二阶倒格矢 $\xi_{\pm 2,0}$ 和 $\xi_{0,\pm 2}$ 。对方格晶格， $\kappa_{2,0} = \kappa_{0,2}$ 。 C_{1D} 是厄米矩阵。

C_{rad} ：辐射耦合矩阵。

$$C_{rad} = \begin{bmatrix} \mu^{(1,0)} & \mu^{(-1,0)} & 0 & 0 \\ -\mu^{(1,0)} & -\mu^{(-1,0)} & 0 & 0 \\ 0 & 0 & \mu^{(0,1)} & \mu^{(0,-1)} \\ 0 & 0 & -\mu^{(0,1)} & -\mu^{(0,-1)} \end{bmatrix}, \quad (11.39)$$

其中辐射耦合系数为

$$\mu^{(r,s)} = -\frac{k_0^4}{2\beta_0} \xi_{p,q} \xi_{-r,-s} \int_{PC} \int_{PC} G(z, z') \phi_0(z') \phi_0(z) dz' dz. \quad (11.40)$$

C_{rad} 有两个关键特征。第一，它是**非厄米矩阵**——这也是辐射损耗的数学来源。Green 函数 $G(z, z')$ 的虚部编码了能量向自由空间的不可逆流出，使得 $\mu^{(r,s)}$ 为复数，其虚部直接给出辐射常数 α_{rad} 。第二， C_{rad} 具有特殊的反对称结构：同一列的上下两个元素符号相反。这意味着沿同一轴的前后向波对（如 R_x 和 S_x ）从辐射通道获得的耦合恰好反号，这也是后面讨论干涉抵消和 BIC 的矩阵语言基础。

C_{2D} ：二维高阶耦合矩阵。

$$C_{2D} = \begin{bmatrix} \chi_y^{(1,0)} & \chi_y^{(-1,0)} & \chi_y^{(0,1)} & \chi_y^{(0,-1)} \\ \chi_y^{(1,0)} & \chi_y^{(-1,0)} & \chi_y^{(0,1)} & \chi_y^{(0,-1)} \\ \chi_x^{(1,0)} & \chi_x^{(-1,0)} & \chi_x^{(0,1)} & \chi_x^{(0,-1)} \\ \chi_x^{(1,0)} & \chi_x^{(-1,0)} & \chi_x^{(0,1)} & \chi_x^{(0,-1)} \end{bmatrix}, \quad (11.41)$$

其中

$$\chi_i^{(r,s)} = -\frac{k_0^2}{2\beta} \sum_{m^2+n^2>1} \left[\zeta_i^{(m,n;r,s)} + \eta_i^{(m,n;r,s)} \right], \quad (11.42)$$

$$\zeta_i^{(m,n;r,s)} = \xi_{p-m,q-n} \xi_{m-r,n-s} \int_{\text{PC}} \int_{\text{PC}} G_{m,n}(z, z') \phi_0(z') \phi_0(z) dz' dz, \quad (11.43)$$

$$\eta_i^{(m,n;r,s)} = -\frac{1}{n_0^2(z)} \xi_{p-m,q-n} \xi_{m-r,n-s} \int_{\text{PC}} |\phi_0(z)|^2 dz. \quad (11.44)$$

这里 ζ 项来自 $E_-^{(m,n)}$ (通过 Green 函数 $G_{m,n}$ 求解), η 项来自 $E_+^{(m,n)}$ (通过横向性约束的代数关系求解)。 $\mathbf{C}_{2D}$ 是厄米矩阵, 因为高阶波的 Green 函数 $G_{m,n}$ 是实数 (倏逝波不携带能量流出)。它主要影响模式频率和分裂, 而不直接贡献辐射损耗。

η 项之所以会带出 $1/n_0^2(z)$, 可以按下面的最小步骤理解。由式 (11.30) 和式 (11.31), $E_+^{(m,n)}$ 先被约束成一个代数消元变量, 其局域幅度正比于 $n_0^{-2}(z)$ 乘以由 Fourier 分量驱动的源项。再把这个 $E_+^{(m,n)}$ 代回目标基本波的投影积分, 就会得到

$$\int_{\text{PC}} \phi_0(z) \frac{1}{n_0^2(z)} \phi_0(z) dz,$$

也就是式 (11.44) 中的标量权重。换句话说, η 不是额外假设出来的项, 而是“先用横向性约束消去 E_+ , 再做纵向投影”后必然出现的局域权重。

对圆孔等高对称孔形, 低阶截断 ($|m|, |n| \leq 3$) 通常已足够使频率收敛; 但对非对称孔形 (如直角等腰三角形), 高阶 Fourier 分量不可忽略, 需截断到 $|m|, |n| \leq 10$ 才能使辐射常数收敛 [41]。

三个子矩阵的厄米性质直接对应物理: $\mathbf{C}_{1D}$ 和 $\mathbf{C}_{2D}$ 是厄米的, 因为面内反馈和倏逝波耦合都不携带能量离开系统; $\mathbf{C}_{\text{rad}}$ 是非厄米的, 因为辐射波把能量带到了自由空间。这一区分正是三维 CWT 能同时给出频率排序和辐射损耗的根本原因。

11.4.7 从四波耦合方程到完整本征问题

将上述三个子矩阵合并, 耦合方程式 (11.16) 及其类似方程最终化为闭合的 4×4 本征问题

$$(\delta + i\alpha)\mathbf{V} = \mathbf{C}\mathbf{V}, \quad \mathbf{C} = \mathbf{C}_{1D} + \mathbf{C}_{\text{rad}} + \mathbf{C}_{2D}, \quad (11.45)$$

其中所有矩阵元都已由光子晶体几何 (通过 $\xi_{m,n}$) 和多层波导结构 (通过 $\phi_0(z)$ 、 $G(z, z')$ 、 $G_{m,n}(z, z')$) 解析确定。因此, 一旦给定孔形、填充比、晶格常数和外延层结构, $\mathbf{C}$ 的每一个元素都可通过数值积分计算, 原则上不需要经验拟合参数。

求解式 (11.45) 的计算代价极低: 它只是一个 4×4 复矩阵的本征值分解, 在普通计算机上约需 1 秒即可完成一组参数的全部四个模式。作为对比, 同等精度的 3D FDTD 仿真通常需要约 4 小时 [41]。这一量级差异正是 CWT 在参数扫描和设计优化中不可替代的原因。

11.4.8 三维 CWT 的核心输出与二维模型的本质差异

求解式 (11.45) 的本征值 $\lambda_j = \delta_j + i\alpha_j$ 和本征向量 $\mathbf{V}_j$, 可同时得到:

1. **模式频移**：由 $\text{Re}(\lambda_j) = \delta_j$ 给出，对应 A/B/C/D 四模的频率排序；
2. **辐射常数**：由 $\text{Im}(\lambda_j) = \alpha_j$ 给出；按功率衰减定义有 $\alpha_{\text{rad},j} = 2\alpha_j$ ，直接量化面发射损耗；
3. **面内场型与偏振**：由本征向量 $\mathbf{V}_j$ 给出，可判断模式的 irrep 归属和远场偏振；
4. **纵向场剖面**：基本波、辐射波和高阶波各自的 z 方向场分布。

与二维 CWT 相比，三维 CWT 的本质进步在于三点 [41]：

- 辐射耦合系数 C_{rad} 从经验拟合参数变为解析可算量；
- 不同波矢分量的纵向场剖面被分别处理，而非统一近似为 $\phi_0(z)$ ；
- 高阶 Fourier 分量被系统纳入，使非对称孔形的建模成为可能。

11.4.9 为什么辐射波与高阶波不能共用基本波的纵向场型

如果只停留在二维有效折射率图像，人们很容易默认所有相关波分量共用同一个 $\phi_0(z)$ 。这在真实 PCSEL 中通常并不成立。三类波在纵向上的物理图像应明确区分：

1. **基本波**的场型接近背景层状波导的基模，峰值常位于有源层附近，并在上下包层中缓慢衰减；
2. **辐射波**在 z 方向上呈振荡形式，并在光子晶体层上下都保持非零幅度，它对应从腔内泄露到自由空间的能量；
3. **高阶波**通常比基本波更强地局域在光子晶体层内，并在层外呈倏逝衰减；波矢阶数越高，局域往往越强。

这也是三维 CWT 相比二维模型最根本的物理进步之一。因为一旦三类波的纵向场型不同，它们与有源区、损耗层、空气/衬底辐射通道的重叠就会不同，从而直接改写

$$\Gamma_{\text{QW}}, \quad \Gamma_{\text{loss}}, \quad \gamma_{\uparrow}, \quad \gamma_{\downarrow}, \quad (11.46)$$

进而改写阈值排序、上下出光比和远场。因此，真实三维结构中的耦合问题不能简单还原为纯二维矩阵；原因是不同波分量对应不同的 z 向场型、边界条件和重叠因子。

式 (11.45) 仍然是无限周期近似下的结果。处理有限孔阵的边缘效应时，需要把四个常数振幅升级为空间缓变包络：

$$R_x(\rho), \quad S_x(\rho), \quad R_y(\rho), \quad S_y(\rho),$$

并为它们加入有限孔阵的边界条件。本章后文将给出对应的有限尺寸 CWT 包络形式。

11.4.10 圆孔与非对称孔形：干涉抵消与辐射打开

三维 CWT 的一个最重要物理预测，涉及孔形对辐射常数的影响。对圆形孔（ C_{4v} 高对称孔），模式 A 和 B 在精确 Γ 点的辐射常数为零 $\alpha_{\text{rad}} = 0$ ；而对非对称孔形（如等边三角形、直角等腰三角形），辐射常数随填充比增大而增大 [41]。

物理机制可用纵向干涉图像来理解：圆孔情形下，沿 $+x$ 与 $-x$ 传播的两支反向基本波本征上处于相位差为 π 的状态，它们向法向衍射的分量在纵向方向相消干涉，净辐射振幅为零——这也是第 9 章所述 symmetry-protected BIC 的具体物理来源。当孔形变为非对称时，反向基本波之间不再严格保持 π 相位差，干涉相消被部分破坏，辐射通道被打开，模式从 BIC 转变为 quasi-BIC [41]。

用公式语言来表达：辐射耦合矩阵 $\mathbf{C}_{\text{rad}}$ 的虚部正比于辐射振幅的平方，

$$\alpha_{\text{rad}} = 2 |\text{Im}[\lambda(\mathbf{C}_{\text{rad}})]| \propto \left| \int_{\text{PC}} \xi_{1,0}(z) G(z, z') \xi_{-1,0}(z') \phi_0(z) \phi_0(z') dz dz' \right|^2. \quad (11.47)$$

对圆孔，实空间积分因为高对称性导致两项相消， $\alpha_{\text{rad}} = 0$ ；对三角孔，对称性破缺使积分非零。这一预测与 3D FDTD 定量吻合，验证了三维 CWT 能正确描述面发射的物理机制。

该平方关系是弱开口极限下的标准结果，而非经验假设。设某个 dark family 在无辐射极限的右本征矢为 $\mathbf{u}_0$ ，则由辐射耦合矩阵引起的一阶本征值修正可写成

$$\delta \lambda_{\text{rad}} \sim (\mathbf{u}_0^L)^\dagger \mathbf{C}_{\text{rad}} \mathbf{u}_0^R.$$

而 $\mathbf{C}_{\text{rad}}$ 本身正比于“源波 $\rightarrow$ 辐射通道 $\rightarrow$ 目标波”的二次过程，因此其反厄米部分天然就是辐射振幅乘其共轭的形式。于是

$$\alpha_{\text{rad}} \propto 2 |\text{Im} \delta \lambda_{\text{rad}}| \propto |a_{\text{rad}}|^2.$$

当对称性破缺使 $a_{\text{rad}} \propto \delta$ 时，便立刻得到 $\alpha_{\text{rad}} \propto \delta^2$ 与 quasi-BIC 的平方定律。

以方格晶格圆孔（ $n_b = 3.57$ ， $n_{\text{air}} = 1$ ， $f = 12\%$ ， $a = 295 \text{ nm}$ ）为例，三维 CWT 给出模式 A 和 B 在精确 Γ 点的辐射常数 $\alpha_{\text{rad}} \approx 0$ ，对应 $Q \rightarrow \infty$ ，与 3D FDTD 结果吻合；而对同等参数的直角等腰三角形孔，辐射常数约为 $50\text{--}200 \text{ cm}^{-1}$ ， Q 有限可设计 [41]。波截断阶数需取 $D \geq 10$ （即保留 $(2D+1)^2 = 441$ 个平面波分量）才能使辐射常数收敛，频率收敛则 $D \geq 5$ 已足够。

11.4.11 品质因子与辐射常数的提取

从式 (11.45) 的本征值可直接提取器件最关心的两个量。设第 j 个本征值为 $\lambda_j = \delta_j + i\alpha_j$ ，并定义复传播常数 $\tilde{\beta}_j = \beta_0 + \lambda_j$ 。若进一步换算到本书统一时间因子 $e^{-i\omega t}$ 下的复频率记号，则

$$\tilde{\omega}_j = \omega_{0,j} - i\gamma_j, \quad \gamma_j = v_{g,j}\alpha_j, \quad (11.48)$$

其中 $v_{g,j}$ 是该模对应的群速度。品质因子由复频率的实部与虚部之比给出：

$$Q_j = \frac{\text{Re}(\tilde{\omega}_j)}{2|\text{Im}(\tilde{\omega}_j)|}. \quad (11.49)$$

辐射常数 (radiation constant) 则定义为单位传播长度上因垂直辐射而损失的功率比例：

$$\alpha_{\text{rad},j} = 2\alpha_j = 2\text{Im}(\lambda_j) = \frac{2\gamma_j}{v_{g,j}} = -\frac{2\text{Im}(\tilde{\omega}_j)}{v_{g,j}}. \quad (11.50)$$

这个定义的优势在于它不依赖等效折射率的模糊定义—— λ_j 直接来自本征值，可与 3D FDTD 的衰减率提取结果做无歧义比较 [41]。

11.4.12 非 Γ 点扩展：能带结构计算

前面的推导都限于精确 Γ 点 ($\mathbf{k} = \mathbf{0}$)。但要计算完整的能带结构，必须把模型扩展到 Γ 点附近的有限波矢 $\delta\mathbf{k}$ 。Liang 等人 [41] 给出了这一扩展的框架。

基本思路是：当工作点偏离 Γ 点一个小波矢 $\delta\mathbf{k} = (\delta k_x, \delta k_y)$ 时，四个基本波的传播方向不再严格沿 $\pm x$ 和 $\pm y$ ，而是各自偏转一个小角度。具体地，四个基本波的波矢变为

$$\begin{aligned} \beta_1 &= (\beta_0 + \delta k_x, \delta k_y), \\ \beta_2 &= (-\beta_0 + \delta k_x, \delta k_y), \\ \beta_3 &= (\delta k_x, \beta_0 + \delta k_y), \\ \beta_4 &= (\delta k_x, -\beta_0 + \delta k_y). \end{aligned} \quad (11.51)$$

由于波矢偏转，每个基本波的 TE 极化方向也随之旋转。设 (m_x, n_y) 为第 (m, n) 个基本波的实际面内波矢分量，则电场不再严格只有一个分量，而是同时具有 E_x 和 E_y ，其比例由偏振方向决定：

$$\frac{E_{x,mn}}{E_{y,mn}} = -\frac{n_y}{m_x}. \quad (11.52)$$

这一修正使得原本在 Γ 点严格解耦的 x -like 和 y -like 偏振在偏离 Γ 点后发生混合。

将修正后的波矢和偏振代入耦合方程，本征问题的形式不变，但耦合矩阵 $\mathbf{C}$ 的每个元素都变成 $\delta\mathbf{k}$ 的函数。沿布里渊区高对称路径 $\Gamma \rightarrow X \rightarrow M \rightarrow \Gamma$ 扫描 $\delta\mathbf{k}$ ，即可得到完整的能带结构 $\omega(\mathbf{k})$ ，包括每条带的辐射常数随波矢的变化。

对方格晶格圆孔 ($f = 12\%$, $\varepsilon_b = 12.7449$, $a = 295 \text{ nm}$)，三维 CWT 计算的能带结构与 3D FDTD 结果在频率上吻合到 $< 1\%$ 。在精确 Γ 点，模式 A 和 B 的辐射常数为零 (BIC)，而偏离 Γ 点后辐射常数迅速增大——这也是 BIC 在动量空间中的特征行为。模式 C 和 D 在 Γ 点即有较大辐射常数，偏离后变化相对平缓。整条能带的计算 (约 100 个 $\mathbf{k}$ 点) 在三维 CWT 中仅需约 2 分钟，而等效的 3D FDTD 需要数百小时 [41]。

对 PCSEL 而言， Γ 点是第一优先检查对象，但不是唯一对象。真正可信的候选模筛选，应当同时检查 exact Γ 及其邻域一小片倒空间中的近平带分支。

11.5 辐射通道为什么在三维 CWT 中必须显式处理

二维最小模型通常把辐射损耗压缩为一个或几个非厄米参数，这在许多前期分析中足够。升级到三维 CWT 后，辐射应被视为若干开放通道的集合，而不是单一数值。

更具体地说，向上辐射和向下辐射可以具有不同的耦合强度、不同的角谱分布，甚至不同的偏振选择规则。若进一步考虑封装、反射层或衬底厚度，外部通道对器件的反馈也可能不再可以忽略。于是，更一般的三维 CWT 往往会把辐射通道视为一组可消去的连续自由度：先写出内腔模与外部辐射模的联合方程，再通过消元把外部通道折算成等效的频移和损耗矩阵。

在形式上，这一步可以压缩为

$$\mathbf{M}_{\text{eff}}(\omega) = \mathbf{M}_{\text{int}}(\omega) + {}_{\text{rad}}(\omega), \quad (11.53)$$

其中 $\mathbf{M}_{\text{int}}$ 描述内部耦合， ${}_{\text{rad}}$ 则是由开放辐射通道引入的自能项（self-energy）。在工程上，不必执着于这个名称本身，更重要的是明白：辐射损耗和频率拉移往往来自同一开放通道，它们不是彼此独立的两个现象。

因此，在某些器件中，小幅外延参数变化会同时改变辐射 Q 、远场主瓣和偏振稳定性；它改变的是开放通道的耦合结构，而不只是单一损耗常数。

11.6 有限尺寸 CWT：从常数振幅到包络方程

无限周期 CWT 给出的 R_x, S_x, R_y, S_y 是常数振幅；但真实器件总是有限尺寸，因此更接近器件语言的写法应为

$$R_x(x, y), \quad S_x(x, y), \quad R_y(x, y), \quad S_y(x, y). \quad (11.54)$$

在最小包络近似下，可把耦合波方程写成

$$(\delta + i\alpha)\mathbf{V} = \mathbf{C}\mathbf{V} + i\mathbf{D}_x\partial_x\mathbf{V} + i\mathbf{D}_y\partial_y\mathbf{V}, \quad (11.55)$$

其中

$$\mathbf{V} = \begin{bmatrix} R_x & S_x & R_y & S_y \end{bmatrix}^T, \quad (11.56)$$

$\mathbf{D}_x, \mathbf{D}_y$ 是把 exact Γ 点近邻色散斜率压缩后的群速度矩阵。它们描述包络沿横向如何传播、扩展并在边缘处与边界条件相互作用。

若把无限周期问题的约化矩阵写成

$$\mathbf{C}(\delta\mathbf{k}) = \mathbf{C}_0 + \delta k_x \mathbf{D}_x + \delta k_y \mathbf{D}_y + \mathcal{O}(|\delta\mathbf{k}|^2), \quad (11.57)$$

则

$$\mathbf{D}_x = \left. \frac{\partial \mathbf{C}}{\partial k_x} \right|_{\Gamma}, \quad \mathbf{D}_y = \left. \frac{\partial \mathbf{C}}{\partial k_y} \right|_{\Gamma}. \quad (11.58)$$

进一步，对某个已选定的超模 j ，其包络沿 x/y 方向看到的有效群速度可写成

$$v_{g,x}^{(j)} = \frac{(\mathbf{u}_j^L)^\dagger \mathbf{D}_x \mathbf{u}_j^R}{(\mathbf{u}_j^L)^\dagger \mathbf{u}_j^R}, \quad v_{g,y}^{(j)} = \frac{(\mathbf{u}_j^L)^\dagger \mathbf{D}_y \mathbf{u}_j^R}{(\mathbf{u}_j^L)^\dagger \mathbf{u}_j^R}. \quad (11.59)$$

因此 $\mathbf{D}_x, \mathbf{D}_y$ 不是额外拍脑袋引入的“速度矩阵”，而是 exact Γ 附近色散雅可比在有限维子空间中的线性化结果。

例子

2×2 子空间的 D_x 具体形式。为说明 D_x 如何从色散导出，以方格晶格四波子空间中的 E doublet 子块为例。设 Γ 点附近 $C(\delta k_x)$ 在该子块内可写成

$$\tilde{C}(\delta k_x) = \begin{bmatrix} 0 & \kappa_2 \\ \kappa_2 & 0 \end{bmatrix} + \delta k_x \begin{bmatrix} c_{xx} & c_{xy} \\ c_{yx} & c_{yy} \end{bmatrix} + \mathcal{O}(|\delta k_x|^2),$$

其中 c_{ij} 来自 Bloch 态对 δk_x 的一阶导数。则

$$[D_x]_{E\text{-block}} = \begin{bmatrix} c_{xx} & c_{xy} \\ c_{yx} & c_{yy} \end{bmatrix}, \quad [D_y]_{E\text{-block}} = \begin{bmatrix} d_{xx} & d_{xy} \\ d_{yx} & d_{yy} \end{bmatrix},$$

其中每个元素均为 Γ 点处 $\partial C/\partial k_x$ 或 $\partial C/\partial k_y$ 的投影值。代入式 (11.59)，沿 x 方向的有效群速度为

$$v_{g,x}^{(E)} \approx \frac{u_{E,x}^* c_{xx} u_{E,x} + u_{E,x}^* c_{xy} u_{E,y} + u_{E,y}^* c_{yx} u_{E,x} + u_{E,y}^* c_{yy} u_{E,y}}{|u_{E,x}|^2 + |u_{E,y}|^2},$$

其中 $u_{E,x}, u_{E,y}$ 是该 doublet 在 E 子空间中的右本征矢量分量。对 singlet（如 A_1 或 B_1 ）， D_x, D_y 的相应子块退化为单个标量 c_{xx} 或 d_{xx} ，直接给出 $v_{g,x} \approx c_{xx}$ 。这个例子说明： D_x 和 D_y 的矩阵元是近 Γ 点色散展开的投影结果，而非独立拟合参数。

以上最小包络形式突出了群速度矩阵的物理来源。若进一步把时域演化、横向衍射以及边缘与增益的空间依赖显式写出，就得到更接近器件语言的包络方程：

$$\partial_t \mathbf{A}(\rho) + \mathbf{v}_g \cdot \nabla_{\parallel} \mathbf{A}(\rho) = iD \nabla_{\parallel}^2 \mathbf{A}(\rho) + [\mathbf{M}_0 + \mathbf{M}_{\text{rad}} + \mathbf{M}_{\text{edge}}(\rho) + \mathbf{M}_{\text{gain}}(\rho)] \mathbf{A}(\rho). \quad (11.60)$$

其中 $\mathbf{v}_g$ 是群速度矢量，描述包络的横向传播； D 是衍射系数，描述横向扩散效应。对 PCSEL 这类强反馈系统， $\mathbf{v}_g$ 通常较小，但在有限尺寸效应显著的情况下（如大孔径单模问题），传播项 $\mathbf{v}_g \cdot \nabla_{\parallel} \mathbf{A}$ 和衍射项 $iD \nabla_{\parallel}^2 \mathbf{A}$ 可能对模式竞争产生关键影响。这里的 $\mathbf{M}_{\text{edge}}(\rho)$ 用来描述边缘泄漏、吸收环或有限孔阵终止； $\mathbf{M}_{\text{gain}}(\rho)$ 则允许增益在横向上不均匀。到了这一步，三维 CWT 已经从“周期结构的低维本征模型”开始向“有限器件的半解析包络模型”过渡。

式 (11.55) 的作用是把有限阵列边缘、吸收环、局域注入窗口与横向热分布纳入同一半解析框架，而非替代全波。由此得到的输出不只是在无限周期下比较本征值，还包含有限阵列模式包络、边缘泄漏趋势和横向损耗通道。

这类有限尺寸 CWT 的优点主要有三点：

1. 比最小无限周期模型更接近真实器件；
2. 比大尺寸 3D FDTD 更适合做参数扫描与灵敏度分析；
3. 能自然输出水平损耗、有限阵列模式包络与边缘泄露趋势，为 第二十五章、第二十七章 提供更像器件的中间模型。

11.7 更高阶 CWT 到底多算出了什么

与最小 CWT 相比，三维 CWT 增加的是几类可计算的物理量。

第一类是上下辐射的差异。二维最小模型只能给一个总的辐射趋势，而三维模型可以分辨向上与向下的功率分配，这对提取效率、远场和封装接口都很重要。

第二类是纵向重叠差异。不同候选模可能与量子阱层、重掺杂层、金属邻近层具有不同的重叠因子。三维 CWT 可以更真实地给出这些差异，从而更可靠地比较净增益排序。

第三类是有限尺寸模式包络。边缘吸收、边缘泄漏、横向热透镜和注入窗口可通过包络方程以较低代价纳入，用于区分大面积单模设计中的边缘/横向非均匀问题和晶格单胞问题。

第四类是对热、电扰动的灵敏度导数。因为三维模型显式保留了层栈和模式重叠，它更适合输出诸如 $\partial\omega/\partial T$ 、 $\partial\gamma/\partial T$ 、 $\partial\omega/\partial n_{\text{eff}}$ 这类导数，供后续热-电-光耦合模型使用。

设两种外延方案具有相同的晶格常数和孔形，但上包层厚度不同。二维有效模型可能仍预测目标 Gamma 点模的频率和四波耦合几乎不变；而三维 CWT 会显示，上包层厚度变化导致纵向场峰值上移，从而同时改变了量子阱重叠、上方辐射通道耦合和接触层损耗重叠。结果是：冷腔频率变化不大，净增益排序却可能明显翻转。

这个例子说明，二维最小模型并非原则上错误，而是在某些问题上分辨率不足。

11.8 什么时候继续升级 CWT，什么时候直接转向全波

模型升级不是层数越多越好。一个常见误区是：二维最小 CWT 出现偏差后，不断加入更多参数和矩阵元，试图把它补成通用模型。这样往往会牺牲可解释性，并得到既不够准确、也不够透明的中间模型。

更实际的判断标准是看你要回答的问题属于哪一层。

如果问题是“哪个傅里叶通道在主导模式分裂”“哪类几何扰动大致会让偏振往哪个方向选”，最小 CWT 通常已经够用。

如果问题是“为什么上、下出光比和偏振稳定性对外延层这么敏感”“为什么重掺杂接触层会改变模式排序”，那就应该升到三维或更高阶 CWT。

如果问题是“有限器件的精确 Q 、远场角分布和边缘模式泄漏究竟是多少”，特别是在几何复杂、边界复杂或热电非均匀很强的情况下，继续堆 CWT 往往不划算，应当直接转向 RCWA、FEM 或 FDTD 做高保真分析。

一个很实用的经验是：当新增自由度能对应到一个明确的物理通道，并且这个通道会改变你的设计决策时，升级是值得的；如果新增自由度只是为了把曲线再拟合漂亮一点，而不改变你接下来会做什么，那通常不值得。

易错点

“高阶 CWT 总是优于低阶 CWT，因此应持续升级”的判断并不可靠。高阶模型的主要风险是参数与物理意义可能失去清晰对应，而非计算量。如果模型需要大量不可独立校准的参数才能工作，设计上往往不如直接使用高保真全波方法。

11.9 三维 CWT 与器件级模型的接口为什么更自然

从后续器件建模的角度看，三维 CWT 的一大优势是它天然输出与外延和热电模型兼容的量。最常用的接口包括：

$$\left\{ \omega_m, \gamma_{m,\uparrow}, \gamma_{m,\downarrow}, \Gamma_{\text{QW},m}, \Gamma_{\text{loss},m}, \frac{\partial \omega_m}{\partial T}, \frac{\partial \gamma_m}{\partial T} \right\}. \quad (11.61)$$

这里， ω_m 是候选模频率， $\gamma_{m,\uparrow}$ 和 $\gamma_{m,\downarrow}$ 分别表示向上、向下辐射损耗， $\Gamma_{\text{QW},m}$ 是与量子阱区的重叠因子， $\Gamma_{\text{loss},m}$ 是与损耗层的重叠因子，而温度导数则为热耦合模型提供线性接口。

这些量连接后续章节：第 16、17 章讨论如何把层栈和材料增益映射到重叠因子；第 18、19 章进一步给出这些量随电注入和温升的变化。因此，三维 CWT 是光学模型与器件物理之间的接口层，而非孤立的复杂光学模型。

回看本章

更高阶或三维 CWT 的意义是显式纳入最小模型中被压缩、但在真实器件里可能主导结果的自由度：纵向层栈、辐射通道、高阶倒格分量和有限尺寸包络，而非装饰最小模型。它仍不替代高保真全波方法，而是在保持可解释性的前提下，将模型分辨率提升到可讨论外延影响、上下辐射不对称和有限器件模式竞争的层次。是否值得升级，取决于新增自由度是否改变设计判断。

练习题

1. [基础] 给出一个二维最小 CWT 明显不够用、但三维 CWT 仍然可能有效的 PCSEL 场景，并说明原因。
2. [基础] 解释为什么式 (11.2) 中需要同时保留倒格索引 m 和纵模索引 n ，而不能只把纵向信息合并进一个有效折射率。
3. [进阶] 结合式 (11.53)，讨论为什么辐射频移和辐射损耗往往来自同一开放通道。
4. [综合]（思路提示：先定层级，再列关键式）设边缘吸收环增强后单模性改善、阈值却升高。结合式 (11.60) 解释这种权衡。

延伸阅读

- 下一章将把这里的“低维内部耦合”和“外部端口通道”联系到更一般的耦合模理论语言，从而把 PCSEL 专用 CWT 与微扰和分裂分析接起来。
- 更一般的周期介质本征方法背景，可参考 [28, 57] 中关于高阶平面波展开、对称性与开放通道的相关章节。
- 三维和高阶 PCSEL CWT 的代表性文献可参考 [41, 42, 72]。

第十二章 耦合模理论、微扰观点与模式分裂

读完本章，你能

- 区分时间耦合模理论（Temporal Coupled-Mode Theory, TCMT）、空间耦合模理论与 PCSEL 专用 CWT 的角色差异。
- 掌握微扰公式如何把“小几何变化”“小折射率漂移”“小损耗改动”转换为频率漂移和损耗变化。
- 理解模式分裂是偏振选择、模竞争和远场改变的直接来源，而非数学细节。

读前准备 建议已掌握 第八章、第十章和第十一章 中关于 CWT、开放系统和对称性破缺的基础内容。

阅读主线 本章所处模型层级：L1→L2

第十章、第十一章 给出 PCSEL 专用的耦合波语言，适合描述晶格散射网络。本章引入更通用的耦合模语言，先说明 TCMT 与 CWT 的关系，再用微扰公式描述结构小改动对频率和损耗的影响，最后将这些公式用于模式分裂与对称性破缺。

12.1 为什么在 CWT 之外还要学耦合模理论

很多读者看到这里会自然提问：既然前两章已经有了 CWT，为什么还要再学一个耦合模理论？答案在于两者关注的对象不同。

CWT 的自然基底是由倒格矢连接的内部平面波分量，用于解释光子晶体如何组织反馈和辐射。TCMT 则把结构压缩为一个或几个与端口耦合的共振模；其基底是已形成的共振超模及其输入输出通道。

因此，当问题关注孔形如何改变 x 向波和 y 向波之间的散射时，CWT 通常更自然；当问题关注某个共振模与上方自由空间、下方衬底或波导端口的耦合时，TCMT 更简洁。两者是对同一系统的不同投影，而非替代关系。

直觉框

CWT 描述腔内场如何由周期散射形成；TCMT 描述已形成的腔模如何与外界端口交换能量。前者强调内部结构，后者强调端口与通道。PCSEL 同时具有周期开放电磁系统和可测量面发射器件两种属性，因此两者都需要。

12.2 时间耦合模理论的基本语言

先考虑最简单的单模情形。若一个共振模的复振幅为 $a(t)$ ，其与外部输入输出端口耦合，则 TCMT 的标准写法是

$$\dot{a} = (i\omega_0 - \gamma_i - \gamma_e) a + \mathbf{k}^\top \mathbf{s}_+, \quad (12.1)$$

$$\mathbf{s}_- = \mathbf{C} \mathbf{s}_+ + \mathbf{d} a. \quad (12.2)$$

这里， ω_0 是共振频率， γ_i 是内部损耗， γ_e 是向外部端口泄漏的外耦合损耗， $\mathbf{s}_+$ 和 $\mathbf{s}_-$ 分别表示输入与输出通道振幅， $\mathbf{C}$ 是直接散射背景矩阵， $\mathbf{k}$ 与 $\mathbf{d}$ 表示模式与端口之间的耦合系数。

为了让这组式子真正能和 RCWA、FDTD/FEM 的输出对接，还必须先说明归一化。教材里最稳妥的约定是把 $a(t)$ 视为能量振幅，于是

$$|a|^2 \propto U_{\text{cav}}, \quad |s_{\pm}|^2 \propto P_{\text{port}}, \quad (12.3)$$

其中 U_{cav} 是腔内储能， P_{port} 是端口功率通量。于是 a 的量纲更接近 $\sqrt{J}$ ， $s_{\pm}$ 的量纲更接近 $\sqrt{W}$ ，而 $\mathbf{k}$ 、 $\mathbf{d}$ 的量纲更接近 $s^{-1/2}$ 。若换用别的归一化，例如把 $|a|^2$ 直接定义为功率，那么耦合系数的量纲和数值都会跟着变。跨章节、跨软件比较 TCMT 参数时，若不先统一这一步，单看数字大小往往没有意义。

实际从数值结果提取 TCMT 参数时，最常用的接口是被动反射/透射谱的共振峰拟合。以单端口近似为例，共振附近的反射系数可写成

$$r(\omega) \approx C_{11} + \frac{id_1 k_1}{\omega - \omega_0 + i(\gamma_i + \gamma_e)}, \quad (12.4)$$

其中极点线宽由 $\gamma_i + \gamma_e$ 决定（无显著背景干涉时对应半高全宽 $2(\gamma_i + \gamma_e)$ ），非共振背景由 C_{11} 给出。通过对 RCWA 或 FDTD 计算的谱线做拟合，可提取 γ_e （外耦合损耗）与 γ_i （内部损耗）之和；再结合被动 Q （ $Q = \omega_0/2(\gamma_i + \gamma_e)$ ）和上下出光比，可进一步分离 γ_e 的上下通道贡献。

这组式子的物理意义如下：第一式表示模幅按自身复频率演化并受外部输入驱动；第二式表示输出同时包含非共振直接散射背景和经由共振模再辐射的部分。

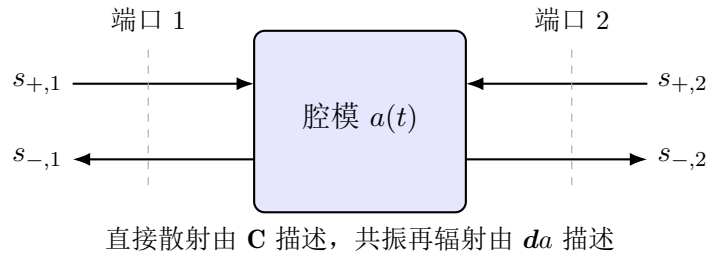


图 12.1: TCMT 的最小“端口-腔模-端口”图像。PCSEL 语境下，端口可以是上方自由空间、下方衬底辐射通道，或等效的外部耦合通道；TCMT 正是把内部共振模与这些外部端口之间的能量交换压缩成有限维输入输出模型。

对 PCSEL 来说，若我们把“上方自由空间出射”“下方衬底辐射”甚至“侧向端口”都视为外部通道，那么 TCMT 很适合描述谱线、共振峰、临界耦合和端口干涉现象。它尤其适用于分析被动散射谱和端口响应，而不要求我们显式追踪内部每一个倒格波分量。

12.3 从单模到多模：模式竞争和分裂如何写进 TCMT

PCSEL 更常见的是多模或近简并模情形。此时把单个振幅 a 升级成模矢量 $\mathbf{a}$ ，就得到

$$\dot{\mathbf{a}} = (\mathbf{i}\Omega - \mathbf{\Gamma}) \mathbf{a} + \mathbf{K} \mathbf{s}_+, \quad (12.5)$$

$$\mathbf{s}_- = \mathbf{C} \mathbf{s}_+ + \mathbf{D} \mathbf{a}. \quad (12.6)$$

其中 Ω 是模间保守耦合矩阵， $\mathbf{\Gamma}$ 是损耗与辐射矩阵。到这里你会发现，这和前两章的 CWT 形式已经非常接近。区别不在数学结构，而在基底和参数来源：

- 在 CWT 中，矩阵索引更偏向内部平面波或“平面波 $\times$ 纵模”基底；
- 在 TCMT 中，矩阵索引更偏向已经形成的共振超模以及它们与端口的关系。

因此，TCMT 与 CWT 可视为同一有限维投影思想在不同基底上的表述。实际工作中，可先用 CWT 识别重要内部耦合通道，再把候选超模参数压缩为 TCMT 参数，用于解释散射谱、反射峰、Fano 线形和端口干涉。

12.4 微扰观点为什么如此重要

在研究和设计阶段，常见问题通常是小改动：孔半径轻微变化、刻蚀深度偏差、温度升高导致折射率上升、重掺杂层吸收增加、金属位置偏移等。若每次都依赖全波重算，代价较高且不利于分析机制。微扰理论的作用，是把这些小改动映射为本征频率和损耗的一阶变化。

对于封闭或近似厄米系统，最熟悉的一阶频移公式是

$$\frac{\delta\omega}{\omega} \approx -\frac{1}{2} \frac{\int_V \delta\varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}{\int_V \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}. \quad (12.7)$$

这个公式的意义并不难：如果某个区域的介电常数增加，而该模在这个区域正好场强很强，那么模式更“愿意”把能量放进去，频率通常会向低频方向漂移。反之，若扰动发生在场很弱的区域，频移就会小得多。

式 (12.7) 的价值在于把设计动作改写为场重叠问题。这与前文 CWT 中耦合系数的结构一致：扰动是否有效，取决于它是否作用在目标模式具有显著场强或双正交重叠的区域。

式 (12.7) 严格依赖厄米归一化和小扰动条件。对强开放的 PCSEL 辐射模，真正严格的处理需要使用准正规模 (quasinormal mode) 或等价的开放系统归一化框架。工程上常把式 (12.7) 当作第一近似使用，但必须知道它在强辐射、强非厄米情况下只是一种近似，不是普适精确公式。

色散材料的微扰修正

公式12.7假设非色散材料 ($\partial\epsilon/\partial\omega \approx 0$)。对色散材料, 严格的一阶频移公式需包含色散修正项 [60, 3]:

$$\frac{\delta\omega}{\omega} \approx -\frac{1}{2} \frac{\int_V \left[\delta\epsilon + \omega \frac{\partial(\delta\epsilon)}{\partial\omega} \right] |\mathbf{E}|^2 dV}{\int_V \left[\epsilon + \omega \frac{\partial\epsilon}{\partial\omega} \right] |\mathbf{E}|^2 dV}. \quad (12.8)$$

在 PCSEL 常见的工作波长附近 (近红外), GaAs 和 InP 的材料色散 $\partial n/\partial\omega$ 贡献通常小于 10%。但对高精度建模 (如波长调谐 > 10 nm) 或宽光谱应用, 建议纳入色散修正以避免系统性偏差。

12.5 先用一个 2×2 非厄米玩具模型把左模/右模讲明白

在直接写开放 Maxwell 的一般公式之前, 先看一个最小非厄米矩阵

$$\mathbf{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad b \neq c \text{ 或 } a, d, b, c \in \mathbb{C}. \quad (12.9)$$

设其某个本征值为 λ 。对应的右本征矢满足

$$\mathbf{M}\mathbf{v}^R = \lambda\mathbf{v}^R,$$

可取

$$\mathbf{v}^R = \begin{bmatrix} b \\ \lambda - a \end{bmatrix}. \quad (12.10)$$

左本征矢则满足

$$(\mathbf{v}^L)^\top \mathbf{M} = \lambda(\mathbf{v}^L)^\top,$$

等价于 $\mathbf{M}^\top \mathbf{v}^L = \lambda\mathbf{v}^L$, 因此可取

$$\mathbf{v}^L = \begin{bmatrix} c \\ \lambda - a \end{bmatrix}. \quad (12.11)$$

对比 式 (12.10) 和式 (12.11) 可见: 除非问题退回厄米或特定 complex-symmetric 情形, 左矢量并不等于右矢量的复共轭。在非厄米问题中, 关闭投影依赖左/右本征矢的双线性配对, 而不是右本征矢自身的厄米内积。

这就是开放 Maxwell 微扰中引入伴随模的最小代数原因。左模可先理解为投影时的测试函数; 在开放系统中, $\mathbf{E}^L$ 与 $\mathbf{E}^R$ 分别由伴随 Maxwell 算子和原始 Maxwell 算子给出。

例子

数值演示: 2×2 非厄米矩阵的左右模 取 $a = 1, d = 2, b = 0.1, c = 0.3$, 则矩阵 $\mathbf{M} = \begin{bmatrix} 1 & 0.1 \\ 0.3 & 2 \end{bmatrix}$ 。本征值为

$$\lambda_{\pm} = \frac{3 \pm \sqrt{1 + 0.12}}{2} \approx \frac{3 \pm 1.058}{2}.$$

对 $\lambda_+ \approx 2.029$, 右模为 $\mathbf{v}_+^R = [0.1, \lambda_+ - 1]^T \approx [0.1, 1.029]^T$, 左模需解 $\mathbf{v}_+^{LT} \mathbf{M} = \lambda_+ \mathbf{v}_+^{LT}$, 得

到 $\mathbf{v}_+^L = [0.3, \lambda_+ - 1]^T \approx [0.3, 1.029]^T$ 。验证双正交: $\mathbf{v}_-^L \cdot \mathbf{v}_+^R = 0.3 \times 0.1 + (\lambda_- - 1)(\lambda_+ - 1) = 0$ (利用韦达定理可严格证明)。这个简单例子说明: 即使在 2×2 情况下, 左模也不等于右模的复共轭或转置。

用非厄米微扰公式估计孔形误差的影响

设设计好的圆孔 PCSEL 因工艺误差变成椭圆, 介电常数扰动为 $\delta\epsilon(\mathbf{r})$ 。若已知理想结构的右模场分布 $\mathbf{E}_n^R(\mathbf{r})$ 和左模分布 $\mathbf{E}_n^L(\mathbf{r})$, 则频率移动的一阶近似为

$$\delta\omega_n \approx -\frac{\omega_n}{2} \frac{\int \delta\epsilon \mathbf{E}_n^L \cdot \mathbf{E}_n^R dV}{\int \epsilon \mathbf{E}_n^L \cdot \mathbf{E}_n^R dV}.$$

注意分子中出现的是左模与右模的内积, 而不是通常的 $|\mathbf{E}|^2$ 。如果错误地使用右模自身做投影, 会得到错误的频移预测, 尤其在强辐射损耗模式下误差更大。

12.6 为什么非厄米一阶扰动不能直接照搬厄米公式

上一节只给出了代数例子。开放系统微扰的关键点是: 非厄米问题中, 右模本身不足以关闭微扰公式。原因不只是频率变为复数, 而是开放 Maxwell 算子不再自伴, 右模之间也不再按厄米内积正交。于是, 厄米情形默认成立的三件事会失效:

1. 归一化不再能简单用 $\int \epsilon |\mathbf{E}|^2 dV = 1$ 关闭;
2. 投影时不能默认用右模自己的复共轭来消去一阶模形修正;
3. 频移与损耗修正必须被统一看成复频率修正, 而不是先算实部、再临时补虚部。

最稳妥的写法, 是先把开放系统本征问题看成一般的非线性非厄米本征问题:

$$\hat{\mathcal{L}}(\tilde{\omega}, p) \mathbf{E}^R = 0, \quad (12.12)$$

其中 p 表示几何、材料或工作点参数。与之配套, 需要引入伴随算子与左模:

$$\hat{\mathcal{L}}^A(\tilde{\omega}, p) \mathbf{E}^L = 0, \quad (12.13)$$

这里 $\hat{\mathcal{L}}^A$ 是伴随算子, 它由“内积里的分部积分”定义:

$$\langle u, \hat{\mathcal{L}}v \rangle = \langle \hat{\mathcal{L}}^A u, v \rangle + \text{边界项} \quad (12.14)$$

直观上, 把算子从右边搬到左边, 在你选定的内积下整理分部积分项, 就得到伴随算子。在封闭厄米问题里边界项为零且 $\hat{\mathcal{L}}^A = \hat{\mathcal{L}}$, 于是左模等于右模; 在开放问题里边界项不再消失, 左模与右模必须区分。

对一般非厄米问题, 真正起投影作用的是左模与右模形成的双正交结构, 而不是右模与自身的厄米内积。可把左模先理解为“做投影时的测试函数”, 这会比“寻找另一支物理场”更直观。现在对式 (12.12) 做一阶微扰。若参数改变 $p \rightarrow p + \delta p$, 则

$$\left[\frac{\partial \hat{\mathcal{L}}}{\partial \omega} \right]_n \delta\tilde{\omega} + \delta\hat{\mathcal{L}} \mathbf{E}_n^R + \hat{\mathcal{L}}_n \delta \mathbf{E}_n^R = 0. \quad (12.15)$$

再用左模对它做投影，由于 $\mathbf{E}_n^L$ 满足伴随本征方程，最后一项被消去，于是得到开放系统中的一阶公式

$$\delta\tilde{\omega}_n = -\frac{\left\langle \mathbf{E}_n^L, \delta\hat{\mathcal{L}} \mathbf{E}_n^R \right\rangle}{\left\langle \mathbf{E}_n^L, \frac{\partial \hat{\mathcal{L}}}{\partial \omega} \bigg|_n \mathbf{E}_n^R \right\rangle}. \quad (12.16)$$

这是厄米频移公式在开放系统中的对应形式。分子仍表示扰动与模式的重叠，分母则不再是普通储能积分，而是由伴随投影和频率导数共同定义的广义归一化量。

严格表述（难度）

式 (12.16) 就是本章后文和练习题默认调用的开放系统一阶微扰总公式。如果读者只想带走一个“开放模微扰”的最小表达，应优先记住这一式，而不是把式 (12.7) 生硬推广到强开放情形。

若材料存在频率色散，这个分母最好至少写出最小形式。对

$$\hat{\mathcal{L}}(\omega) = \nabla \times \mu_r^{-1} \nabla \times - \frac{\omega^2}{c^2} \varepsilon_r(\omega)$$

这类 Maxwell 算子，有

$$\left\langle \mathbf{E}_n^L, \frac{\partial \hat{\mathcal{L}}}{\partial \omega} \mathbf{E}_n^R \right\rangle \approx -\frac{1}{c^2} \int \frac{\partial(\omega^2 \varepsilon_r(\omega))}{\partial \omega} \mathbf{E}_n^L \cdot \mathbf{E}_n^R dV. \quad (12.17)$$

若再退回无色散近似， $\partial(\omega^2 \varepsilon_r)/\partial \omega \rightarrow 2\omega \varepsilon_r$ ，分母才会进一步退化成更像“储能积分”的形式。也正因为如此，开放系统分母本质上是包含材料色散信息的广义归一化量，而非普通厄米范数。

式 (12.16) 比式 (12.7) 更一般，因为它保留了开放系统中不能省略的两点：左/右模区分，以及算子对频率的显式依赖。只要材料色散、辐射边界或 PML 已进入模型，这两点通常都不能省略。

12.7 双正交、伴随模与 QNM 归一化在这里各自扮演什么角色

到这里很容易产生另一个疑问：左模到底是不是右模的复共轭？答案是，一般并非。对真正的厄米问题，左模与右模在适当内积下确实可以重合；但对开放系统，它们只在更特殊的条件下才会简化。对互易、无磁光非互易的很多 Maxwell 问题，算子常常在某个无复共轭的双线性配对下呈现 complex-symmetric 结构，此时左模更接近“转置伙伴”，而不是“共轭伙伴”。这也是为什么 QNM 归一化中常出现 $\mathbf{E} \cdot \mathbf{E}$ 而不是 $|\mathbf{E}|^2$ 。

在实际数值中，常见的一种 QNM 广义归一化可写成

$$\langle\langle \Psi_m, \Psi_n \rangle\rangle = \int_{\Omega+\text{PML}} \left[\mathbf{E}_m \cdot \frac{\partial(\omega\varepsilon)}{\partial \omega} \mathbf{E}_n - \mathbf{H}_m \cdot \frac{\partial(\omega\mu)}{\partial \omega} \mathbf{H}_n \right] dV, \quad (12.18)$$

其中 $\Psi_n = (\mathbf{E}_n, \mathbf{H}_n)$ ，积分区域包含物理区与 PML。这个式子在正文里不要求读者立即拿出手算，但必须知道它传达的三条信息：

1. 开放系统归一化通常不再使用普通能量范数；
2. 材料色散会直接进入归一化分母，而不是只在“材料参数表”里出现；
3. 一旦归一化换了，微扰公式的分母也必须随之更换，不能继续沿用封闭腔形式。

更系统的函数空间直觉、QNM 定义和归一化讨论见 第三章、第二十八章。本章此处的目标，是让读者在使用微扰语言时，不会再把开放系统误当成“只是多了一个复数频率的封闭腔”。

12.8 开放系统中微扰不只改频率，还会改损耗

对 PCSEL 而言，只讨论频移是不够的，因为很多设计动作更直接改写的是损耗，而非共振频率本身。举例来说，轻微打破对称性可能几乎不改变两支模式的频率，却显著改变它们与辐射通道的耦合强度，于是损耗分裂比频率分裂更明显。

在开放系统里，更一般地可以把复频率写成

$$\tilde{\omega} = \omega - i\gamma,$$

因此，微扰对应复频率变化 $\delta\tilde{\omega}$ ：其实部给出频移，虚部给出衰减变化。式 (12.16) 在开放系统中输出的是统一的频移与线宽变化。模式工程不仅改变频率，也改变模式与开放通道的耦合；这解释了 第八章 中偏振选择常主要体现为损耗分裂而非频率分裂。

从设计角度看，这一结论意味着：只检查能带或共振频率分裂是不够的，还必须检查辐射损耗是否重排；后者常直接决定阈值和单模性。

12.9 哪些工程公式可以继续用，哪些地方必须升级到开放系统严格形式

到这里，读者最需要的是一条清楚的使用边界，而不是更多公式。对 PCSEL 设计，实践上可分成三层。

第一层是**可以把厄米公式当作趋势代理**的情形：模式泄露较弱、候选模远离强辐射通道重组区、扰动主要发生在腔内高场区域而不是出光边界附近、并且目标模与竞争模已经明显分离。在这种情况下，式 (12.7) 往往能较可靠地判断“往哪个方向漂移”。

第二层是**必须至少用左/右模公式检查一次**的情形：模式本身就是强开放模、设计变量直接作用于出光孔形或上下辐射不对称、材料色散不可忽略、或你已经开始关心虚部变化、偏振损耗分裂和 quasi-BIC 附近的辐射调控。此时，若仍只用式 (12.7)，最容易把“频移方向判断正确”误当成“损耗排序也正确”。

第三层是**微扰理论本身不宜作为主模型**的情形：近简并模强混合、模式身份交换、扰动显著改变阵列边界或电流孔径、热反馈导致空间分布整体重构，或系统靠近异常点（exceptional point）等强非厄米重组区域。此时应回到完整全波或多物理场本征问题重算，而不是继续扩展一阶公式。

易错点

开放系统里最常见的误判是“用对了方向判断，却错用了定量结论”，而非“公式完全错了”。尤其在 quasi-BIC、偏振分裂和热漂移问题上，单靠厄米频移公式往往能看出趋势，却未必能给出正确的损耗排序和阈值排序。

12.10 两模耦合为什么足以抓住“分裂”的核心

尽管真实 PCSEL 可能涉及多模网络，但理解模式分裂时，一个两模模型通常已经足够说明问题。设两个近简并模式的复频率分别为 $\tilde{\omega}_1$ 和 $\tilde{\omega}_2$ ，耦合强度为 κ_{12} 和 κ_{21} ，则耦合矩阵为

$$\mathbf{M} = \begin{bmatrix} \tilde{\omega}_1 & \kappa_{12} \\ \kappa_{21} & \tilde{\omega}_2 \end{bmatrix}. \quad (12.19)$$

其本征值为

$$\tilde{\omega}_{\pm} = \frac{\tilde{\omega}_1 + \tilde{\omega}_2}{2} \pm \sqrt{\kappa_{12}\kappa_{21} + \left(\frac{\tilde{\omega}_1 - \tilde{\omega}_2}{2}\right)^2}. \quad (12.20)$$

数学基础：互易性定理的适用条件与非互易机制

经典 Lorentz 互易定理表述为：对各向同性、无磁光效应的线性介质，若 $(\mathbf{E}_1, \mathbf{H}_1)$ 和 $(\mathbf{E}_2, \mathbf{H}_2)$ 是同一频率下的两组源产生的场，则

$$\int_V (\mathbf{E}_1 \cdot \mathbf{J}_2 - \mathbf{E}_2 \cdot \mathbf{J}_1) dV = \oint_S (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) \cdot d\mathbf{S}. \quad (12.21)$$

在无耗散的封闭系统中，这进一步使耦合矩阵满足 $\kappa_{12} = \kappa_{21}^*$ （厄米性）；互易性本身主要给出转置对称约束。

互易性成立的条件：

1. 介质张量对称： $\hat{\varepsilon} = \hat{\varepsilon}^T$, $\hat{\mu} = \hat{\mu}^T$ （无磁光/手性非互易）
2. 线性响应：无频率转换或非线性混合
3. 封闭系统或适当的辐射边界条件

开放系统中的有效非厄米耦合：即使介质本身满足互易性，开放系统的辐射通道也会诱导有效耗散耦合，使系统矩阵呈现非厄米结构。这是因为：

- 模式通过共同辐射通道发生间接耦合，耦合强度依赖于远场干涉条件
- 辐射损耗矩阵 $\mathbf{\Gamma}_{\text{rad}}$ 一般不是对角的，导致有效哈密顿量非厄米
- 在 biorthogonal 基底，左模与右模不构成简单的复共轭关系

这并不等同于破坏 Lorentz 互易性：在互易条件下，散射矩阵仍满足 $\mathbf{S} = \mathbf{S}^T$ （并可在合适基底写作 $\kappa_{12} = \kappa_{21}$ ）；变化的是“保守厄米”到“开放非厄米”的动力学结构。

式 (12.20) 概括了模式分裂的基本机制：当两模原本简并时，非零耦合会解除简并；若 $\tilde{\omega}_1$ 与 $\tilde{\omega}_2$ 含有不同虚部，分裂还会改写损耗排序。因此，在 PCSEL 中，分裂不仅是谱学现象，也会影响模式选择。

12.11 对称性破缺怎样导致模式分裂

现在把上述两模语言落回 第八章 的物理图像。设某个理想对称器件在 Gamma 点处支持一对简并或近简并模式，它们对应两种正交偏振或两种不同对称性表示。在完全对称时，系统对这两种模式没有偏好；一旦加入椭圆孔、双晶格扰动、上下一侧更强的辐射出口，或者热、电注入造成横向不均匀，对称性平衡就会被打破，于是出现两类可能的后果。

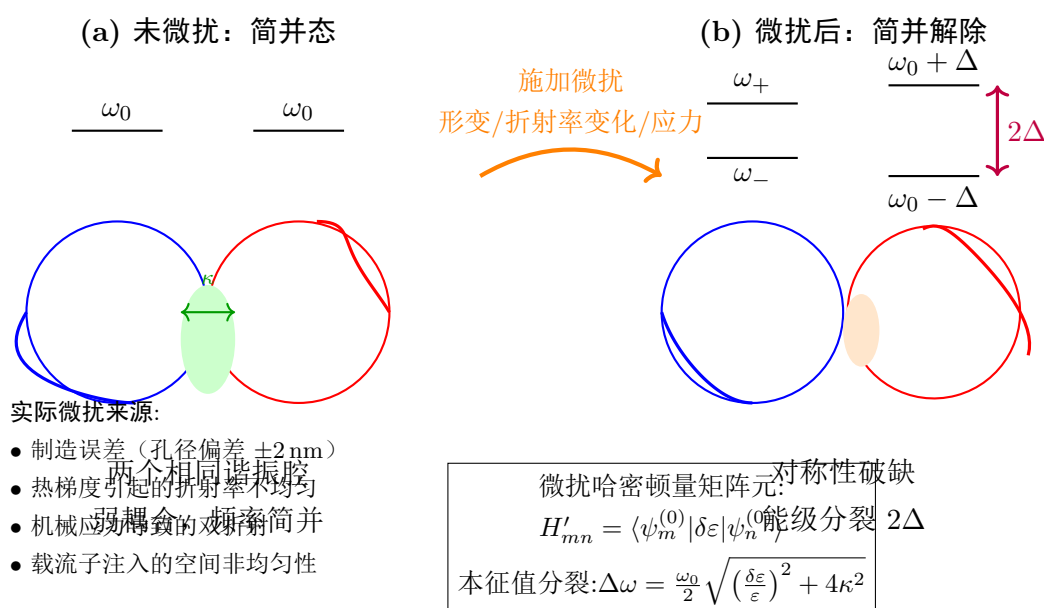


图 12.2: 耦合谐振腔系统中微扰诱导的模式分裂现象。(a) 未微扰情况：两个完全相同的环形谐振腔通过倏逝波耦合，由于系统具有镜像对称性，两个局域模式形成简并对。(b) 施加微扰后：一侧谐振腔发生轻微形变或折射率变化，对称性破缺导致简并解除，能级分裂为 $\omega_{\pm} = \omega_0 \pm \Delta$ 。底部公式给出了考虑耦合作用后的精确分裂表达式。实际 PCSEL 中的微扰来源包括制造误差、热梯度、机械应力和载流子注入的非均匀性等。

第一类后果是频率分裂：某一支模式的有效折射率或光程改变得更多，所以其共振频率偏移更大。

第二类后果是损耗分裂：某一支模式与辐射通道或损耗层的重叠增加得更多，所以其衰减更快或更慢。

在实际 PCSEL 设计中，这两类分裂常同时出现，只是主导程度不同。对于单模和偏振工程，关键在于它们如何改写净增益排序；分裂的重要性不在于谱线分开本身，而在于它决定哪一支模在阈值附近占优。

设方格晶格中一对原本简并的正交偏振 Gamma 点模，在引入轻微椭圆孔后出现分裂。若椭圆长轴沿 x 方向，那么沿 x 方向偏振的模式可能首先看到更强的保守耦合修正，于是频率

偏移；与此同时，它也可能因为近场对称性改变而更强地耦合到上方辐射通道，于是损耗也改变。最终究竟哪一支模式先起振，不能只看频率谁更接近增益峰，而要看复频率与模增益共同决定的净阈值排序。

例子

[本书算例] 把椭圆度灵敏度翻译成可测分裂。在 $\lambda_0 = 940 \text{ nm}$ 处， $f_0 = c/\lambda_0 \approx 319 \text{ THz}$ 。假设一次局部全波扫描得到：孔的 1% 椭圆度使两支偏振模产生 $\Delta f/f_0 = 5 \times 10^{-5}$ 的分裂，则

$$\Delta f \approx 16 \text{ GHz}, \quad \Delta \lambda \approx \frac{\lambda_0^2}{c} \Delta f \approx 0.047 \text{ nm}.$$

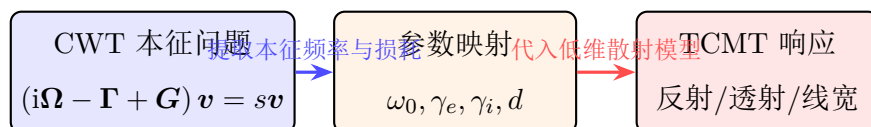
这里的 5×10^{-5} 是给定结构的模型输入，不是普适常数；算例的作用是把“轻微椭圆”变成可与光谱分辨率和增益带宽比较的数字。

这个例子说明，微扰分析和两模分裂模型可直接连接对称性破缺与器件表现。

12.12 TCMT、CWT 和微扰三者怎样配合使用

现在可以把这三套语言的分工说得更清楚。

图12.3给出了 CWT 与 TCMT 两种语言之间的映射关系：CWT 聚焦于晶格内部的倒格矢散射网络，适合回答”哪种散射通道主导模式形成”；TCMT 聚焦于共振模与外部端口的能量交换，适合回答”模式如何与自由空间、衬底或波导耦合”。



工作流：先由 CWT 得到目标模的复频率和模矢量，再把辐射损耗、内部损耗和远场归一化信息压缩成 TCMT 参数。这样可以把“微观的耦合矩阵”转成“可直接和线宽、反射谱、耦合强度比较”的解析模型。

图 12.3: CWT 与 TCMT 的语言映射。(a)CWT 视角：关注晶格内部由倒格矢驱动的散射过程，基底是行波或驻波分量，耦合系数 κ 来源于介电常数分布的 Fourier 分量；(b)TCMT 视角：关注已形成共振模与外部端口的耦合，基底是共振超模，耦合系数描述能量进出速率；(c) 两者的衔接：CWT 的特征模可以作为 TCMT 的输入，CWT 的内部散射强度可以映射为 TCMT 的有效耦合率和损耗率。

图12.3 的用途是区分建模对象：解释孔形如何影响偏振时，用 CWT 追踪散射通道；解释金属厚度如何影响阈值时，用 TCMT 追踪损耗通道；同时涉及两者时，再通过映射关系连接。

如果你的问题是”哪一种倒格矢散射通道在控制 Gamma 点模式的重组”，优先使用 CWT。如果你的问题是”某个共振模式和上方自由空间、下方衬底或旁路端口怎样干涉并形成谱

线”，优先使用 TCMT。

如果你的问题是”结构改动很小，我想先判断频率和损耗会朝哪个方向变化”，优先使用微扰公式。

三者之间通常是按问题层级配合使用，而非并列选择。常见路径是：先用 CWT 找到主导模和主导耦合；再在该模附近用微扰公式评估小改动；最后将候选参数区间送入 TCMT 或全波散射分析，检查端口响应和输出行为。这种串联方式比无约束参数扫描更便于解释机制。

把 tutorial 中的 Hermitian/非 Hermitian 控制翻译成三种语言

Noda 等人的高功率 PCSEL 路线中经常出现两个参数族：一个是由面内 180° 与 90° 衍射决定的 Hermitian 耦合，另一个是经由辐射波产生的非 Hermitian 耦合 [50, 74]。这套说法可以直接映射到本章三种语言。

在 CWT 语言中，Hermitian 耦合主要来自介电常数 Fourier 分量把四个基本 Bloch 波彼此连接；它决定 Γ 点附近模式的保守分裂、群速度曲率和有限尺寸包络。非 Hermitian 耦合则来自同一组 Bloch 波共享上方空气、下方衬底或背面反射器形成的辐射通道；它进入有效矩阵的虚部，改写辐射常数和损耗排序。

在 TCMT 语言中，前者更像共振模之间的内部耦合矩阵 Ω ，后者更像通过共同端口诱导的耗散矩阵 Γ_{rad} 。在微扰语言中，双晶格间距、孔径比例、背面 DBR 相位和 p 包层厚度都是在同时改变复频率的实部与虚部，而非单纯“调频率”的旋钮。若只追踪频率分裂而不追踪辐射常数曲率，就会漏掉大面积单模设计真正依赖的损耗裕量。

因此，本书采用的读法是：Hermitian/非 Hermitian 控制是 CWT、TCMT 与开放系统微扰在高功率 PCSEL 上的一次交汇，而非另一套独立理论。它的任务是把目标模与竞争模的复频率排序做成可设计量，而不是只把某个被动 Q 或某个耦合系数推到极值。

一个常见错误是把 TCMT 参数、CWT 参数和全波仿真输出混在一起直接比较大小，却不检查它们的归一化约定是否一致。对开放系统而言，耦合系数、模幅和端口振幅的定义如果不统一，数值上的”差一个常数”会迅速演变成物理解释上的误判。

12.13 为什么本章内容会反复回到后面的器件章节

从表面上看，微扰与模式分裂像是“纯光学细节”；但在真实 PCSEL 器件中，它们会持续被电注入和温升重新激活。温度上升引起的折射率漂移、本征增益峰漂移、自由载流子吸收上升，都会通过本章的语言转化为频移、损耗变化和模式分裂再排序。也正因为如此，后面的热-电-光耦合章节并非另起炉灶，而是把本章的线性小扰动语言变成空间分布和工作点依赖的器件语言。

因此，本章给出的是一种后续章节会反复使用的分析流程：对小改动，先判断其主要改变频率、损耗还是重叠，而非只适用于冷腔的数学技巧；对模式切换，先判断哪个简并被打破、哪个通道被重新加权。

回看本章

TCMT、CWT 与微扰理论是互补的三种语言。CWT 用于追踪光子晶体内部耦合网络，TCMT 用于描述共振模与外部端口的交换，微扰理论用于把小结构或材料改动映射为频率和损耗的一阶变化。对开放 PCSEL 问题，严格的一阶扰动应写成左/右模投影形式，而不能简单照搬厄米频移公式；厄米公式在弱开放条件下仍可作为趋势代理，但在强辐射、quasi-BIC、近简并重组或显著色散区，需要使用开放系统归一化甚至完整重算。模式分裂的重要性在于它会改写偏振、远场和阈值排序。有效设计分析应根据问题在这三种语言之间切换。

练习题

1. [基础] 用自己的话解释 TCMT 与 CWT 的区别，并说明为什么它们在 PCSEL 中都不缺。
2. [基础] 从式 (12.16) 的推导思路出发，解释为什么开放系统里必须引入左模或伴随模，而不能只用右模自己关闭投影。
3. [进阶] 结合式 (12.7) 与式 (12.16)，说明为什么“把材料改在场强最大的地方”通常最有效，但并不一定总是最优设计策略。
4. [进阶] 从式 (12.20) 出发，讨论在什么情况下损耗分裂会比频率分裂更直接地影响模式选择。
5. [综合] (思路提示：先定层级，再列关键式) 设某 PCSEL 在低温下稳定单偏振工作、升温后出现偏振切换。请用“微扰 + 模式分裂”的语言给出一条可能的解释链。

延伸阅读

- 下一章开始进入数值方法部分。PWEM 负责给出候选模目录，本章的分裂与微扰观点则解释这些候选模在后续全波分析中的频率和损耗差异。
- 一般光学腔与激光模理论可参考 [61, 10] 中关于耦合模、模竞争和频率漂移的相关章节。
- TCMT 与光子晶体 slab 共振的经典文献可参考 [13, 14, 63]。
- 开放系统归一化与非厄米微扰可继续阅读 [38, 48, 60, 34]，并与第二十八章对照。

第三部分

数值方法

第十三章 平面波展开法 (PWEM) 与光子能带计算

本章结束时，你将亲手算出图13.4：不是照着文献描一条带，而是从圆孔几何、倒格矢、Fourier 系数和矩阵截断逐项生成它，并能指出图中哪一处简并可信、哪一处还需要收敛检查。

读完本章，你能

- 从头理解平面波展开法 (Plane-Wave Expansion Method, PWEM) 为什么能用来计算光子晶体能带，以及它和第五章中能带概念之间的关系。
- 从二维理想光子晶体的标量 Maxwell 本征问题出发，推导 PWEM 的矩阵本征方程，明确每个矩阵元来自哪里、代表什么耦合。
- 学会为一个正方晶格圆形孔洞光子晶体搭建最小 PWEM 计算：定义单胞、倒格矢、高对称路径、傅里叶系数与截断规则。
- 学会读 PWEM 算出的能带图，知道它能提供哪些候选模信息，又不能替代哪些开放系统与器件级结论。

读前准备 建议已掌握第三章至第五章中关于周期 Maxwell 方程、Bloch 定理、倒格矢与能带图的概念。若对 Gamma 点、带边与候选模目录仍不熟悉，应先回看第五章。

本章结构

本章所处模型层级：L1

第五章已经把“什么是光子能带”“为什么 PCSEL 关心 Gamma 点候选模”“能带图能告诉我们什么、不能告诉我们什么”讲清楚了。但概念本身还不等于计算。在实际计算中，读者必须进一步回答：一张能带图究竟是怎样从 Maxwell 方程一步一步算出来的？在最简单的二维教学模型中，矩阵本征问题长什么样？对一个正方晶格圆孔光子晶体，倒格矢、高对称路径和圆孔的傅里叶系数该怎样写？本章就回答这些问题。本章按以下顺序展开：先从二维理想光子晶体的物理模型出发，再导出 PWEM 的代数形式，然后给出一个可复现的最小计算例子，最后解释如何读图，以及这些结果应当怎样交给第十四章、第十五章。

13.1 为什么在能带概念之后必须把它真正算出来

第5章已经强调，能带图不只是曲线展示，而是周期 Maxwell 本征谱的可视化。既然如此，研究工作就不能停留在“知道应该看能带图”这一步，还必须知道能带图是怎样从方程里长出来的。否则，读者容易落入两种相反的误区。

第一种误区是“软件主义”。即把 PWEM 当成出图工具，只关心如何让程序画出几条曲线，却不知道这些曲线代表的本征问题是什么。这样处理后，一旦几何稍有变化、极化定义稍有不同，结论就会迅速失去可解释性。

第二种误区是“纯概念主义”。即能把 Gamma 点、带边、简并这些词说得很熟，却从未亲手把几何、倒格矢、傅里叶系数和矩阵截断联系起来。这样做的直接后果是：看到一张能带图时，知道它重要，却不知道它为什么可信、哪些地方可能不收敛、也不知道该怎样把结果交给后续模型。

PWEM 是避免这两种误区的第一步。它把第5章的能带概念落实成一个可以逐项检查的矩阵本征问题。它不负责回答全部 PCSEL 问题，但它适合做一件事：在无限周期、被动、线性的框架里，把候选模目录算清楚，把高对称点附近的模式组织方式算清楚，把后续 CWT 和 RCWA 应该盯住哪些模族这件事先算清楚。

直觉框

如果说第5章是在教你“怎么看地图”，那么本章就是在教你“图的计算来源”。只有既会读图、又知道图从哪里来，后面看到 RCWA、FDTD/FEM 和器件模型时，才不会把不同层级的结果混在一起。

13.2 从二维理想光子晶体到可计算本征问题

为了把思路讲得最清楚，本章先回到一个教学型理想模型：介电常数只随平面坐标 (x, y) 周期变化，而与 z 无关。这样的二维光子晶体当然不是完整的 PCSEL slab，但它已经保留了周期性、Bloch 分类和能带形成的基本数学结构，因此适合作为 PWEM 入门模型。

13.2.1 二维模型中的极化分离为什么是教学上有价值的

在严格二维、各向同性、无磁并且介电常数只依赖 (x, y) 的情况下，Maxwell 方程可以分成两类标量问题。一类以 E_z 为唯一非零电场分量，另一类以 H_z 为唯一非零磁场分量。不同文献对这两类问题采用的 TE/TM 命名并不完全一致，因此本章尽量直接使用“ E_z 标量问题”和“ H_z 标量问题”这样的表述，避免符号标签先把读者绕晕。

之所以要先用这种理想二维模型，并非因为真实 PCSEL 就是二维标量问题，而是因为它可以最透明地展示能带如何由平面波耦合而来。等读者把这一层吃透后，再回到 slab 的矢量问题，就更容易看出哪些是保留下来的主骨架，哪些是由垂直约束、开放边界和外延层引入的新物理。

二维标量 PWEM 是“真实 PCSEL 光学问题的教学代理模型”，不是完整器件模型。它严格成立的前提是：结构在 z 方向均匀、材料线性无磁、问题可按二维极化分离。真实 PCSEL

slab 中常说的 TE-like/TM-like, 只能看成这一理想图像的延伸标签, 而不能当作完全严格的矢量分类。

13.2.2 选取 H_z 标量问题作为教学主线

对具有空气孔和高折射率背景的二维光子晶体, 常见做法是先写 H_z 标量问题。令场具有谐波时间因子 $e^{-i\omega t}$, 并设

$$\mathbf{H}(\mathbf{r}) = \hat{\mathbf{z}}H_z(x, y), \quad \mathbf{E}(\mathbf{r}) = E_x(x, y)\hat{\mathbf{x}} + E_y(x, y)\hat{\mathbf{y}}. \quad (13.1)$$

在无源条件下, 频域 Maxwell 方程可化为

$$\frac{\partial H_z}{\partial y} = -i\omega\varepsilon_0\varepsilon_r E_x, \quad (13.2)$$

$$-\frac{\partial H_z}{\partial x} = -i\omega\varepsilon_0\varepsilon_r E_y, \quad (13.3)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = i\omega\mu_0 H_z. \quad (13.4)$$

用前两式消去 E_x, E_y , 可得到仅关于 H_z 的标量本征方程

$$\nabla_{\parallel} \cdot [\eta(\mathbf{r}_{\parallel})\nabla_{\parallel} H_z(\mathbf{r}_{\parallel})] + \left(\frac{\omega}{c}\right)^2 H_z(\mathbf{r}_{\parallel}) = 0, \quad (13.5)$$

其中 $\mathbf{r}_{\parallel} = (x, y)$, $\nabla_{\parallel} = \hat{\mathbf{x}}\partial_x + \hat{\mathbf{y}}\partial_y$, 并定义

$$\eta(\mathbf{r}_{\parallel}) = \frac{1}{\varepsilon_r(\mathbf{r}_{\parallel})}. \quad (13.6)$$

式 (13.5) 就是我们后续做平面波展开的起点。

这一步的物理意义非常重要。它说明周期结构并非直接“修改了光速”, 而是通过位置相关的 $\eta = 1/\varepsilon_r$ 改变了场梯度之间的耦合方式。也正因为如此, 后面的矩阵元里才会自然出现几何傅里叶系数与波矢点积的组合。

13.2.3 同一个二维模型里还有另一类标量问题

若选择 E_z 为标量主变量, 则会得到另一种形式的本征方程

$$\nabla_{\parallel}^2 E_z(\mathbf{r}_{\parallel}) + \left(\frac{\omega}{c}\right)^2 \varepsilon_r(\mathbf{r}_{\parallel}) E_z(\mathbf{r}_{\parallel}) = 0. \quad (13.7)$$

式 (13.7) 与 式 (13.5) 的结构并不相同, 因此两类极化的能带通常也不同。很多教材在讨论带隙时会分别画出两套带图, 就是这个原因。对 PCSEL 初学者来说, 关键是理解: 二维能带本来就带有极化依赖, 且这种依赖在真实 slab 中会进一步变成更复杂的矢量模式问题, 而非一开始就记住哪一个一定叫 TE、哪一个一定叫 TM。

从数值收敛角度, H_z 标量问题 (式 (13.5)) 通常比 E_z 问题 (式 (13.7)) 收敛更快。原因在于 H_z 方程对介电常数的逆 $\eta = 1/\varepsilon_r$ 做展开, 而在高折射率对比界面处 $1/\varepsilon_r$ 的不连续性比 ε_r 本身更温和, 使傅里叶级数截断引入的吉布斯效应更小。这一经验与 Fourier factorization

和平面波收敛文献相一致 [40, 29]。因此教学上常以 H_z 问题为主线；而在折射率对比度很高（如空气孔深刻蚀、 $\Delta n/\bar{n} > 0.3$ ）时，对收敛性的检查尤为重要。

若把 E_z 同样写成平面波展开，则会得到另一类矩阵本征问题：

$$|\mathbf{k} + \mathbf{G}|^2 e(\mathbf{k} + \mathbf{G}) = \left(\frac{\omega}{c}\right)^2 \sum_{\mathbf{G}'} \varepsilon_{\mathbf{G}-\mathbf{G}'} e(\mathbf{k} + \mathbf{G}'). \quad (13.8)$$

其中 $e(\mathbf{k} + \mathbf{G})$ 表示 E_z 分量在平面波基底 $e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}_{\parallel}}$ 上的傅里叶展开系数。对读者来说，式 (13.8) 最值得记住的是一个结构事实： H_z 与 E_z 两类标量问题并不共享同一套矩阵，因此高折射率对比结构里“展开 ε 还是展开 ε^{-1} ”会直接影响收敛与带图位置，而非每个下标。这也是为什么后面的 RCWA 章节还要专门讨论 Fourier factorization 规则。

13.3 Bloch 形式与平面波展开：PWEM 的数学骨架

既然介质是周期性的，那么根据第 4 章的 Bloch 定理，任意本征解都可以写成 Bloch 形式。对式 (13.5) 而言，最自然的写法是

$$H_{z,n\mathbf{k}}(\mathbf{r}_{\parallel}) = e^{i\mathbf{k}\cdot\mathbf{r}_{\parallel}} u_{n\mathbf{k}}(\mathbf{r}_{\parallel}), \quad u_{n\mathbf{k}}(\mathbf{r}_{\parallel} + \mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r}_{\parallel}), \quad (13.9)$$

其中 $\mathbf{R}$ 是任意晶格平移矢量， $u_{n\mathbf{k}}$ 是具有晶格周期的周期函数。由于周期函数可以再作傅里叶展开，于是可写成

$$H_{z,n\mathbf{k}}(\mathbf{r}_{\parallel}) = \sum_{\mathbf{G}} h_n(\mathbf{k} + \mathbf{G}) e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}_{\parallel}}, \quad (13.10)$$

而介质函数的倒数展开为

$$\eta(\mathbf{r}_{\parallel}) = \sum_{\mathbf{G}} \eta_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}_{\parallel}}. \quad (13.11)$$

这里每一个 $\mathbf{G}$ 都是倒格矢， $h_n(\mathbf{k} + \mathbf{G})$ 是待求的平面波系数， $\eta_{\mathbf{G}}$ 则完全由单胞几何和材料参数决定。

把式 (13.10) 和式 (13.11) 代回式 (13.5)，并逐项比较具有相同指数因子的平面波系数，可得

$$\sum_{\mathbf{G}'} \eta_{\mathbf{G}-\mathbf{G}'} (\mathbf{k} + \mathbf{G}) \cdot (\mathbf{k} + \mathbf{G}') h_n(\mathbf{k} + \mathbf{G}') = \left(\frac{\omega_n(\mathbf{k})}{c}\right)^2 h_n(\mathbf{k} + \mathbf{G}). \quad (13.12)$$

这就是二维 H_z 标量问题的 PWEM 矩阵本征方程。若把所有保留下来的平面波系数排成向量 $\mathbf{h}_n$ ，则它可以简写为

$$\mathbf{A}(\mathbf{k}) \mathbf{h}_n = \left(\frac{\omega_n(\mathbf{k})}{c}\right)^2 \mathbf{h}_n. \quad (13.13)$$

13.3.1 矩阵元为什么正好长成这个样子

初学者最容易跳过的地方，就是式 (13.12) 里矩阵元的物理解释。这里至少有三层信息。

第一， $\mathbf{k} + \mathbf{G}$ 与 $\mathbf{k} + \mathbf{G}'$ 是两支不同的平面波分量，因此矩阵的行和列索引本质上是在给“参与耦合的平面波通道”编号。

第二, $\eta_{\mathbf{G}-\mathbf{G}'}$ 是介质函数倒数的傅里叶系数, 它只在几何具有相应倒格矢分量时才不为零。因此, 几何的周期性是非常具体地决定了哪些平面波之间可以通过哪一种倒格矢差值发生耦合, 而非抽象地“影响了场”。

第三, 点积

$$(\mathbf{k} + \mathbf{G}) \cdot (\mathbf{k} + \mathbf{G}') \quad (13.14)$$

反映的是梯度算符作用在平面波上之后带来的波矢权重。也就是说, PWEM 是把“介质周期性”和“场的空间变化率”一起编码进矩阵元, 而非简单把介质傅里叶系数拼成一个矩阵。为说明 PWEM 计算流程, 全书的图表汇总部分给出 PWEM 工作流程图 (见图13.1)。该图系统展示了从输入结构参数、构建介电函数傅里叶分量、组装本征矩阵、 $\mathbf{k}$ 空间采样、求解大规模稀疏矩阵本征值问题, 到最终输出能带结构和进行后处理的完整计算链条。图中还标注了计算复杂度信息 ($O(N_{\text{pw}}^3)$) 和实用建议。需要说明的是, 收敛所需的截断数没有普适常数: 本章附带的教学脚本刻意只用 25 个分量以便逐项对照, 收敛性一节的误差-截断曲线 (参考解取 $N_{\text{PW}} = 225$) 才是确定截断的正确方式。

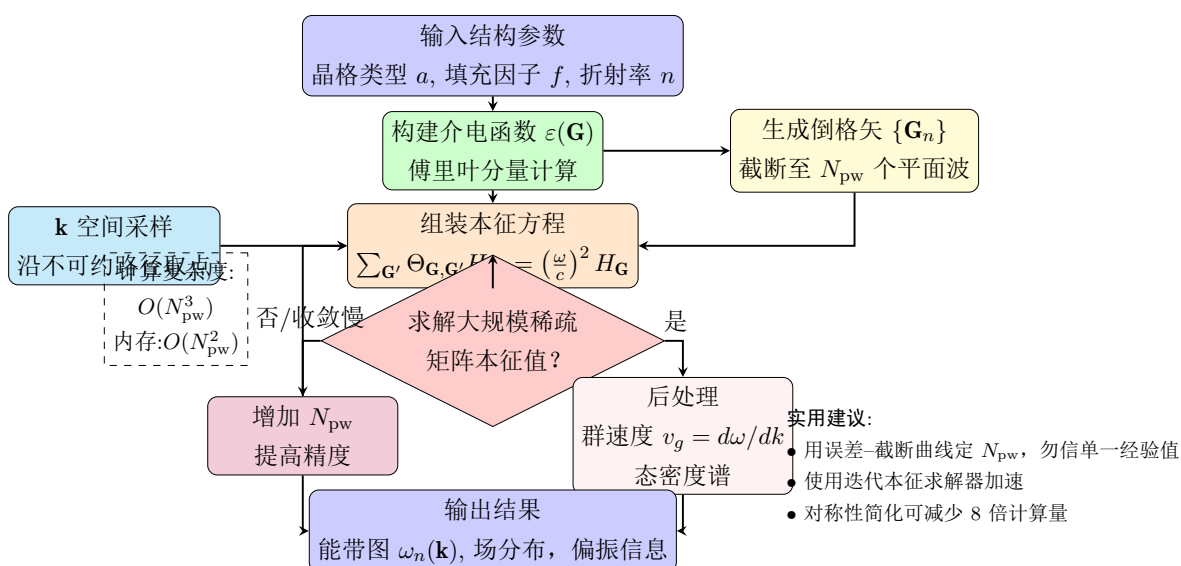


图 13.1: PWEM 的标准计算流程: 由几何与材料参数构造倒格矢和介电函数傅里叶分量, 离散 Maxwell 本征方程并沿高对称路径求解本征频率与本征矢量; 若结果未收敛, 则增大 N_{pw} 重新计算。输出的能带、模场和群速度等量可用于 PCSEL 的初始选模与参数扫描。

PWEM 的本质是“把一个模式理解成许多平面波在周期介质中彼此耦合后的自治组合”, 而非“用很多平面波去拟合一个模式”。能带就是这些耦合组合允许出现的本征频率集合。

13.4 正方晶格圆形孔洞: 把单胞几何写成傅里叶系数

现在把抽象公式落到一个最常用、也最适合入门的教学型几何上: 正方晶格空气圆孔, 背景为高折射率介质。这个几何与图5.4的示意一致。

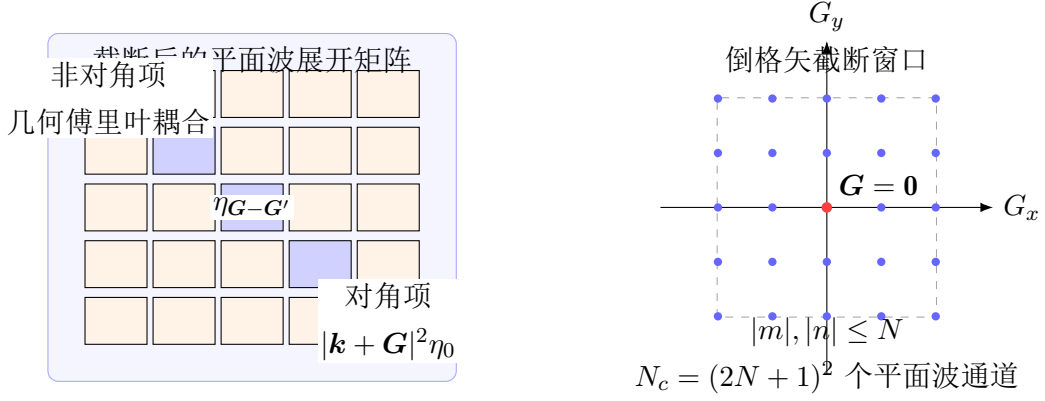


图 13.2: 平面波展开法的两个核心离散对象。左图强调矩阵的对角项主要来自“单个平面波通道的波矢权重”，非对角项主要来自介质傅里叶系数引入的通道耦合；右图强调实际求解并非保留无限多个倒格矢，而是在 $\mathbf{G}$ 空间截断出一个有限窗口。

13.4.1 几何与倒格空间定义

设正方晶格常数为 a ，单胞面积为

$$A_c = a^2. \quad (13.15)$$

实空间基矢取为

$$\mathbf{a}_1 = a\hat{\mathbf{x}}, \quad \mathbf{a}_2 = a\hat{\mathbf{y}}, \quad (13.16)$$

倒格矢基矢则为

$$\mathbf{b}_1 = \frac{2\pi}{a}\hat{\mathbf{x}}, \quad \mathbf{b}_2 = \frac{2\pi}{a}\hat{\mathbf{y}}. \quad (13.17)$$

任意倒格矢都可写成

$$\mathbf{G}_{mn} = m\mathbf{b}_1 + n\mathbf{b}_2, \quad m, n \in \mathbb{Z}. \quad (13.18)$$

若圆孔半径为 r ，孔内相对介电常数为 ε_h ，背景相对介电常数为 ε_b ，则填充因子为

$$f = \frac{\pi r^2}{a^2}. \quad (13.19)$$

对很多教学例子，常取孔内为空气，即 $\varepsilon_h \approx 1$ ；背景可以取某个高折射率半导体对应的介电常数 $\varepsilon_b = n_b^2$ 。这里故意先不把具体数字写死，因为几何推导本身与材料数据库无关。

13.4.2 圆孔指标函数的傅里叶系数

PWEM 实际使用的关键在于：必须把式 (13.11) 里的 $\eta_{\mathbf{G}}$ 明确写出来。为此，把单胞中的介质函数倒数写成

$$\eta(\mathbf{r}_{\parallel}) = \eta_b + (\eta_h - \eta_b)\chi_{\text{hole}}(\mathbf{r}_{\parallel}), \quad \eta_h = \frac{1}{\varepsilon_h}, \quad \eta_b = \frac{1}{\varepsilon_b}, \quad (13.20)$$

其中 χ_{hole} 是圆孔指标函数，孔内取 1、孔外取 0。于是

$$\eta_{\mathbf{G}} = \eta_b \delta_{\mathbf{G},0} + (\eta_h - \eta_b)\chi_{\text{hole},\mathbf{G}}. \quad (13.21)$$

对单胞中心处的圆孔, 指标函数的傅里叶系数可直接积分得到

$$\chi_{\text{hole},0} = f, \quad (13.22)$$

而对任意非零倒格矢 $\mathbf{G} \neq 0$, 有

$$\begin{aligned} \chi_{\text{hole},\mathbf{G}} &= \frac{1}{A_c} \int_{\text{disk}} e^{-i\mathbf{G} \cdot \mathbf{r}} d^2r \\ &= \frac{1}{A_c} \int_0^r \int_0^{2\pi} e^{-i|\mathbf{G}|\rho' \cos \phi} \rho' d\phi d\rho' \\ &= \frac{2\pi}{A_c} \int_0^r J_0(|\mathbf{G}|\rho') \rho' d\rho' \\ &= f \frac{2J_1(|\mathbf{G}|r)}{|\mathbf{G}|r}. \end{aligned} \quad (13.23)$$

这里用到了

$$\int_0^{2\pi} e^{-i|\mathbf{G}|\rho' \cos \phi} d\phi = 2\pi J_0(|\mathbf{G}|\rho')$$

以及

$$\int_0^r J_0(|\mathbf{G}|\rho') \rho' d\rho' = \frac{r J_1(|\mathbf{G}|r)}{|\mathbf{G}|}.$$

也就是说, 圆孔 Fourier 系数里之所以自然出现 Bessel 函数, 不是额外假设, 而只是极坐标积分对圆对称几何的直接结果。于是, $\eta_{\mathbf{G}}$ 可写成

$$\eta_0 = \eta_b + (\eta_h - \eta_b)f, \quad (13.24)$$

$$\eta_{\mathbf{G}} = (\eta_h - \eta_b)f \frac{2J_1(|\mathbf{G}|r)}{|\mathbf{G}|r}, \quad \mathbf{G} \neq 0. \quad (13.25)$$

式 (13.24) 和式 (13.25) 是正方晶格圆孔二维光子晶体最常用的解析傅里叶系数。它们把几何参数 r/a 和材料参数 $\varepsilon_h, \varepsilon_b$ 直接带入 PWEM 矩阵, 因此是“从几何到能带”的一组可直接计算的接口。

若圆孔不在原点, 而是位于单胞中的某个位置 $\mathbf{r}_c$, 则式 (13.25) 还要再乘上相位因子 $e^{-i\mathbf{G} \cdot \mathbf{r}_c}$ 。这个位置相位对多孔单胞、双原子单胞或故意引入对称性破缺的结构尤为重要。

13.5 从二维教学模型到真实 slab: 修正等效介电常数

二维 PWEM 假设结构沿 z 方向均匀, 不能直接用于有限厚度 slab。为近似三维效应, 可先用传递矩阵法计算纵向基模场剖面 $\phi_0(z)$ 及等效折射率 n_{eff} , 再定义光子晶体层约束因子

$$\Gamma_{\text{PC}} = \int_{\text{PC layer}} |\phi_0(z)|^2 dz, \quad (13.26)$$

然后通过联立

$$n_{\text{eff}}^2 = f\tilde{\varepsilon}_a + (1-f)\tilde{\varepsilon}_b, \quad \tilde{\varepsilon}_b - \tilde{\varepsilon}_a = \Gamma_{\text{PC}}(\varepsilon_b - \varepsilon_a) \quad (13.27)$$

确定修正介电常数 $\tilde{\varepsilon}_a, \tilde{\varepsilon}_b$ 。这两条式子的物理意义值得在本章就说清楚。

第一, $n_{\text{eff}}^2 = f\tilde{\varepsilon}_a + (1-f)\tilde{\varepsilon}_b$ 强制二维模型与真实纵向基模共享同一个平均传播常数。也就是说, 二维模型是先把真实 slab 波导的基模传播常数锁定, 再去反推横向周期结构应该看见多大的平均介电环境, 而非任意找一对“好看的等效介电常数”。

第二, $\tilde{\varepsilon}_b - \tilde{\varepsilon}_a = \Gamma_{\text{PC}}(\varepsilon_b - \varepsilon_a)$ 说明真实折射率反差并不会被导模百分之百看见。只有落在光子晶体层中的那一部分场能量, 才真正参与横向布拉格散射。因此 Γ_{PC} 的作用, 是把“有限厚度刻蚀”和“模场并不完全待在光子晶体层内”这两个三维事实压缩进二维模型, 而非额外引入一个经验修正因子。

因此, 这个修正后的二维模型更准确的用途是: 它能较可靠地给出候选带边、频率排序和高对称点附近的模式家族, 但它仍然不能直接给出辐射常数、上下出光比和真实阈值排序。只要问题开始显式依赖法向辐射、外延不对称、复杂孔形或有限尺寸, 就必须继续交给 第十章、第十一章、第十四章和第十五章。

13.6 如何搭建一个最小 PWEM 计算

现在把上面的方程压缩成一个初学者可以直接照着执行的基本流程。为了避免把软件界面和物理步骤混为一谈, 本节使用 solver-agnostic 的写法, 也就是无论你最终用自写代码、MATLAB、Python 还是某个已有求解器, 主流程都应当是相同的。

13.6.1 步骤一: 定义教学型几何与参数表

先固定一个最小例子。最自然的选择是: 正方晶格、单个圆形空气孔、背景为高折射率介质。参数表可写成 表13.1 的形式。

表 13.1: 正方晶格圆孔二维光子晶体的最小 PWEM 教学参数表。这里给出的是“示例参数集”, 目的是说明如何搭建模型, 而不是宣称任何最终器件性能。

参数	符号	示例写法	说明
晶格常数	a	设计自由参数	决定归一化频率与目标波长映射
孔半径	r	$0.30a$	常见教学型填充比起点
背景折射率	n_b	3.4	可代表高折射率半导体背景
孔内折射率	n_h	1.0	空气孔近似
倒格矢截断	N	3, 5, 7, 9	表示保留满足 $ m , n \leq N$ 的倒格矢
路径	—	$\Gamma \rightarrow X$ $\rightarrow M \rightarrow \Gamma$	读取主要高对称点信息

这里需要特别强调: a 在归一化带图中并不需要先给定绝对数值。因为纵轴通常使用

$$\Omega = \frac{\omega a}{2\pi c} = \frac{a}{\lambda}, \quad (13.28)$$

所以只要先得到 Ω , 后面就可以再通过目标工作波长反推出所需的晶格常数 a 。这也是能带图在概念设计阶段特别高效的原因之一。

13.6.2 步骤二：生成倒格矢基底并做截断

利用式 (13.18) 生成所有满足 $|m|, |n| \leq N$ 的倒格矢。对标量 H_z 问题，这会产生

$$N_G = (2N + 1)^2 \quad (13.29)$$

个平面波基函数。随着 N 增大，矩阵维度迅速增长，因此必须把“能算出来”和“已经收敛”区分开。一个安全的基本操作是：至少对目标频段做两到三组不同 N 的对比，而不是只跑一次就接受结果。

13.6.3 步骤三：计算傅里叶系数并组装矩阵

对每一个倒格矢差值 $\mathbf{G} - \mathbf{G}'$ ，用式 (13.24) 和式 (13.25) 计算 $\eta_{\mathbf{G}-\mathbf{G}'}$ ，然后按式 (13.12) 组装矩阵元

$$A_{\mathbf{G}, \mathbf{G}'}(\mathbf{k}) = \eta_{\mathbf{G}-\mathbf{G}'}(\mathbf{k} + \mathbf{G}) \cdot (\mathbf{k} + \mathbf{G}'). \quad (13.30)$$

对每一个给定的 $\mathbf{k}$ ，这都是一个实对称或厄米型矩阵本征问题。求解其一组最小本征值，就得到该 $\mathbf{k}$ 处的本征频率平方。

13.6.4 步骤四：沿高对称路径逐点求解

对正方晶格，最常用的高对称路径是

$$\Gamma = (0, 0) \rightarrow X = \left(\frac{\pi}{a}, 0\right) \rightarrow M = \left(\frac{\pi}{a}, \frac{\pi}{a}\right) \rightarrow \Gamma. \quad (13.31)$$

把这条路径离散成若干采样点，对每个采样点都构造矩阵、求本征值，再按频率大小规律排序并连续追踪，就得到带图。这里的“连续追踪”重要，因为不同 $\mathbf{k}$ 点上按数值大小规律排序的第 n 个本征值，不一定总对应同一条物理能带。实际处理时，往往还要结合模的对称性或本征向量重叠来辅助跟踪。

13.6.5 步骤五：做收敛检查，而不是直接相信第一张图

PWEM 的第一张带图通常只能叫“初稿”。要让它进入后续建模链，至少应完成三项检查。

1. 基底收敛：提高 N 后，目标频段内的候选带位置是否稳定；
2. 路径分辨率收敛：增加路径采样点后，平带和简并附近的细节是否稳定；
3. 参数一致性检查：若以不同等价写法生成单胞或改变原点，带图是否只发生应有的标记变化，而不发生物理上不应出现的漂移。

如果这三项还没有做，图就只适合作为内部草图，不适合拿来指导 CWT 截断或 RCWA 的频段设置。

以表 13.1 中的示例参数为起点，可以先取 $N = 3$ 跑出一张粗略带图，再逐步提高到 $N = 5, 7, 9$ 。通常你会观察到：随着截断增加，低阶带先趋于稳定，高阶带稳定得更慢；Gamma 点和其他高对称点附近的简并关系会逐渐变清楚；而是否存在明显带隙、哪一段带在目标频段附近最平，则必须以收敛后的图为准，而不是以粗网格初稿为准。

把收敛讨论画成对数图: N_{PW} 与带隙精度

为了让“收敛”从定性描述变成可执行判据, 图13.3 给出本书教学脚本的误差-截断对数图。误差定义为

$$\varepsilon_{\text{gap}}(N_{\text{PW}}) = \frac{|\Delta\Omega_{23}^{(N_{\text{PW}})} - \Delta\Omega_{23}^{(\text{ref})}|}{\Delta\Omega_{23}^{(\text{ref})}},$$

其中参考解取高截断 $N_{\text{PW}} = 225$ 。

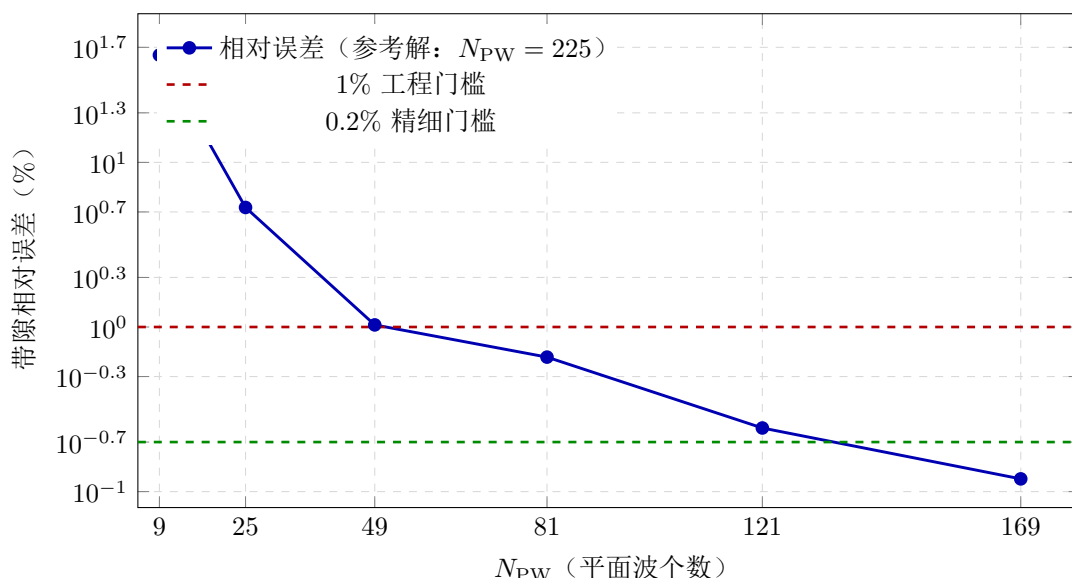


图 13.3: PWEM 截断误差随 N_{PW} 的对数衰减趋势 (本书教学脚本数据)。误差定义为 X 点带隙代理 $\Delta\Omega_{23}$ 相对高截断参考解 ($N_{\text{PW}} = 225$) 的偏差。该图直观说明“先做误差-截断曲线, 再给出带图结论”这一数值卫生原则。

这张图有两个直接用途。第一, 给出“什么时候可以停止加基底”的硬门槛: 若只做候选筛选, 1% 误差常已可用; 若要做带隙微调和近简并排序, 通常应压到 0.2% 量级。第二, 提醒读者“经验值不是常数”: 本书这个中等折射率对比教学例子在 $N_{\text{PW}} \approx 100$ 已接近稳定, 但高对比、尖锐几何边界或金属参与结构通常要更高截断 [29, 40]。因此, 较稳妥的报告方式是“报告误差-截断曲线”, 而不是只报一个 N_{PW} 数字。

13.7 本书附带的 Python 计算脚本与实际带图

前面的五步已经把 PWEM 的数学与建模流程讲清楚。为了避免这一章再次退化成“公式讲完就结束”, 本书项目中同时附带了一个可直接运行的 Python 示例:

- 脚本文件: `code/pwem_square_hole/run_pwem_square_hole.py`
- 输出目录: `code/pwem_square_hole/output/`
- 频带数据文件: `pwem_square_hole_hz_bands.csv`

- 汇总信息文件: `pwem_square_hole_summary.json`

它是已经在本项目中实际运行过的脚本，而非伪代码。脚本使用 Python 标准库实现二维标量 H_z PWEM。

这一份教学型计算采用的参数与 表13.1 一致，具体为

$$\frac{r}{a} = 0.30, \quad \varepsilon_b = 3.4^2, \quad \varepsilon_h = 1, \quad |m|, |n| \leq 2, \quad (13.32)$$

并沿 $\Gamma \rightarrow X \rightarrow M \rightarrow \Gamma$ 路径计算前六条能带。这里故意把截断保持在教学型规模，也就是 25 个倒格矢分量。这样做的目的是让读者能够把“式子里的每一项”与“实际算出来的图”直接对上，而非追求最终设计精度。

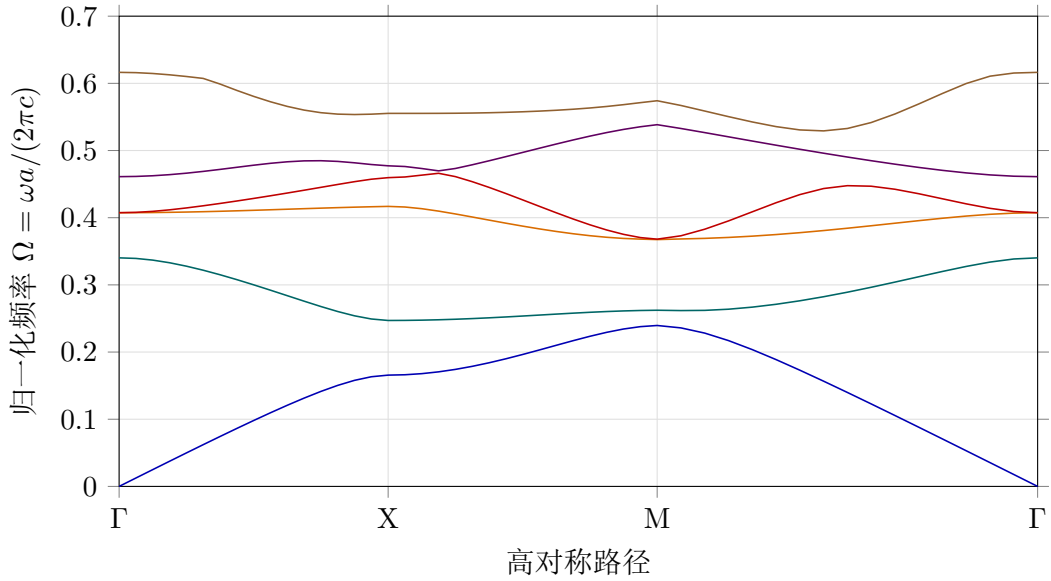


图 13.4: 由本书项目附带的 Python 脚本实际计算得到的二维标量 H_z PWEM 能带图。结构为正方晶格圆形空气孔，高折射率背景取 $\varepsilon_b = 3.4^2$ ，孔半径取 $r/a = 0.30$ ，截断为 $|m|, |n| \leq 2$ 。该图是教学型被动无限周期结果，不是 slab PCSEL 的最终器件结论。

现在不只概括“这图有三点要看”，而是把它作为一张教学型带图逐段阅读。

先读横轴，再读路径闭合是否自洽 横轴是沿 $\Gamma \rightarrow X \rightarrow M \rightarrow \Gamma$ 的累计路径长度，而非实空间距离。图中四个关键位置分别是 $s = 0, 0.5, 1.0$ 和 1.7071 。由于最后回到同一个 Γ 点，图的左右两端应当给出同一组本征频率。对 图13.4 来说，这个闭合检查成立：左端与右端都回到

$$\Omega_\Gamma \approx (0, 0.3402, 0.4074, 0.4074, 0.4611, 0.6164). \quad (13.33)$$

这一步非常基础，却是读者检查代码和路径设置是否正确的第一道门槛。若路径末端回不到起点对应的同一组频率，通常是路径、排序或数值实现出了问题，而非物理新现象。

把图和表合起来读：先看排序是否稳定，再看间距在哪里缩小 表13.2 给出的是帮助初学者形成数值直觉，而非另一种“装饰性数据”。比如，第二条带在 Γ 点位于 0.3402，到 X 点降

表 13.2: 本书附带 Python 脚本在三个高对称点给出的前六条带归一化频率。这个表的作用是把“图上的大概位置”落实成可以逐项核对的数字，而非代替带图。

带序号	Γ	X	M
1	0.0000	0.1657	0.2396
2	0.3402	0.2470	0.2622
3	0.4074	0.4168	0.3674
4	0.4074	0.4596	0.3681
5	0.4611	0.4772	0.5385
6	0.6164	0.5553	0.5741

到 0.2470，再到 M 点略回升到 0.2622。只看曲线容易得到“它往下弯了”的印象；把数字放出来以后，就能进一步判断：它在 $\Gamma \rightarrow X$ 段的变化远大于在 $X \rightarrow M$ 段的变化，因此目标频段若放在第二带附近，真正敏感的方向更可能出现在靠近 X 点之前，而不是整个高对称路径平均同样敏感。

再读斜率：带图上的上升和下降到底意味着什么 沿高对称路径画出的带图，本质上是在观察 $\omega_n(\mathbf{k})$ 沿某一条一维路径的切片。因此，曲线的局部斜率并非完整的二维群速度矢量，但它至少告诉我们群速度在当前路径方向上的投影符号。对本图而言，第一条带在 $\Gamma \rightarrow X$ 段持续上升，说明其群速度投影为正；第二条带在同一路段持续下降，说明投影为负。第三、第四条带在离开 Γ 后一上一下地分开，则说明原先共享同一高对称点频率的两支候选模，在该方向上获得了不同的色散响应。对 PCSEL 读者来说，这一步很关键，因为后续讨论“哪一支更容易被微扰推到目标窗口里”时，实际被比较的是这种局部色散响应，而不是一张图上曲线谁更好看。

还要读带间距：哪几支带彼此最危险 看带图不能只盯住单条曲线，还要盯住相邻曲线的最小间距。就这张图而言，第三、第四条带在 Γ 点严格简并，而在 M 点又几乎重新并拢；第一、第二条带虽然在整个路径上没有真正简并，却在 $X \rightarrow M$ 段显著靠近。前一种“先简并再分裂再并拢”的模式，通常提示后续对称性破缺会显著改写模态组合；后一种“始终分离但局部靠近”的模式，则提醒我们在参数扫描时要留意带排序交换或近邻竞争。带图的设计价值在于，它能在这一阶段就把“哪些模式以后可能互相抢位置”提前暴露出来。

再读第一段 $\Gamma \rightarrow X$ ：谁在上升，谁在下降，谁在分裂 从 Γ 走向 X 时，第一条带由 0 单调上升到约 0.1657，第二条带则由约 0.3402 下降到约 0.2470。这种一升一降的组织方式告诉我们：即使是最简单的教学型结构中，低阶带也是在不同对称方向上重新分配色散斜率，而非彼此平行地整体平移。

第三和第四条带在 Γ 点处从

$$\Omega \approx 0.4074 \quad (13.34)$$

起始时完全重合，但走向 X 后分裂为约 0.4168 和 0.4596。这是本图最值得初学者盯住的第一处现象：高对称点的简并一旦离开高对称位置，就会按照允许的耦合通道重新组织。第五章

讲“简并与近简并”时，如果没有真实图支撑，容易停留在概念词条；现在这张图把那个概念具体化了。

再读第二段 $X \rightarrow M$ ：哪些带在靠近，哪些带在重新并拢 从 X 走向 M 的过程中，第一和第二条带继续靠近，到 M 点分别达到约 0.2396 和 0.2622。它们没有真正简并，但频率间隔已经显著缩小。这类“靠近但未相交”的行为在真实设计里重要，因为稍微加入几何微扰、有效折射率变化或有限基底误差，就可能让近邻带的排序和间隔重新调整。

更值得注意的是第三和第四条带。在 X 点它们分得较开，而到 M 点时已经几乎重新并拢为

$$\Omega_M^{(3,4)} \approx 0.3674, 0.3681. \quad (13.35)$$

这说明：简并非某一支带永久携带的标签，而是和对称点绑定的结构信息。对 PCSEL 读者来说，这为后续模式分裂、偏振选择和对称性破缺分析提供原始素材。若某组候选模在某个高对称点天然接近，那么它们通常就是后续最容易被微扰重排的一组对象。

最后读第三段 $M \rightarrow \Gamma$ ：哪些带在回归原先的 Gamma 点家族 从 M 返回 Γ 的路上，第一条带逐步落回 0，第二条带回到约 0.3402，第三和第四条带也重新回到 Gamma 点简并值约 0.4074。这段路径的意义不只是“把图画完整”，而是帮助读者确认某些模族到底是不是围绕同一个 Gamma 点家族在演化。对本图而言，第三和第四条带在 Γ 点具有相同的本征频率，构成了一组典型的 Gamma 点近邻候选模族；如果后续要做最小 CWT、对称性破缺或偏振选择分析，这一组模就是自然的重点对象。

把频率数字翻译成几何尺度 带图若只停在无量纲频率，就还没有真正进入设计语言。利用式 (13.28)，可以把图上的频率直接换成晶格常数。比如，若我们希望把图中第二条带在 Γ 点附近的候选位置

$$\Omega \approx 0.3402 \quad (13.36)$$

映射到真空波长 940 nm，则应取

$$a \approx 0.3402 \times 940 \text{ nm} \approx 320 \text{ nm}. \quad (13.37)$$

这并非说“取 320 nm 就得到可工作的 PCSEL”，而是说明：带图首先提供的是几何缩放起点。后续还要问 slab 有效折射率是否一致、辐射通道是否合适、有限尺寸和外延层是否会改写这个候选窗口。

这张图对后续建模最实际的三种用途 第一种用途是给 CWT 选候选模。对本图而言，第三、第四条带在 Γ 点简并、在离开 Γ 后分裂，适合进入“哪一组模最容易被对称性破缺重排”的分析。

第二种用途是给 RCWA 设频段窗口。比如，如果后续重点想研究 Γ 点附近的第二至第四条带，那么 RCWA 的频率扫描就不应再盲扫整个宽频范围，而应围绕

$$\Omega \approx 0.34 \text{ 到 } 0.46 \quad (13.38)$$

的窗口定向展开。

第三种用途是做收敛与代码一致性核查。由于本图采用的只是 $|m|, |n| \leq 2$ 的教学型小基底，所以它特别适合检查“公式是否写对”“路径是否写对”“本征值排序是否写对”。它不适合作为最终设计图，但适合作为第一张必须先算对的图。

这张图不能被过度解读的地方 虽然图13.4已经是脚本实际算出的结果，但它仍然只是二维标量、被动、无限周期、小基底截断的结果。因此，读者看到它时要主动克制四种过度解读。

第一，不能把第一条带在 Γ 点从零开始误读成“零阈值模式”。那只是二维标量本征问题中最低分支的数学起点。

第二，不能把第三、第四条带在 Γ 点的简并直接等同于“真实 slab PCSEL 中一定会出现的两支激射模”。真实器件里还要经过纵向层栈、辐射通道和有限尺寸筛选。

第三，不能把图中某一段看起来较平，就直接称为“最低阈值带”。平带只说明色散斜率变小，不自动给出净模增益排序。

第四，不能把这张图上的任何频率直接当作器件工作频率。真实工作状态下，温升、载流子分布、增益峰位和折射率回写都会让实际激射频率偏离冷腔被动结果。

图13.4的严格身份是“二维标量、被动、无限周期、有限基底截断”的 PWEM 结果。它能证明本章的推导确实可以落成可运行代码，也能帮助读者练习读图；但它并不自动给出垂直辐射损耗、有限孔阵边界效应、阈值增益排序或器件级阈值电流。

13.8 如何读一张 PWEM 算出来的能带图

第5章已经从概念上讲过“怎么看能带图”。现在再往前走一步，专门从计算结果的角度说一遍：一张由 PWEM 算出的带图，应该按照什么顺序去读，才不至于被图形表面牵着走。

第一步：先确认你看的图已经收敛

这是最容易被忽略、却最不应该省略的一步。若两次截断阶数不同的计算在目标频段内仍明显偏移，那么后续任何“这里有带隙”“这里是候选 Gamma 点模”的判断都站不住。对真正的研究流程而言，收敛检查是读图资格的一部分，而非额外工作。

第二步：把归一化频率翻译回设计波长

若目标工作波长为 λ_0 ，而 PWEM 算得某个候选模位于归一化频率 Ω_0 ，则两者关系由式 (13.28) 给出：

$$a = \Omega_0 \lambda_0. \quad (13.39)$$

这一步看似简单，却是从“无量纲色散图”走向“器件几何尺寸”的关键桥梁。很多新手会直接比较两张不同 a 的带图而不做无量纲转换，结果把缩放效应和真实模式重排混在一起。

第三步：重点看 Gamma 点附近哪些带落在目标窗口

对 PCSEL 来说, PWEM 图最有价值的区域通常是 Gamma 点附近的目标频段, 而非整条路径平均对待。需要重点记录的是: 目标工作波段附近在 Gamma 点有几支候选带、它们彼此相隔多远、有没有简并或近简并、以及它们离相邻竞争带有多近, 而非“全图最高最低的带”。

第四步：不要只看频率, 还要看本征向量组成

PWEM 的一个常被低估的输出, 是本征向量 $\mathbf{h}_n$ 。它告诉你某个模式主要由哪些平面波分量组成。若目标模在 Gamma 点附近主要由少数一级倒格矢分量组成, 那么它通常很适合进入最小 CWT 模型; 若某个模式严重依赖大量高阶分量才能表达, 则简单 CWT 可能就不够了, 需要更高阶模型或直接交给 RCWA/FEM 处理。

第五步：把“图上能看见的”和“图外必须继续验证的”分开

即使 PWEM 图已经完全收敛, 它仍然首先只是被动无限周期的带结构图。因此, 你最多可以从中稳健地读出: 候选频段、Gamma 点模族、带边位置、可能的带隙、简并结构、以及对几何参数变化的一阶趋势。你不能由此直接得到: 辐射损耗、法向远场、有限阵列边缘泄漏、阈值增益排序, 也不能得到阈值电流和热稳定工作窗口。

易错点

把 PWEM 图上“Gamma 点附近最平的一条带”直接称为“最终最低阈值模”, 是入门阶段最常见、也最危险的偷换。PWEM 只负责给你候选模目录, 真正的辐射损耗要交给 RCWA 或开放本征问题, 真正的有限器件排序要交给 FDTD/FEM, 真正的阈值电流和工作稳定性要交给外延、电学和热学模型。

13.9 从 PWEM 结果中我们究竟能得到什么

如果把本章结果纳入整本书的工作流, 那么 PWEM 至少提供六类高价值输出。

第一, 目标波段附近的候选模目录。这决定后续 RCWA 和 CWT 应该重点跟踪哪几支带。

第二, Gamma 点附近的简并与分裂结构。这为后续偏振工程、对称性破缺和模式选择提供最初的分类。

第三, 带边和平带信息。这为理解候选模的色散特征、群速度趋势和态密度变化提供线索。

第四, 几何参数的一阶灵敏度。通过扫描 r/a 、介质对比度或单胞微扰, 可以快速判断哪类参数最值得继续优化。

第五, 本征向量的主导平面波组成。这能直接指导第十章、第十一章 CWT 截断时保留哪些耦合分量。

第六, 从归一化频率到实际波长的尺度映射。这使能带设计可以在无量纲层面高效探索, 再映射回具体工作波长。

但同样必须把边界说清楚。PWEM 无法单独提供:

- 真实 slab 开放结构中的辐射损耗和光锥耦合强度;
- 有限孔阵中的边缘模式、横向包络和真实远场;
- 有源区重叠、阈值增益、阈值电流和热漂移后的模式排序;
- 电注入不均匀性、热热点和模式竞争导致的器件级工作状态。

这一边界并非 PWEM 的缺陷, 而是它的职责范围。PWEM 的职责是把候选模目录、几何趋势和模式组织结构提供给后续模型, 而不是替后续模型抢结论。

13.10 PWEM 结果应该怎样交给后续章节

为了明确本章与后续章节的接口, 本章最后必须明确: PWEM 是后续几章的输入接口, 而非终点。

对第十章、第十一章的 CWT 来说, PWEM 最重要的输入是: 目标模主要由哪些倒格矢分量组成、哪些简并在 Γ 点附近最需要保留、以及某种几何微扰会优先影响哪类模族。这些信息决定最小耦合波模型应怎样截断。

对第十四章来说, PWEM 提供的是频段窗口、候选模标签和参数扫描方向。换句话说, PWEM 告诉 RCWA “去哪里看”, RCWA 再回答 “这些模式怎样与开放辐射通道交换能量”。

对第十五章来说, PWEM 提供的是有限器件仿真的初始候选频段、目标模族和对称性预期。这样全波仿真就不必在过大的参数空间里盲扫。

对第十七章至第二十章这些外延与器件模型来说, PWEM 更像一个前端几何筛选器。它只能告诉你某个平面周期方案在色散层面是否值得继续, 并不能跳过纵向模、量子阱重叠、电流扩展和热回写这些器件级步骤。

表 13.3: PWEM 结果向后续方法交付的最小接口量。

PWEM 输出	下一步最直接使用它的方 法/章节	交付时最该保留的字段
候选频段与归一化频率窗口	第十四章、第十五章	目标带号、 $\Gamma/X/M$ 位置、对应 Ω 与换算后的目标波长区间
Γ 点简并结构与 irrep family	第八章至第十章	singlet/doublet 标记、近简并支路、预期偏振/辐射家族
主导倒格矢分量与本征向量权重	第十章、第十一章	哪几个 $\mathbf{G}$ 分量占主导、是否适合四波或更高阶 CWT 截断
对几何参数的一阶趋势	第十四章、第二十五章	随 r/a 、 a 、填充比变化的带移动方向与模族重排趋势
被动无限周期筛选结论	第十七章、第十九章	仅作为 “几何值得继续” 的前端证据, 不越权声明阈值电流或工作窗口

若某一组正方晶格圆孔参数在 PWEM 中显示：目标波长附近存在两支接近的 Gamma 点候选带，其中一支主要由四个一级倒格矢分量组成，另一支则混入更多高阶分量，那么前者通常更适合先送入最小 CWT 和 RCWA，后者则提醒研究者该结构可能已经超出简单四波模型的舒适区。这就是“PWEM 输出被后续方法使用”的典型例子。

13.11 为什么真实 PCSEL 不能停在二维 PWEM

本章必须在结束前再把一个界限说清楚：二维 PWEM 是极好的入门工具，也是高效的前端筛选工具，但真实 PCSEL 的关键物理并不止于二维周期色散。

PCSEL 之所以特别，不只是因为它有二维光子晶格，还因为它是 slab 开放系统、有 Gamma 点法向辐射、有真实外延层、有量子阱重叠、有电注入路径和温升反馈。二维 PWEM 把这些东西全部暂时搁置，目的是先让读者把“能带从何而来”学会。只要这个前提记得牢，它就会是有效的前端工具；一旦忘了这个前提，它就会立刻从“好工具”变成“误导源”。

因此，正确的学习顺序是“先用 PWEM 建立被动无限周期候选模地图，再用 RCWA 处理开放单胞，再用 FDTD/FEM 处理有限器件，再用器件模型处理注入与热”，而非“用 PWEM 取代后续所有方法”。这条链条一旦打通，PWEM 才有明确的工作流位置。

回看本章

PWEM 的核心思想，是把周期介质中的 Maxwell 本征问题投影到平面波基底上，从而把“模式存在性与色散关系”变成一个可计算的矩阵本征问题。对正方晶格圆形孔洞二维光子晶体而言，几何通过圆孔指标函数的傅里叶系数进入矩阵元，而能带则通过沿高对称路径逐点求解该矩阵获得。对 PCSEL 研究来说，PWEM 的价值在于高效建立候选模目录、Gamma 点简并结构和参数扫描方向；它的边界则在于不处理开放辐射、有限尺寸、有源阈值和器件级热电问题。把这两面同时看清，PWEM 才能成为整本书建模链中的可靠起点。

练习题

1. [基础] 从式 (13.5) 出发，完整推导式 (13.12)，并解释为什么矩阵中会自然出现卷积型傅里叶系数 $\eta_{G-G'}$ 。
2. [基础] 对正方晶格圆孔单胞，证明式 (13.23) 中的圆孔指标函数傅里叶系数确实包含 Bessel 函数 J_1 。
3. [进阶] 设目标工作波长为 940 nm，若某候选模在 PWEM 图中对应归一化频率 Ω_0 ，写出晶格常数 a 与 Ω_0 的关系，并解释为什么这一步不能跳过。
4. [进阶] 说明为什么“Gamma 点附近最平的带”不能直接被称为“最低阈值激射模”，并列出至少三个后续必须继续验证的问题。
5. [综合] (思路提示：先定层级，再列关键式) 设计一份你自己的 PWEM 输出报告模板，要求它能被第十章、第十四章直接使用。

延伸阅读

- 立即接着读 第十四 章。这样最容易看清：PWEM 给出的是被动无限周期候选模目录，而 RCWA 开始处理开放周期结构中的辐射与共振。
- 若想进一步打牢本章的数学基础，可回到 第二十八 章，重点看 Fourier-Bloch 展开与卷积结构。
- 若想把二维教学模型和真实 PCSEL slab 联系起来，可回看 第十七 章，再思考二维 PWEM 中的有效折射率代理到底保留了什么、忽略了什么。
- 若想追溯 PWEM 的经典起点，可参考 [21, 37, 29]。

第十四章 RCWA 在周期开放结构与阈值分析中的应用

读完本章，你能

- 理解为什么在 第5章 建立能带概念、并在 第十三章 用 PWEM 算出候选模目录之后，还必须引入 RCWA。
- 掌握 RCWA 的散射矩阵视角，明白它为什么特别适合处理层状周期开放结构。
- 学会区分“周期单胞被动共振结论”“光学阈值增益估计”和“有限器件级结论”这三个层次。

读前准备 建议已掌握 第七章、第十章和第十三章 中关于候选模筛选、耦合波图像和阈值条件的内容。

本章结构

本章所处模型层级：L1→L2

第5章 已经说明怎样阅读能带图，第十三章 又把它落实成了 PWEM 计算，于是到这里读者已经知道“候选模在哪里、怎样随几何变化、主导平面波分量大致是什么”。但这仍然不是面发射激光器的完整光学问题，因为 PWEM 还没有真正处理开放边界，也不能直接回答法向辐射如何离开器件。本章要做的事，就是把问题推进到“周期开放结构的散射与共振”层级。我们先说明为什么 RCWA 特别适合 PCSEL，再从分层和傅里叶展开建立散射矩阵框架，随后解释共振极点、被动谱和光学阈值估计之间的关系，最后明确它的边界：RCWA 适用范围宽，但它默认描述的仍是周期单胞，不是完整有限器件。

14.1 为什么在 PWEM 之后自然轮到 RCWA

PWEM 解决的是被动本征问题，能告诉我们 Gamma 点附近有哪些候选模、主要由哪些平面波分量组成。但 PCSEL 的关键物理之一——模式如何向自由空间辐射——在 PWEM 的封闭本征框架里根本没有出现。

因此，接下来必须问的不再是“有哪些本征模”，而是“入射波和出射波如何经过层状周期结构发生散射”。这是 RCWA 的天然视角：它把周期结构看成散射系统，直接处理开放边

界、入射通道与辐射耦合。对 PCSEL 来说，这一视角能直接回答三类问题：某个候选模是亮模还是暗模；某种几何扰动主要增强的是上方辐射还是下方辐射；以及在给定小信号增益下，开放极点何时靠近实轴。

与第 12 章的对比 PWEM 擅长快速扫描晶格参数并找到候选模频段，但它默认封闭边界，无法区分向上辐射和向下辐射的贡献。RCWA 则以开放散射问题为出发点，可以直接计算 S_{11} （反射）和 S_{21} （透射），从而给出上下出光比例。

14.2 RCWA 的基本思路：平面内做傅里叶，纵向做分层

RCWA 的名称里虽然有“耦合波”，但它和前面 CWT 的工作方式并不相同。CWT 的核心是主动保留少数主导波分量，用低维耦合模型换取解释性；RCWA 则更接近一种严格的半解析数值方法，它在每一层中对平面内周期方向做傅里叶展开，而在纵向把结构切成若干层逐层求解。

设器件在 x 和 y 方向周期，在 z 方向分层。对每一层 ℓ ，把介电常数与场都在平面内写成傅里叶级数，就得到一个有限维模态系统。该层中的场可写成“向上”和“向下”传播的模叠加：

$$\mathbf{f}_\ell(z) = \mathbf{W}_\ell^+ \mathbf{a}_\ell^+ e^{+i\beta_\ell z} + \mathbf{W}_\ell^- \mathbf{a}_\ell^- e^{-i\beta_\ell z}, \quad (14.1)$$

其中 $\mathbf{f}_\ell$ 是该层选取的场变量向量， $\mathbf{W}_\ell^\pm$ 是本征模基底矩阵， β_ℓ 给出纵向传播常数， $\mathbf{a}_\ell^\pm$ 分别表示上、下行模振幅。

注记

纵向传播常数矩阵 β_ℓ 的定义。在周期层中，纵向传播常数应由层本征值问题给出：

$$\mathbf{M}_\ell \mathbf{W}_\ell = \mathbf{W}_\ell \beta_\ell, \quad \beta_\ell \equiv \beta_\ell.$$

这里 $\mathbf{M}_\ell$ 是由该层 Fourier 卷积矩阵构造的 RCWA 层算子， $\mathbf{W}_\ell$ 为本征模基底矩阵。仅在均匀层（无谐波耦合）时，才可退化为单谐波形式

$$\beta_n = \sqrt{k_0^2 \varepsilon_\ell - |\mathbf{k}_{\parallel,n}|^2}.$$

当 β_n 为纯虚数时，对应倏逝波；当为实数时，对应传播波。

这一步的物理意义非常明确：在每一层里，平面内周期导致的多谐波耦合已经被吸收进层本征模中；而真正的结构复杂性，被转化成“层与层之间如何通过边界条件连接”。于是，复杂三维问题被拆成两部分：层内求模，层间拼接。图14.1直观展示了如何将连续的倾斜光栅剖面离散化为多层阶梯近似结构，这是 RCWA 处理任意形状周期结构的核心技巧。

图14.1揭示的离散化思想是 RCWA 能够处理复杂周期结构的关键：任何连续的剖面都可以用足够多的薄层来逼近，而每一层内部的平面内周期性使得我们可以用傅里叶级数展开。这种“分层 + 展开”的策略，使 RCWA 可以在保持半解析离散结构的同时处理复杂剖面。

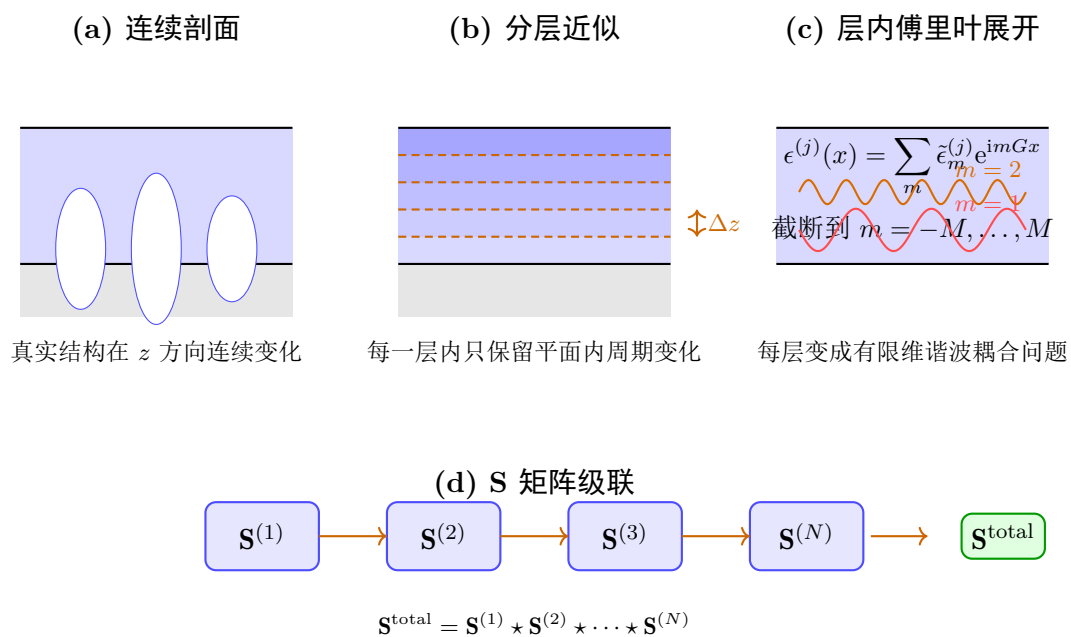


图 14.1: RCWA 中的阶梯近似离散化策略。(a) 原始连续剖面: 倾斜光栅或曲面结构, 其介电常数分布在空间中连续变化; (b) 阶梯近似: 将结构沿 z 方向切割为 N 层, 每层内部介电常数在平面内周期分布但在 z 方向恒定, 层与层之间通过边界条件连接; (c) 收敛性: 随着层数 N 增加, 阶梯近似逐渐逼近真实剖面, RCWA 解收敛到严格解。离散化层数必须通过“层数-目标量”收敛曲线确定; “倾角较小时 20–50 层常可收敛”仅可作为 [工程经验] 初值, 不可替代收敛检验 [40, 44]。

14.3 为什么散射矩阵是 RCWA 的真正核心

如果只是逐层传递振幅，最直接的想法是用传输矩阵（transfer matrix）。但对高对比度、强倏逝波或厚层结构，这样做在数值上常常不稳定，因为指数增长和指数衰减模式会在矩阵乘法中导致严重病态。RCWA 在工程上较稳定的做法，是使用散射矩阵（scattering matrix）递推。

把“不稳定”说得更定量一点：若某层的纵向本征常数写成 $k_z = i\kappa$ ，则该层传播因子会同时含有

$$e^{+ik_z d} = e^{-\kappa d}, \quad e^{-ik_z d} = e^{+\kappa d}.$$

在简单传输矩阵连乘中，指数衰减项和指数增长项会被强行放进同一矩阵乘法链，结果条件数往往随层数和倏逝强度指数恶化。散射矩阵递推的价值就在于：它把入射、反射和透射通道分开追踪，避免把“物理上应当衰减的分量”和“数值上会炸掉的增长因子”直接混在同一个乘法序列里。

把整个结构视为一个多端口散射体，就有

$$\begin{bmatrix} \mathbf{b}_{\text{up}} \\ \mathbf{b}_{\text{down}} \end{bmatrix} = \mathbf{S}(\omega, \mathbf{k}_{\parallel}) \begin{bmatrix} \mathbf{a}_{\text{up}} \\ \mathbf{a}_{\text{down}} \end{bmatrix}, \quad (14.2)$$

其中 $\mathbf{a}_{\text{up}}$ 与 $\mathbf{a}_{\text{down}}$ 表示从上、下方入射的通道振幅， $\mathbf{b}_{\text{up}}$ 与 $\mathbf{b}_{\text{down}}$ 则是对应的出射振幅。 $\mathbf{S}$ 编码了在给定频率和面内 Bloch 波矢下，整个周期单胞如何把入射通道重排为出射通道。

若把式 (14.2) 写成块矩阵形式，

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix},$$

则 $\mathbf{S}_{11}$ 表示“从上方入射再向上方反射”的块， $\mathbf{S}_{21}$ 表示“从上方入射后向下方透射”的块；其余两块则对应从下方入射的情形。把这层物理含义紧贴在式 (14.2) 后面，是为了让后面的 $\mathbf{S}_{11}$ 、 $\mathbf{S}_{21}$ 不再像纯软件符号。

若只保留零级上传播与向下传播通道，那么最常见的两个标量散射参量就是

$$S_{11} = \frac{b_{\text{up},0}}{a_{\text{up},0}}, \quad S_{21} = \frac{b_{\text{down},0}}{a_{\text{up},0}}, \quad (14.3)$$

其中下标 0 表示零级衍射通道，通常要求端口振幅已经按单位功率通量归一化。这样处理后， $|S_{11}|^2$ 与 $|S_{21}|^2$ 才能直接对应反射率与透射率的主通道贡献。对存在多个衍射级次的周期结构，较稳妥的做法是分别统计各传播级次的功率通量，再把它们与同一个入射通量相除。明确这一步后，反射谱、透射谱和上下出光比例都可以对应到具体衍射通道的功率比，而不是只对应求解器输出的一条曲线。

对 PCSEL 来说，式 (14.2) 的威力在于它直接把“内部结构”和“外部辐射通道”连在一起了。于是，上下辐射不对称、偏振选择、法向出光窗口和反射谱中的共振尖峰，都能在同一框架下被读取。不过，开放散射仍是被动层面的问题，它不会自动给出激射结论（红线 #2，见第 2.10 节）。

Fourier factorization: 由界面连续性决定的离散化规则 RCWA 还有一个经常被忽略、却直接决定收敛质量的技术点: Fourier factorization。对具有介电常数跳变的周期结构, 像 εE 这样的乘积项并不能一律用“傅里叶系数逐项卷积”来处理。Li 的 factorization 规则要求根据场分量在界面处的连续性, 选择对 ε 还是对 ε^{-1} 构造 Toeplitz 卷积矩阵 [40]。记 $[\varepsilon]_{mn} = \hat{\varepsilon}_{m-n}$ 为对介电函数傅里叶系数构造的 Toeplitz 矩阵, $[\varepsilon^{-1}]$ 为对 $1/\varepsilon$ 同样构造的矩阵。若乘积中的场分量在界面处连续 (如 TE 偏振的电场切向分量), 可直接使用 Laurent 规则

$$\widehat{\varepsilon E_t} \approx [\varepsilon] \hat{E}_t; \quad (14.4)$$

若法向位移矢量更自然连续 (TM 偏振下电场法向分量在界面跳跃), 则应使用 inverse rule

$$\widehat{D_n} = \widehat{\varepsilon E_n} \approx [\varepsilon^{-1}]^{-1} \hat{E}_n. \quad (14.5)$$

用错规则的后果通常是伪收敛: Gibbs 震荡增强, 收敛速度从指数级降为代数级, 并出现系统性频移, 而非发散。以 GaAs/空气二元光栅 (折射率对比约 3.4:1) 为例, 错误的卷积规则可使 TM 偏振共振峰位产生 5–10 nm 的系统性偏移, 且把傅里叶阶数从 $N = 5$ 提高到 $N = 21$ 后偏移仍不消失。因此, 如果 RCWA 预测的共振波长与实验测量总有几纳米的固定偏差, 在排除工艺误差后, 应优先检查 factorization 规则是否与界面连续性一致。式 (14.4) 和式 (14.5) 想让读者记住的只有一件事: Fourier factorization 是由界面连续性决定的离散化规则, 而非可由提高截断阶数自动消除的数值细节。

14.4 被动谱、共振与极点: 谱线为什么会尖起来

很多读者第一次用 RCWA 时, 最直观的输出是一条反射谱或透射谱。看到尖峰或窄谷之后, 容易把它直接称为“高 Q 模”或“低阈值候选模”。这种说法并非总错, 但如果不说明其背后的极点结构, 就会显得很经验化。

从散射矩阵观点看, 开放共振模对应于复频率平面上散射矩阵的极点。通俗地说, 如果系统内部存在一个能量停留时间较长的开放模, 那么当实频扫描靠近该模的共振位置时, 谱线就会出现尖锐变化。谱峰、谱谷、相位跃迁、Fano 线形, 本质上都与极点及其和背景散射的干涉有关。

因此, RCWA 中看到尖谱线时, 真正应该问的不是“峰高不高”, 而是:

- 该特征是否随傅里叶阶数和分层数收敛;
- 它对应的是哪一支候选模, 和 PWEM/CWT 的哪一支模族能对上;
- 其线宽对应的被动损耗量级是否与其他方法一致;
- 它主要通过哪个辐射通道向外耦合。

只有把这些问题答清楚, 被动谱才能转化为器件设计信息。

14.4.1 随书实算：从 PWEM 频窗到 RCWA 反射谱

图14.2是脚本 `code/rcwa_square_hole/run_rcwa_square_hole.py` 用 `grcwa` 实际计算的结果。模型取正方晶格圆形空气孔, $r/a = 0.30$, 高折射率层 $n = 3.4$ 、厚度 $0.5a$, 上方为空气、下方衬底 $n = 1.5$; 请求 49 个 Fourier 级次, 求解器实际保留 45 个。被动扫描在 $\Omega = a/\lambda \approx 0.43$ 附近出现明显的 Fano 型反射结构, 且全频窗内 $|R + T - 1|$ 的最大误差小于 10^{-12} 。

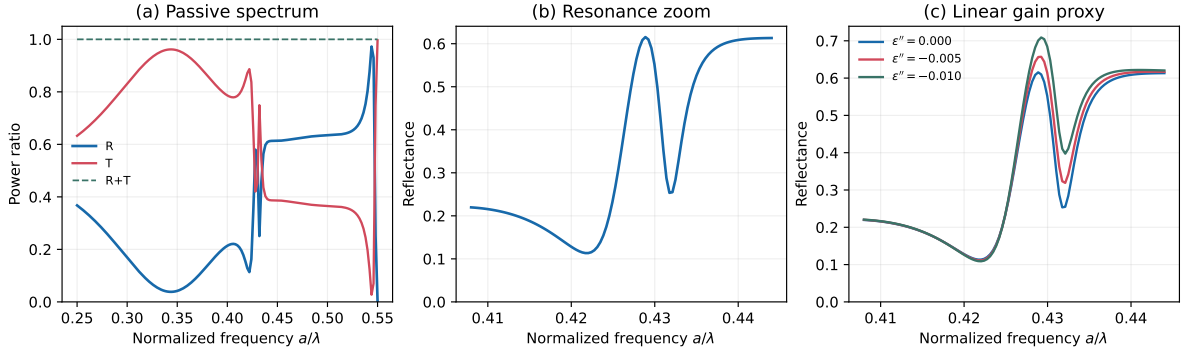


图 14.2: 正方晶格圆孔 slab 的可复现 RCWA 算例。左: 被动反射、透射及能量守恒检查; 中: 候选共振附近的定向细扫; 右: 在高折射率区域加入三组小负介电常数虚部后的线性增益代理。右图只说明给定实频谱怎样随小信号增益改变, 不是自洽激光阈值解。

这份算例把“PWEM 告诉 RCWA 去哪里看”变成了可执行接口: 先在候选窗口做无源宽扫, 再对谱线结构细扫, 最后才改变增益代理。复现时至少应同时提高 Fourier 阶数与实空间采样网格; 若峰位、线形或 $R + T$ 守恒随截断显著变化, 就不能把该谱线用于模式排序。

14.5 RCWA 怎样进入光学阈值分析

前面 第七章 已经明确区分了阈值增益和阈值电流。RCWA 在这里的作用, 是估计周期开放结构层面上的光学阈值增益, 而非直接给出 I_{th} 。最常见的做法, 是在有源层引入小信号增益对应的等效虚部, 例如写成

$$\varepsilon(\omega) = \varepsilon'(\omega) + i\varepsilon''_g(g), \quad (14.6)$$

然后扫描增益参数 g , 寻找使目标共振从净衰减变为恰好补偿损耗的条件。

如果只停在这一步, 读者容易把“扫到线宽为零”误听成“已经建立了完整激光阈值理论”。教材写法必须把这件事分层说清。对 PCSEL 而言, 至少应区分以下四个层次:

1. **冷腔开放共振层:** 先在无增益条件下找到复频率极点 $\tilde{\omega}_0 = \omega_0 - i\gamma_0$, 它描述的是开放系统如何储能与泄能。
2. **线性小信号阈值代理层:** 在给定小信号增益模型后继续追踪极点 $\tilde{\omega}(g)$, 并用 $\Im\tilde{\omega}(g_{th}) = 0$ 估计光学阈值增益。

3. **阈值处自洽有源场层**：真正的阈值问题要求场、极化和反转分布同时满足自洽条件，尤其当增益色散明显时，实部与虚部不能被拆成两个彼此独立的旋钮。
4. **阈上稳态激光层**：一旦超过阈值，增益饱和、空间烧孔、模式竞争和注入非均匀性都会进入主导地位，这时需要更完整的有源激光形式化，例如基于 Maxwell-Bloch 方程的稳态或时域求解。

本章 RCWA 所做的，严格地说位于第二层。它有用，但它仍是线性小信号阈值代理，不是阈上稳态激光解。

在这个过程中，可以从几种等价角度理解“达到阈值”：

- 被动极点在复频率平面上向实轴靠近；
- 某个共振的净线宽收缩到零；
- 在小信号线性框架中，目标模的总损耗被增益恰好补偿。

这一步得到的是 g_{th} 这类光学量。它对设计极其有用，因为它告诉我们结构本身对目标模和竞争模提出了多高的增益要求。但必须再次强调， g_{th} 不是 I_{th} 。从前者到后者，还需要材料增益模型、漂移-扩散输运、复合和热反馈。

为什么这只是线性小信号阈值代理 式 (14.6) 默认我们已经把有源区压缩成了一个“可调复介电常数”模型。只要这一模型仍停留在线性响应层面，那么我们实际上做的是：给开放冷腔模外加一个尚未饱和的增益补偿，并观察净衰减是否被抵消。这个判据与第七章的阈值条件相容，但它并没有告诉我们阈上光场会稳定在什么幅度、是否发生多模竞争、反转是否仍近似均匀，也没有告诉我们空间烧孔会不会把模式排序翻转。

增益色散为什么要求更谨慎 在许多工程扫描中，人们把 ϵ_g'' 当成独立参数扫动，而把 ϵ' 固定在冷腔值附近。这样做的目的是把问题简化成“先补偿损耗，再看哪一支模最早到零线宽”。作为前端筛选，这种近似经常足够。但若增益谱较窄、工作点离谱峰不远、或载流子引起的色散不可忽略，那么 $\epsilon'(\omega)$ 与 $\epsilon''(\omega)$ 由因果材料模型联系在一起，不能任意独立指定。实际有源介质是会同时改变频率位置、线宽和模场重叠的复响应系统，而非简单的“只加负虚部”。

阈值频率并不必然等于冷腔共振频率 把阈值理解成“在冷腔频率上把线宽调到零”也过于粗糙。更准确的说法是：随着增益参数变化，极点本身会沿复频率平面移动，

$$\tilde{\omega}(g) = \omega_r(g) - i\gamma(g). \quad (14.7)$$

阈值条件是 $\gamma(g_{th}) = 0$ ，而真正的阈值频率应写成

$$\omega_{th} = \Re \tilde{\omega}(g_{th}), \quad (14.8)$$

它不一定与冷腔共振频率 ω_0 完全相同。造成这种偏移的原因包括：增益色散引起的频率拉动、场分布随增益微调而改变，以及不同辐射通道和材料色散共同作用下的复频率漂移。对 PCSEL 来说，这种偏移尤其重要，因为 Γ 点附近的模式排序本来就可能很近，微小的频移就足以改变“哪一支模先达到阈值”的判断。

它与更完整 active lasing formulations 的关系 从更宽的理论坐标系看，本章 RCWA 增益扫描与更完整的有源激光稳态理论之间是“前端代理”和“自治主问题”的关系。前者把目标模的阈值判定简化成线性极点继续问题，因此很适合做候选模筛选和参数排序。后者则把 Maxwell 方程与极化、反转和饱和非线性联立起来，直接求解阈值处或阈上稳态激光解；典型代表包括 Maxwell–Bloch 方程的直接时域求解，以及稳态自治形式如 steady-state *ab initio* laser theory (SALT) [66, 67]。本书在这里不展开 SALT 的完整推导，但读者必须知道：RCWA 扫 $\varepsilon'' < 0$ 到零线宽，是为真正的有源激光问题提供一个计算成本较低、可解释、可筛选的线性近似入口，而非激光理论本体。

14.6 为什么 RCWA 特别适合做“单胞层面的设计判断”

一旦接受“RCWA 主要给光学阈值增益和辐射通道信息”，它在工作流中的位置就非常清晰了。它适合做以下几类判断：

第一，改变孔半径、占空比、刻蚀深度或双晶格扰动时，目标模与竞争模的被动损耗顺序怎样变化。

第二，上下辐射比例、不同偏振通道强弱和法向出光窗口如何随单胞几何变化。

第三，某种外延层微调是否主要通过改变纵向重叠影响辐射，而不是简单改频率。

第四，在后续有限器件仿真之前，哪些参数区间根本不值得继续深挖，因为开放周期单胞层面就已经显示目标模没有优势。

这也是 RCWA 的工程价值所在：它以远低于全尺寸三维时域仿真的代价，把大量“不可能成功”的单胞方案尽早筛掉。

例子

若两种孔形方案在 PWEM 中都显示存在相近的 Gamma 点候选模，但 RCWA 显示方案 A 的目标偏振通道在法向辐射上更有序、而竞争模辐射损耗更大，那么方案 A 更值得进入有限器件验证。此时即便还不能宣布“器件一定低阈值”，也已经完成了一次高价值的前端筛选。

14.7 RCWA 不能替你解决什么问题

RCWA 适用范围宽，但它解决的是“周期单胞开放散射”问题，而不是“完整器件工作”问题。这个边界必须明确。

首先，它默认平面内无限周期。因此，有限孔阵边缘泄漏、横向模包络和大面积器件的边界损耗不在默认描述范围内。若这些效应开始主导，必须转向第 15 章的 FDTD/FEM 或更高阶包络模型。

其次，它通常在小信号线性框架里处理增益。于是，增益饱和、空间烧孔和工作点非线性不会自动出现在结果里。

进一步说，即便把小信号增益继续到“零线宽”，也仍未自动得到阈上稳态解。若问题已经进入强色散、强模竞争或反转明显不再静止的区域，就必须转向更完整的有源激光稳态理

论，或至少用它们来校准 RCWA 前端筛选的可信边界。

再次，它不包含真实电流路径。均匀增益分布在 RCWA 中只是一个光学代理模型，而不是电注入器件的默认现实。

最后，它也不解决温升问题。很多结构在室温被动谱上看起来很好，一旦工作后温升导致折射率和增益谱漂移，目标模顺序就可能改变。这类问题必须交给热-电-光耦合章节处理。

易错点

把 RCWA 单胞反射谱里的最窄特征直接当作“最终器件最低阈值模式”，或把“扫描 $\epsilon'' < 0$ 直到线宽为零”直接当作“激光阈值理论已经完成”，都是典型的层级偷换。前者忽略了有限尺寸、注入非均匀和温升；后者忽略了增益色散、频率拉动、饱和和模式竞争。

14.8 2024–2026 前沿补充：modified RCWA 正在覆盖 full-PCSEL 结构

近年一个值得单独标注的进展是 modified RCWA 对 full-PCSEL / VCSEL 结构的扩展：通过双向散射矩阵与分层拼接策略，模型不再局限于“仅从上/下单向入射”的传统流程，而可以更直接地处理“腔内源同时向上传播”的器件场景 [71]。这类方法在共振波长、发散角等关键量上与 FDTD 给出了可比精度，同时保留了更高扫描效率。

这对本章方法边界的影响是双面的：

- 正面：RCWA 可承担更大比例的前端筛选任务，尤其适合大参数空间下的快速候选淘汰；
- 边界：当问题进入强非线性增益、强有限尺寸边缘效应或多物理场强耦合时，仍需回到全波与器件级模型完成回写链。

因此，modified RCWA 应被视为“前端能力增强”，不是“对后端验证层的替代”。

14.9 收敛与数值卫生：哪些检查是必须交代的

RCWA 对数值细节敏感。一个结果是否可信，不取决于图形是否平滑，而取决于它是否通过基本的数值检查。至少应当检查四件事。

第一，傅里叶阶数收敛。目标模的谱位、线宽、反射峰形和上下辐射分配，不能随着阶数增加继续大幅变化。

第二，纵向分层收敛。对刻蚀轮廓、锥形孔或多层外延而言，过粗的分层会把真实边界模糊掉。

第三，能量守恒或功率平衡检查。在无源条件下，反射、透射和吸收的总和应当满足相应守恒关系。

第四，结论对微小参数扰动的鲁棒性。如果某个“最好方案”只在一个极细的参数点上存在，而周围稍微变化就消失，那么它通常不是可靠设计。

这些检查之所以重要，不只是为了数值严谨，也是为了避免把求解器伪影误读成器件物理。

14.10 RCWA 在整条 workflows 中的准确位置

现在可以把 RCWA 放回全书主线。第十三章的 PWEM 给出候选模目录；本章的 RCWA 则对这些候选模做开放周期结构层面的精修。接下来第十五章的 FDTD/FEM 会把有限尺寸和更真实几何带回来，再由第十七章至第二十章把这些光学结果接进外延、增益、电注入和热耦合模型。

也就是说，RCWA 是一个关键的中间关口，而非终点。它回答的问题是：如果先只讨论单胞级开放光学，哪些方案值得继续投入更昂贵的器件级验证。把这个关口守好，后面的工作量会大幅下降；把它省掉，后面的全器件仿真往往会变成盲目试错。

回看本章

RCWA 的核心是散射矩阵视角：把周期分层开放结构看成一个输入输出系统，直接分析内部模式如何与外界辐射通道交换能量，而非某个软件界面操作。对 PCSEL 而言，它特别适合处理法向辐射、偏振通道、被动共振和光学阈值增益估计。但本章给出的阈值分析属于线性小信号阈值代理：它通过继续追踪开放极点来估计哪一支模最先补偿净损耗，而不是直接求出阈上稳态激光解。只有把这一层级和更完整的有源激光理论分清，RCWA 才能被放在正确位置上使用：作为“开放单胞精修器”和“后续器件仿真的前置筛选器”，而不是最终器件结论的替代品。

练习题

1. [基础] 用自己的话解释为什么 PWEM 之后自然需要 RCWA，而不是直接跳到器件级电热模型。
2. [基础] 说明散射矩阵方法为什么在数值上通常比简单传输矩阵更稳定。
3. [进阶] 讨论从 RCWA 得到光学阈值增益后，为什么仍然不能直接报告阈值电流。
4. [进阶] 解释为什么“在固定 ϵ' 下扫描负虚部直到零线宽”属于线性小信号阈值代理，而不是完整激光阈值理论。
5. [综合]（思路提示：先定层级，再列关键式）设计一套你认为最基本的 RCWA 收敛检查清单，并说明每一项各自防止什么误判。

延伸阅读

- 下一章将把视角从“周期单胞开放结构”推进到“有限尺寸真实器件”，也就是 FDTD 和 FEM 的领域。读完下一章后再回看本章，会更清楚两者的分工边界。

- 若想打牢开放周期结构的基础，可先对照 [28, 57] 中关于周期散射和开放通道的章节阅读。
- 若想补充 RCWA 的经典原始文献，可参考 [44, 40]。
- 若想理解本章线性增益扫描与更完整有源激光理论之间的关系，可进一步阅读 [66, 67]。

第十五章 FDTD 与 FEM 在 PCSEL 建模中的作用

读完本章，你能

- 理解为什么在 RCWA 之后仍然需要 FDTD 与 FEM，以及它们分别擅长解决哪类 PCSEL 问题。
- 掌握几何设置、边界条件、激励源、结果提取和收敛检查的基本逻辑，而不是停留在软件界面操作层面。
- 清楚地区分冷腔分析、带损耗分析和加入增益后的分析在数学意义上有什么不同。

读前准备 建议已掌握 第七章、第八章和第十四章 中关于开放边界、损耗、远场和阈值的基础概念。

本章结构

本章所处模型层级：L1→L2

上一章的 RCWA 解决的是周期单胞开放结构问题，但真实 PCSEL 最终总要回到有限尺寸器件。边缘、有限孔阵、非理想窗口、局部几何细节和复杂外延层堆栈，都会把问题推向更一般的全波求解。本章先解释为什么 FDTD 和 FEM 会在这里登场，再分别说明它们的控制方程和数值哲学，随后讨论边界条件、激励源、共振参数提取与远场计算，最后严格说明它们能给出什么、又不能单独给出什么。

15.1 为什么有限尺寸一旦重要，全波方法就不可回避

对 PCSEL 来说，使单胞方法不再充分的，不只是“器件变大了”，而是有限尺寸带来的物理变了。无限周期模型里，模式由 Bloch 波矢标记；有限器件里，模式必须同时满足边缘终止、横向包络和开放边界条件。于是，边缘泄漏、横向模式竞争、注入窗口截断和热透镜效应都会显式出现。

此时，如果还只依赖周期单胞方法，就会把有限器件里决定模式排序的一部分物理漏掉。FDTD 与 FEM 的意义在于，它们不再默认平面内无限周期，而是直接对给定有限几何上的

Maxwell 方程做求解。也正因为如此，它们成为从“单胞光学”走向“真实器件光学”的关键台阶。

直觉框

RCWA 擅长告诉你”无限周期单胞在开放边界下会怎样散射”；FDTD 和 FEM 则开始回答”这个真实大小、真实边界、真实窗口的器件会怎样支持模式并向外辐射”。这两层都属于光学，但问题层次已经不同。

表 15.1: FDTD、FEM 与 RCWA 在 PCSEL 分析中的典型分工。

维度	FDTD	FEM	RCWA
基本思路	时域推进，直接求解 Maxwell 方程的瞬态响应	由频域弱形式离散，适合复杂边界和本征值问题	平面内 Fourier 展开并沿纵向分层传播，适合周期结构
适用问题	有限尺寸阵列、宽带响应、近远场提取	冷腔本征模、复频率、局域场与多物理场耦合	能带初筛、反射/透射谱、周期单胞参数扫描
对周期结构的处理	可以处理，但计算域通常较大	可以处理，但需显式设置 Bloch 或开放边界	最自然，通常作为周期单胞筛选的首选方法
对复杂几何的处理	正交网格下容易引入 staircase 误差	非结构网格适应性较好，几何保真度高	可处理分层剖面，但对分层近似和周期假设敏感
计算代价	内存和时间开销通常最高，三维大尺寸问题尤其明显	单次求解较重，但目标频点或目标模分析效率较高	单次扫描较快，适合较大范围参数搜索
主要输出	时域衰减、宽带响应、近场和远场	本征频率、品质因子、模场分布、局域增益重叠	反射/透射、衍射级次、耦合强度趋势
在 PCSEL 工作流程中的角色	验证有限孔阵、边缘泄漏和实际出光表现	比较候选模、阈值相关损耗与多物理场耦合机制	早期筛选晶格常数、填充因子和扰动参数
局限性	网格和时间步长受限，长时间衰减问题成本较高	模型搭建和网格质量要求较高	不适合直接描述有限尺寸器件与强非周期扰动

对 PCSEL，常见组合流程是：先用 PWEM / RCWA 做带结构和单胞参数初筛，再用 FEM 提取候选模的复频率与场分布，最后用 FDTD 检查有限尺寸、边缘泄漏和远场表现。

表15.1系统对比了 FDTD、FEM 和 RCWA 三种方法的关键特征。对 PCSEL 建模而言，这三种方法并非相互排斥，而是各有舒适区：RCWA 适用于无限周期结构的快速扫描和初步设计；FDTD 适用于有限尺寸阵列的宽带响应和瞬态过程；FEM 则适用于复杂几何结构的高精度本征模分析。在实际工作流程中，通常会先用 RCWA 或 CWT 筛选候选结构，再用 FDTD 或 FEM 进行精细验证，最后结合 TCMT 进行物理诠释。

15.2 同一个 Maxwell 方程，为什么会分化出两种数值哲学

无论 FDTD 还是 FEM，最终处理的都是 Maxwell 方程。若写成频域形式，可用

$$\nabla \times \mu_r^{-1} \nabla \times \mathbf{E} - k_0^2 \varepsilon_r \mathbf{E} = i\omega \mu_0 \mathbf{J}, \quad (15.1)$$

其中 $k_0 = \omega/c$, $\mathbf{J}$ 是外加源项。FEM 直接对式 (15.1) 的弱形式进行离散；FDTD 则回到时域，对 Maxwell 旋度方程逐时步推进。两者的根本差异，是看问题的方式不同，而非“谁更高级”。

FDTD 通过时间推进观察场如何建立、衰减和传播。它适合宽带激励和瞬态过程，因为一次时域仿真可以在后处理时取出多个频率分量。

FEM 则更像是频域定点瞄准。它适合在指定频率附近做高精度场分布分析，也适合直接构造本征值问题来求开放腔的复频率与模场。

概括地说，FDTD 侧重时间演化和宽带响应，FEM 侧重指定频率或指定模态附近的精细求解。为了帮助读者直观理解 FDTD 仿真的完整配置，图15.1展示了一个典型的三维 PCSEL FDTD 仿真区域设置示意图，包括 PML 吸收边界层、PCSEL 有源层结构、PhC 孔洞阵列、偶极子激励源、时间监视器、频谱监视器以及近远场变换面的位置关系。图中参数（例如 PML 厚度）应视为 [工程经验] 起点，需由收敛扫描确认，不是固定标准 [8, 64]。

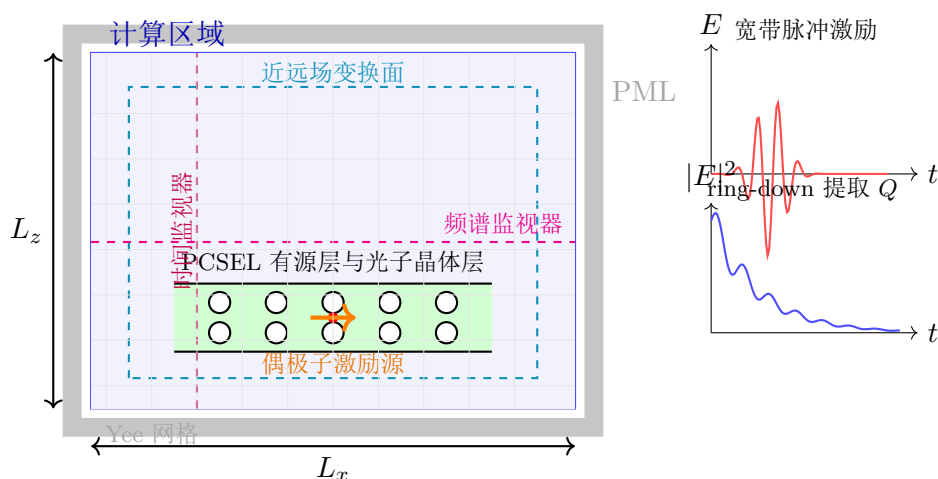


图 15.1: 三维 FDTD 仿真区域的标准配置。(a) 整体布局：PML 吸收边界层包围整个计算区域，PCSEL 有源层位于中心，周围是空气或包层介质；(b) 激励源配置：偶极子源放置在有源区内，模拟自发辐射初始激发；(c) 监视器布置：时间监视器记录特定点的场随时间演化，频谱监视器提取共振频率和 Q 值，近远场变换面用于计算远场辐射图案。关键参数如“PML 厚度 12–20 层、PML 与器件间距约半个波长、网格分辨率 $\leq a/10$ ”均应作为 [工程经验] 初值，需通过误差门限校核 [64]。

图15.1展示的给出的不只是几何布局，也包括边界、源和监视器之间的基本约束：PML 确保开放边界条件正确实现，监视器位置决定了能否准确提取共振参数，而激励源的选择直接影响哪些模式会被有效激发。对于 PCSEL 这类高 Q 值共振系统，合理的仿真配置不能由单纯细化网格替代。

15.3 FDTD 的强项：宽带、瞬态与模建立过程

FDTD 的核心是把时域 Maxwell 方程写在交错的 Yee 网格上，用电场和磁场在半时间步上交替更新。该方法适用于 PCSEL 中的三类问题。

第一，它天然适合宽带问题。用一个短脉冲或调制源激励器件，就可以在一次仿真中得到一段频谱中的响应。这对于寻找候选共振频率和观察多模并存较方便。

第二，它能够直接看模如何建立与衰减。如果想判断某个共振是快速泄漏还是长寿命，时域包络会给出直观的图像。

第三，它容易和时域监视器结合，提取瞬态远场、延迟响应或短时干涉现象。这对某些动态调制或脉冲工作问题尤其有用。

但 FDTD 的代价也很明确。为了保证精度，它需要足够细的空间网格和满足稳定条件的时间步长；三维器件一旦横向尺寸变大，存储和时间成本会迅速上升。PCSEL 又往往同时具有微细孔结构和较大横向尺寸，这是 FDTD 容易变得昂贵的原因。

15.4 FEM 的强项：复杂几何、本征模与局域高精度

FEM 的优势在另一边。它通过构造弱形式，把几何复杂性和材料非均匀性更自然地吸收到网格与形函数中。对于 PCSEL 来说，这带来几个实际好处。

第一，复杂孔形、锥形刻蚀、圆角、分段金属和非规则边界都能更自然地表示，而不需要被强行贴合到正交网格。

第二，FEM 适合做本征频率求解。通过加入 PML 等开放边界，直接求解复本征频率，就能得到冷腔共振频率和辐射 Q 。

第三，FEM 在局域场细节分析上往往更强。若你需要精确比较量子阱区、掺杂区、金属邻近区的场重叠，或者精确获取某一支模式的近场矢量结构，FEM 常常更适用。

当然，FEM 也不是没有代价。它的前处理通常更重，对网格质量和求解器参数更敏感；而且如果目标是扫很宽的频段，频域逐点扫描会比 FDTD 更耗时间。

FDTD 与 FEM 的区别，不是“一个近似，一个精确”。两者都是数值离散方法，都需要收敛检查和边界审计。主要区别在于：FDTD 的误差更多地表现在时空离散和长时间积分上，FEM 的误差更多地表现在网格质量、弱形式离散和开放边界本征值的谱污染上。

15.5 几何设置为什么不能只靠“画出来”

全波方法里，几何设置从来不是一个机械步骤。对 PCSEL 而言，几何模型至少同时承载三类信息：

- 晶格和孔形决定的面内反馈与辐射通道；
- 外延层堆栈决定的纵向约束和损耗重叠；
- 有限孔阵、边缘吸收和窗口定义决定的横向模式边界。

如果只把“周期孔阵”画出来，而不把真实外延层、接触层和边缘终止方式带进去，那么算出的模式常常只是“美观的理想光学对象”，而不是器件会真的支持的模式。也就是说，几何是否足够真实，直接决定你在算“物理问题”还是在算“几何练习题”。

这种纵向误差在 PCSEL 上尤其严厉，因为其外延层既参与光学，也参与后续电学和热学接口。光场哪怕只是在纵向上偏移一点，量子阱重叠和掺杂吸收就都可能随之改变。

15.6 边界条件：PML 不是附件，而是模型的一部分

只要求解开放器件，边界条件就是结果的一部分，不是外加包装。对 FDTD 和 FEM 来说，最常见的开放边界处理都是 PML。它的目标是在有限计算域内尽量模拟“波向无穷远走掉而不反射回来”。

但初学者容易把 PML 当成一个固定配件，仿佛加上去就自然没问题。实际上，PML 厚度、距离器件的间隔、坐标拉伸参数和离散质量，都会直接影响共振频率、 Q 值和远场。若 PML 太近，器件近场和倏逝场会直接碰到吸收层；若 PML 太薄或参数不当，又会产生伪反射，污染高 Q 模和远场主瓣。

因此，对任何声称给出辐射 Q 或远场的 FDTD/FEM 结果，都应该追问：边界离器件多远？PML 厚度怎么选？结果对这些参数是否收敛？如果这些问题答不出来，那么结果的可信度就不够。

表 15.2: FDTD/FEM 中开放边界设置的最小量化检查表。数值是第一次建模时可执行的起点，而非普适常数。

项目	建议起点	通过门槛
PML 与器件最近距离	至少取 $d_{\text{PML}} \gtrsim \lambda_0/n_{\text{eff}}$ ；若存在强倏逝尾则进一步外推	共振频率、 Q 和远场主瓣对该距离变化小于预设容差
PML 厚度	至少 $L_{\text{PML}} \gtrsim \lambda_0/(2n_{\text{min}})$ ，且不少于 8–12 个网格层	继续增厚时，反射谱、 Q 与远场变化应进入平台区
PML 吸收强度 / 拉伸参数	先用默认稳定值，再单因子扫描使理论边界反射压到 10^{-6} 量级以下	不应出现随吸收参数变化而大幅漂移的伪模或伪反射

高 Q 腔的 PML 特殊要求。对 $Q > 10^4$ 的高 Q 腔（如 BIC 或准-BIC 模），常规 PML 设置可能导致显著的数值误差：

- **目标反射率：**应达到 $R_{\text{PML}} < 10^{-8}$ 而非普通的 10^{-6}
- **PML 厚度：**建议增加到 λ_0/n_{min} 以上，或使用渐变 PML(graded PML)
- **距离要求：**倏逝尾场必须在到达 PML 前衰减到峰值的 10^{-6} 以下
- **验证方法：**扫描 PML 参数（厚度、吸收强度、距离）时， Q 值变化应小于 1%

若 PML 设置不足，高 Q 模的数值 Q 值会被人为限制在 10^4 – 10^5 量级，无法反映真实物理极限。

更具体地说，若要报告一个可追溯的开放边界结果，至少应给出三条收敛句：第一，PML 与器件距离从 $d_{\text{PML}}^{(1)}$ 增加到 $d_{\text{PML}}^{(2)}$ 时，目标模频率与 Q 的变化量；第二，PML 厚度从 $L_{\text{PML}}^{(1)}$ 增加到 $L_{\text{PML}}^{(2)}$ 时的变化量；第三，远场主瓣角度与主瓣半高全宽是否对监视面位置稳定。若这三条信息没有出现，那么“边界已经足够远”通常只是主观判断。

若使用 FDTD，还必须同时给出空间网格与时间步长的起点。对三维 Yee 网格，稳定性的最小 Courant 条件可写成

$$\Delta t \leq \frac{1}{c} \left(\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} + \frac{1}{\Delta z^2} \right)^{-1/2}. \quad (15.2)$$

工程上通常再乘一个安全因子 $S < 1$ ，例如取 $\Delta t = S\Delta t_{\max}$ 、 $S \sim 0.5$ （[工程经验]）。同时，空间网格也应至少满足

$$\Delta_{\min} \lesssim \min \left(\frac{\lambda_0}{10n_{\max}}, \frac{r_{\text{feature}}}{5} \right),$$

其中 r_{feature} 代表最小孔边圆角、最窄脊宽或最薄有效层厚度。对量子阱、金属邻近层和陡侧壁附近，还应再做局部加密。这样写的意义是提醒：时间步长不是独立自由度，它被最细网格直接锁定，而非要求读者死记一个数字；而最细网格又必须由波长分辨率和最小几何特征共同决定。只要孔边、薄量子阱附近或金属邻近层要求更细的 $\Delta x, \Delta y, \Delta z$ ， Δt 也必须同步减小。否则，即便频谱看上去平滑，结果也可能只是数值色散和时间离散误差共同伪装出来的“稳定解”。

15.7 激励源的选择，其实是在决定你想看什么

FDTD 和 FEM 并不自动知道你关心哪一支模式。源的放置方式、方向、时域或频域特性，实际上是在告诉求解器“请优先激发哪些对称性和哪些频段”。

如果目标是搜寻候选共振频率，宽带脉冲源通常很合适；如果目标是选择性激发某种偏振或某类对称性模式，则应使用与目标模式近场结构匹配的电流源、偶极源或端口源；如果目标是得到稳态散射响应，则频域连续波源更合适。

这一步之所以重要，是因为错误源设置并不会总是导致仿真报错，而更常见的是：它悄悄让你只看到某几支容易被激发的模式，误以为系统里就只有这些模式。对接近简并的 PCSEL 候选模来说，这种误导尤其危险。

若目标是比较一对近简并正交偏振 Gamma 点模，而激励源只沿单一方向放置且恰好与其中一支模式强耦合，那么仿真中另一支模式可能几乎不出现。此时若不更换源方向或改用本征模分析，就可能错误得出“器件天然单偏振”的结论。

15.8 怎样从时域或频域结果里提取共振频率与品质因子

对冷腔分析来说，最常提取的两个量是共振频率和品质因子。FDTD 中，常见做法是在脉冲激励后观察某个监视点的场随时间衰减，再对包络做指数拟合，从而得到衰减率 γ ，进而写成

$$Q = \frac{\omega_0}{2\gamma}. \quad (15.3)$$

这个一维算例由 `code/fp_decay_q/run_fp_decay_q.py` 生成，是检查 Q 定义、拟合窗口和单位换算的校准件：储能按 $U(t) \propto \exp(-t/\tau_U)$ 衰减时， $Q = \omega_0\tau_U$ 。它没有模拟 PCSEL 的三维有限阵列，但能在进入 FDTD 前排除“把场幅寿命和能量寿命混成两倍”这类基础错误。输出的 CSV、JSON 和图都保存在 `code/fp_decay_q/output/`。

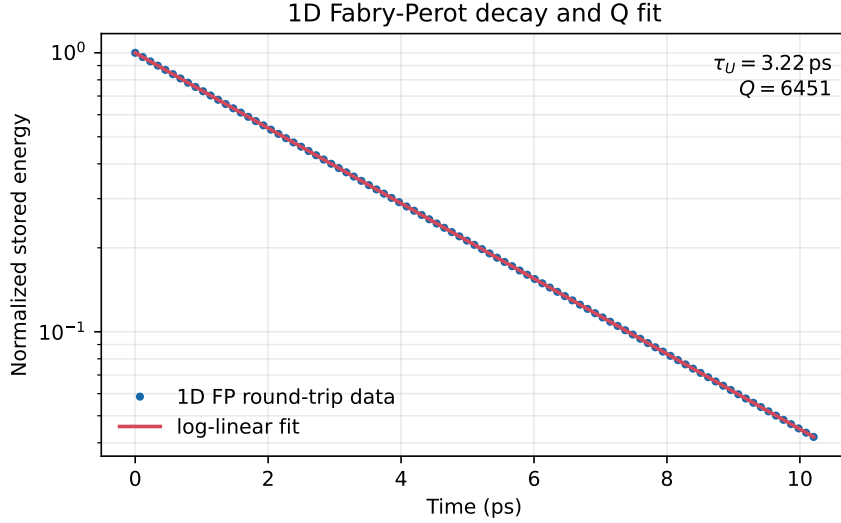


图 15.2: 一维 Fabry-Perot 腔的可复现时域衰减标定。随书脚本逐回程传播储能, 取 $\lambda_0 = 940 \text{ nm}$ 、 $n = 3.4$ 、 $L = 5 \mu\text{m}$ 、 $R_1 = R_2 = 0.985$ 和 $\alpha_i = 5 \text{ cm}^{-1}$, 对 $\ln U(t)$ 做线性拟合得到 $\tau_U = 3.22 \text{ ps}$ 与 $Q = 6451$ 。

模拍频对 Q 提取的影响。当结构支持两个或多个近简并模（频率差 $\Delta\omega \lesssim \gamma$ ）时，时域衰减信号会出现模拍（mode beating）现象：

$$E(t) \sim A_1 e^{-\gamma_1 t} e^{-i\omega_1 t} + A_2 e^{-\gamma_2 t} e^{-i\omega_2 t},$$

此时简单的单指数拟合会给出错误的 γ 和 Q 值。正确处理方法包括：

- 谐波反演（**harmonic inversion**）：将信号拟合成多个指数衰减项的叠加
- 滤波隔离：在频域用窄带滤波器分离出单个模式后再做时域拟合
- 本征模求解：直接使用 FEM 或 FDTD 的本征模求解器获取各模的复频率
- 长时间窗：确保时间窗长度 $T \gg 2\pi/\Delta\omega$ 以分辨拍频周期

若忽略模拍效应，近简并模的 Q 值可能被高估或低估一个数量级以上。

FEM 中，则常通过开放本征值求解直接得到复频率

$$\tilde{\omega} = \omega_0 - i\gamma, \quad (15.4)$$

再由式 (15.3) 得到 Q 。两条路在理想条件下是一致的，但它们对数值误差的敏感方式不同。因此，若条件允许，让时域和频域对同一几何给出一致的趋势，是有价值的互证。

这里再次要强调：冷腔频率和冷腔 Q 是重要输入，但仍不是器件工作点本身。把它们直接当作激射频率和工作阈值，是层次上的错误。

15.9 近场、远场与模场：怎样读出可用于后续建模的量

全波仿真会产生大量场图和谱图，但关键是明确应从中提取哪些量。

首先是模场分布。它是为了和后续的量子阱重叠、损耗层重叠、偏振结构和边缘泄漏联系起来，而非为了“看起来像不像文献图”。

其次是远场。远场与偏振可工程化是 PCSEL 的主要性能指标之一，因此必须确认远场主瓣、旁瓣和角宽是否对网格、边界和监视面位置稳定。

再次是功率流和损耗分解。如果可能，应分辨功率是主要向上辐射、向下辐射，还是从侧边泄漏。否则即便 Q 给出来了，也难以理解损耗来源。

最后是与器件层的接口量，例如与量子阱区的重叠因子、与重掺杂层或金属层的场重叠。这些量往往比单一“最大场强”更重要，因为它们直接进入第十七章至第二十章的器件模型。

若要把近场变成可比较的远场，至少要记住角谱或近远场变换的最小形式。对监视面上的切向场，远场复振幅可形式化写成

$$\mathbf{F}(k_x, k_y) \propto \iint_{S_{\text{mon}}} \mathbf{E}_t(x, y) e^{-i(k_x x + k_y y)} dx dy, \quad (15.5)$$

其观察角满足

$$k_x = n_{\text{out}} k_0 \sin \theta \cos \phi, \quad k_y = n_{\text{out}} k_0 \sin \theta \sin \phi. \quad (15.6)$$

式 (15.5) 和式 (15.6) 的意义是提醒两个常见事实：第一，远场角谱本质上是监视面切向场的二维傅里叶变换，而非要求读者手写近远场变换程序；第二，若监视面过近、截面过小或含有强反应近场，角谱会被直接污染。

易错点

“谱污染”在这里通常是指你以为看到的是器件的远场或本征谱，实际上看到的是边界反射、监视面截断、时间窗不够长或错误激励源带来的伪结构，而非指图像不够平滑。只要改变监视面位置、时间窗或 PML 设置后主结论就翻转，就应先怀疑谱污染，而不是先怀疑新物理。

15.10 冷腔分析和带增益分析的区别

对初学者来说，一个特别容易混淆的问题是：在全波方法里“加一点损耗或增益”看起来只是改材料参数，为什么物理上层级却变了？原因是冷腔分析求解的是被动线性系统，而加入增益后，系统开始逼近一个有源开放系统。

如果只是加入一个均匀、小信号的虚部增益并保持线性，那么你得到的仍然只是“光学阈值趋势”或“复频率如何变化”的信息；一旦再加入载流子分布、增益饱和和温升反馈，问题就不再是单纯电磁本征问题，而变成光-电-热耦合问题。也就是说，数学上的变化不仅是“参数值改了”，而是“方程组和闭合条件都改了”。

很多工作中会把“冷腔全波仿真 + 均匀增益补偿到零衰减”的结果直接称为“器件阈值仿真”。这种说法是不严谨的。它最多是光学线性阈值估计，不是电注入器件级阈值电流预测。

表 15.3: 冷腔全波分析与带增益线性分析的结论边界。

层级	主要输入/输出	不能越级宣称什么
冷腔全波	输入为实介电常数、开放边界与有限器件几何; 输出为复频率、 Q 、近场、远场、损耗分解	不能直接宣称阈值增益、阈值电流或工作频率
小信号带增益全波	在冷腔基础上加入均匀或给定分布的小信号增益代理; 输出为复频率趋向、线性阈值趋势、竞争模排序变化	不能直接宣称阈上稳态、增益钳位、热稳定工作窗口
电热光耦合工作点	输入为注入、热边界、增益压缩与回写模型; 输出为阈值电流、工作频率漂移、模式稳定性	仍需实验回验才能宣称器件最终性能

15.11 收敛检查：哪些量必须扫，哪些量必须复核

全波结果是否可信，不能靠肉眼判断。最基本的检查至少包括四类。

第一类是空间离散收敛。FDTD 需要扫描网格尺寸；FEM 需要扫描网格密度和局部加密策略。

第二类是边界收敛。PML 厚度、边界距离和吸收参数必须被扫过，而不能只选一组看上去顺眼的数。

第三类是时间或求解容差收敛。FDTD 中要检查时间步长和总仿真时长；FEM 中要检查求解残差、特征值搜索区间和迭代容差。

第四类是结果提取方式收敛。不同监视点、不同拟合窗口、不同近远场变换面，不应导致结论大幅波动。

若这四类检查不能通过，即使场图平滑，也不应进入器件级推断。

对本科生或第一次上手的研究者，更可执行的说法是：至少固定一份最小收敛账本。FDTD 至少记录 Δx 、Courant 因子、总时间窗 T_{sim} 、PML 厚度与距离；FEM 至少记录网格最小尺寸、单元阶次、特征值搜索区间、残差容差和 PML 设定。只有账本存在，后续“结果为什么可信”才不是回忆题。

15.12 本章在整条器件主线中的位置

在全书结构中，本章连接周期单胞光学与有限器件光学。第十四章 给出开放周期单胞层面的结论，而本章把问题推进到有限尺寸全波层面。接下来 第十六章 会总结为什么单一方法不能覆盖所有问题，并给出推荐的互证工作流。再往后 第十七章至第二十章 则把这些光学结果转交给外延、增益、电注入和热耦合模型，进入器件级模型。

因此，FDTD 和 FEM 在书中的角色，是把“候选模光学”转化为“真实器件光学”的关键桥梁，而非简单增加两个求解器名称。

回看本章

FDTD 与 FEM 都是在有限尺寸、开放边界下求解 Maxwell 方程的全波方法，但它们擅长的对象不同。FDTD 长于宽带和瞬态，FEM 长于复杂几何和高精度本征模分析。对 PCSEL 而言，它们的主要价值在于把无限周期单胞结论推进到真实有限器件，并输出与外延、增益、电注入和热耦合模型兼容的场分布与损耗信息。它们适用对象不同；二者都不能替代器件级多物理场回写，也不能单独给出最终工作结论。

练习题

1. [基础] 解释为什么在 RCWA 之后，有限尺寸器件分析不能被简单省略。
2. [基础] 比较 FDTD 与 FEM 在“宽带扫频”“复杂孔形”“高 Q 本征模”和“瞬态建立过程”四个问题上的优劣。
3. [进阶] 设计一套用于 PCSEL 冷腔 Q 提取的收敛检查方案，至少包含网格、边界和结果提取三类参数。
4. [综合]（思路提示：先定层级，再列关键式）给出一个“冷腔全波结果很好看，但器件最终仍会失败”的可能场景，并说明缺失了哪一层物理。

延伸阅读

- 下一章将把 CWT、RCWA、FDTD 和 FEM 统一放到同一个工作流里比较。读完下一章后，再回看本章会更清楚每种方法分别对应哪一类问题。
- 若想补充数值方法基础，可参考 [64, 27] 中关于时域方法、有限元方法与开放边界处理的章节，并结合 [8] 回看 PML 的来历。

第十六章 CWT、RCWA、FDTD、FEM 的比较与互证

读完本章，你能

- 清楚区分 CWT、RCWA、FDTD、FEM 各自解决的物理问题，而不是只记住它们的软件标签。
- 建立多方法互证意识，理解为什么单一方法几乎不可能覆盖 PCSEL 的全部核心问题。
- 掌握一条从概念设计到器件级验证的推荐工作流，并知道每一层输出该如何交给下一层。

读前准备 建议已掌握第十章、第十一章、第十四章和第十五章中关于半解析模型、开放周期结构和全波有限器件求解的内容。

本章结构

本章所处模型层级：L1→L3

到这里为止，书里已经出现了四类常用工具：CWT、RCWA、FDTD 和 FEM。初学者常见错误，是误以为其中一种就足够覆盖全部问题，而非不会用它们。本章的任务就是把这四种方法放到同一张地图上，先比较它们分别擅长什么，再说明它们的互证原则，最后给出一条真正适合 PCSEL 的研究工作流。这样安排是为了避免结论层级混乱，而非为了增加方法数量。

16.1 为什么单一方法几乎注定不够

PCSEL 同时是三种对象：它是周期开放电磁系统，是半导体有源腔，也是带真实注入与热分布的器件。任何一种方法，只要天然擅长其中一层，就通常会在另外两层变弱。

CWT 擅长把复杂物理压缩成可解释的耦合通道，因此适合建立直觉、做灵敏度分析和前期筛选；但它的代价是近似较强，不能天然承担高保真几何与开放边界的全部负担。

RCWA 擅长周期开放单胞，特别适合研究辐射通道、偏振和开放共振；但它默认的是无限周期单胞，不天然处理有限器件边缘。

FDTD 和 FEM 擅长真实有限器件全波分析，能把边缘、复杂层栈和真实几何带回来；但它们计算代价更高，也通常不像 CWT 那样自动给出清晰的机制分解。

于是，单一方法不够是因为 PCSEL 的问题本来就不是单层问题，而非因为“每种方法都不行”。方法必须分层合作，才和物理对象的层次相匹配。

直觉框

成熟的 PCSEL 建模不应寻找“最强求解器”，而应建立一条从粗到细、从机理到验证、从光学到器件的层级链。方法之间是接力关系，而非竞赛关系。

16.2 先按“问题类型”而不是“方法名字”来比较

最有用的比较方式，是先问“我现在究竟在回答什么问题”，而非问“哪个方法更准”。如果问题是候选模目录、主导耦合通道和几何灵敏度，那么 CWT 最合适；如果问题是开放周期单胞的辐射、反射谱和光学阈值增益趋势，RCWA 最合适；如果问题是有限器件边缘、近远场、真实 Q 和复杂几何，FDTD/FEM 更合适；如果问题进一步变成阈值电流、热稳定性和模式切换，则必须把这些光学结果交给器件模型。

也就是说，方法比较的第一原则是“按问题选择方法”，而不是“按方法反推问题”。这条原则一旦建立，许多常见争论其实可被重新表述。比如，“RCWA 准还是 FDTD 准”这种问题本身就不完整，因为两者默认解决的问题层级并不一样。

表 16.1: 几类常用方法的典型计算量级与适用尺度。表中时间与自由度属于 [本书算例]，仅用于排程直觉，不用于跨软件绝对比较。

方法	典型自由度	典型对象尺度	单次求解量级
PWEM	$\sim 10^2-10^3$	二维单胞、被动 无限周期	秒级到分钟级
CWT	$\sim 10^0-10^1$	少数主导波与 参数扫描	秒级
RCWA	$\sim 10^3-10^4$	三维层状周期 单胞	秒级到分钟级
3D FDTD	$\sim 10^7-10^8$	数十周期有限 孔阵、宽频瞬态	小时级到天级
3D FEM	$\sim 10^6-10^7$	有限器件、本征 模或窄频响应	小时级到天级

benchmark 前提（用于解释表 16.1）：以单机 16 核 CPU、64 GB 内存、双精度求解为参考，RCWA 与 FDTD/FEM 的网格/谐波截断分别取“达到章节内误差门限”的最小设置；若改用 GPU、不同求解器实现或更严格收敛标准，耗时可变化一个数量级以上。因此表中“秒级/小时级/天级”应读作 [工程经验] 区间，而非方法学常数。RCWA 资源量级对截断阶数与因子化规则的敏感性，可参见经典收敛文献 [45, 40]。

把表 16.1 放在这里的目的是，说明工作流必须分层：不同方法面对的物理问题不同，计算代价也相差数个数量级，而非制造“谁更高级”的错觉。若把本来应由 PWEM 或 RCWA 完成的候选筛选全部交给 3D FDTD/FEM，常见后果是更慢，也更难定位错误，而非更严谨。

为了帮助读者更直观地理解方法分工，图16.1给出一个六维雷达 [教学示意] 图。图中多边形仅表达“优势方向”和“取舍关系”，不代表跨软件、跨硬件可复现实测分数；若要做发布级定量比较，必须同时给出 benchmark 四元信息（硬件、求解器版本、几何与边界、误差门槛）以及原始数据表。

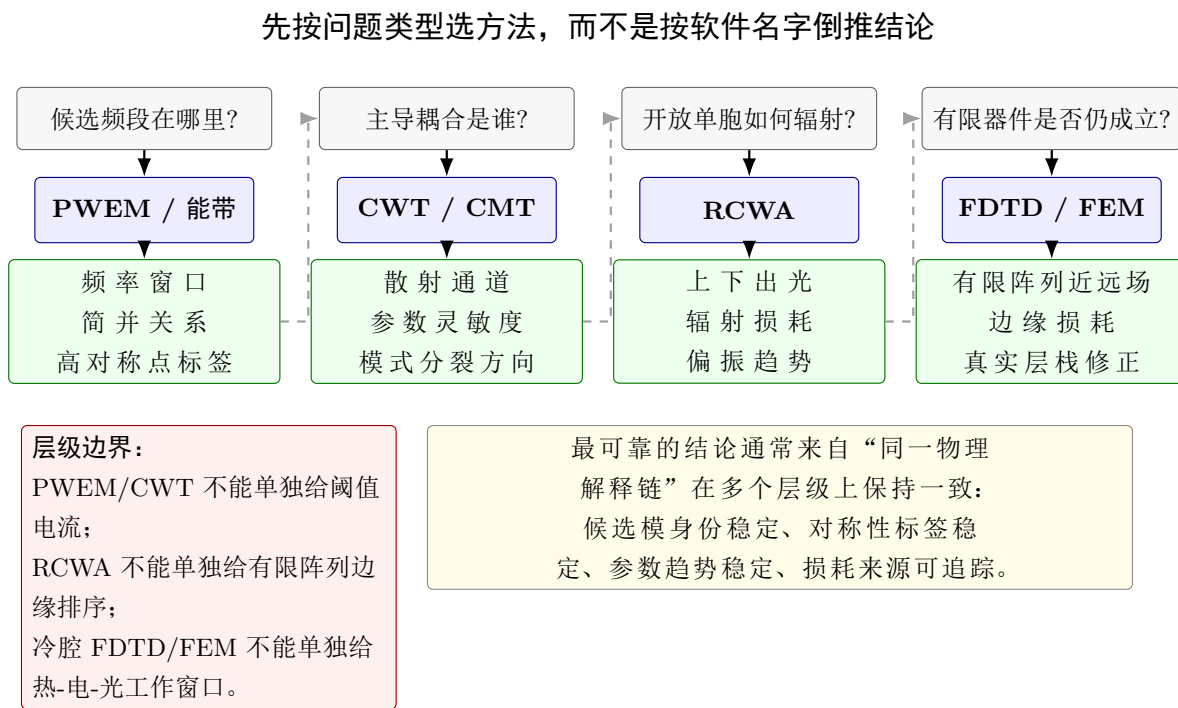


图 16.1: PCSEL 数值方法的分工地图。图中强调从问题类型进入：能带/PWEM 给候选目录，CWT/CMT 给机理压缩，RCWA 给开放周期单胞辐射，FDTD/FEM 给有限器件与真实层栈验证。它是教学示意图，不代表跨软件 benchmark 分数。

16.3 四种方法各自最擅长什么

为了把分工说得更清楚，可以用一句话概括四种方法的适用场景。

CWT 最擅长回答：“为什么这类几何扰动会主要改这个耦合通道，而不是另一个？”

RCWA 最擅长回答：“这个周期单胞在开放边界下会怎样把能量耦合到各个辐射通道？”

FDTD 最擅长回答：“这个有限器件在宽频和瞬态条件下，实际会怎样建立、衰减和辐射？”

FEM 最擅长回答：“这个复杂几何、复杂层栈的有限器件在某个频率或某支模式附近，精确场分布和复本征值是什么？”

如果把四者的适用场景混在一起用，就会出现常见误判。例如：用 CWT 试图给出精确远场角宽，用 RCWA 试图给出有限器件边缘模泄漏，用 FDTD 的一组冷腔结果直接断言器件阈值电流。这些都是问题与方法没有对齐，而非方法本身的问题。

表 16.2: 各方法最常见的“明确不适用”场景。表中“不适用”指不应让该方法单独承担最终结论。

方法	不应单独承担的结论	原因
CWT	精确远场角宽、有限阵列边缘泄漏、器件级阈值电流	基底维数过低，且有限尺寸与电热回写未显式进入
PWEM	开放辐射损耗、真实有限器件排序、阈值结论	典型形式是无损被动无限周期本征问题
RCWA	有限孔阵边缘主导问题、横向包络与边缘吸收主导问题	单胞周期模型本身不含真实有限尺寸终止
FDTD/FEM	电注入阈值电流、阈上稳态、热稳定工作窗口冷腔	若无电热回写，只能给冷腔或小信号光学层级结论

16.4 比较方法时，应比较哪些量

多方法互证并不意味着所有数值都要逐位相同。不同方法的基底、边界和离散方式不同，过分追求“完全一致”往往会掩盖更重要的信息。更有意义的是比较那些应当在方法变化下保持稳定的物理量或趋势。

第一类是候选模的排序。例如目标频段内究竟是哪几支模式最值得关注，这种排序不应因方法切换而完全翻转。

第二类是对称性与偏振标签。虽然数值细节可能不同，但目标模属于哪类对称性表示、主要偏向哪种偏振，应该能在方法间对得上。

第三类是参数微扰趋势。若增大孔半径在 CWT 中预测会增强某种辐射通道，那么 RCWA 与全波方法至少应在趋势上给出相容结论。

第四类是损耗来源分解。即使绝对 Q 不完全一致，也应能判断损耗主要来自向上辐射、向下辐射、边缘泄漏还是掺杂吸收重叠。

换句话说，互证的重点是“物理解释链条在不同分辨率下是否一致”，而非“每个小数点都一样”。

跨方法比较时，最危险的是层级混淆，而非少量数值偏差。若一种方法给的是无限周期被动结果，而另一种方法给的是有限器件冷腔结果，再拿二者的绝对阈值优劣直接比较，本身就没有严格可比性。先对齐问题层级，再比较结果，才有意义。

16.5 数值误差从哪来：互证前的最后一项功课

每种方法都有自己的主导误差源与收敛行为：PWEM 的平面波截断与有效折射率近似、CWT 的模型降阶、RCWA 的 Fourier 收敛与 Gibbs 现象、FDTD 的时域误差累积、FEM 的网格与弱形式实现。逐方法的误差源清单、三层收敛门槛口径（探索级/本书算例/发布级）、不确定度传递与报告规范，统一收入 第三十章 的数值不确定度部分（第30.12.1小节）。互证时只需带着一纪律律：任何进入比较的数字，都必须附有收敛证据和误差量级。

16.6 为什么互证必须按“由粗到细”的顺序组织

如果一开始就直接上三维全器件全波扫描，当然不是做不到，但通常代价极高，而且一旦结果不理想，很难知道问题究竟出在晶格、外延还是边界设置。相反，若先由粗到细组织 workflow，每一层都能承担一个清楚的筛选任务，后续搜索空间会大幅缩小。

最自然的顺序通常是：先用第 5 章的能带阅读框架配合第 13 章的 PWEM 找到候选模，再用 CWT 识别主导耦合通道与参数方向，再用 RCWA 精修单胞开放散射与光学阈值趋势，再用 FDTD/FEM 把有限尺寸和真实外延层栈带回来，最后把这些结果交给电注入和热耦合模型。这条顺序并非人为规定，而是与问题层级天然匹配。

这样做的另一个好处是，任何阶段一旦出问题，都能更容易定位。例如，如果 CWT 和 RCWA 已经都显示某方案目标模并无明显优势，那就没必要花大代价做全器件验证；反过来，如果 RCWA 很乐观、但全波有限器件结果很差，问题通常就出在边缘、有限尺寸或纵向层栈细节，而不是单胞本身。

16.7 一条适合 PCSEL 的推荐 workflow

把前面的思路压缩成一条实际可执行的 workflow，可以概括为五步。

第一步，建立候选模地图。使用第 5 章的能带概念和第 13 章的 PWEM 计算，在目标波长附近找出 Gamma 点模族、简并关系和对称性标签。

第二步，做机理级预设计。使用 CWT 或更高阶 CWT，把注意力集中到少数关键耦合通道，判断哪些几何扰动最可能改变模式分裂、偏振选择和辐射窗口。

第三步，做单胞开放精修。使用 RCWA，围绕前两步锁定的候选模区间，提取被动谱、上下辐射分配和光学阈值增益趋势。

第四步，做有限器件全波验证。使用 FDTD/FEM，把有限孔阵边缘、真实层栈、窗口和边界条件带进模型，检查真实近场、远场、有限尺寸 Q 和模式排序。

第五步，做器件级回写链。把第四步输出的模式重叠、损耗和导数信息交给增益模型、电学输运与热耦合模型，得到阈值电流、工作点漂移和高功率稳定性判断。

这条 workflow 最大的价值，不在于“把所有工具都用一遍”，而在于每一步都有清楚的淘汰与交接标准。

16.8 2024–2026 前沿补充：代理模型与 AI 设计器如何接入互证链

前沿工作显示，PCSEL 的快速设计工具正在从“低阶解析近似”扩展到“解析模型 + 数据驱动代理”的混合范式：一方面，快速工具评估研究给出不同近似器在精度/速度上的可操作对照 [36]；另一方面，基于 Transformer 的代理模型已能在复杂晶格参数空间中实现高吞吐预测 [53]。这并不改变本章的主结论，反而强化了“层级互证”的必要性：代理模型可以提前压缩搜索空间，但不能绕过物理保真层验收。

本书建议把代理模型接入点固定在“阶段 B 与阶段 C 之间”：

- 先用代理模型做粗筛（候选区间、非可行区排除）；

表 16.3: 方法链中的交接表: 每一层最值得交给下一层的中间量。

当前层	主要中间量	下一层如何使用
PWEM	候选频段、Gamma 点 模目录、简并与对称性 标签	交给 CWT 选择最小耦合子空间, 交给 RCWA 锁定扫描频段
CWT	主导耦合通道、模式分 裂方向、参数灵敏度符 号	交给 RCWA 选重点参数和候选模, 减少盲扫
RCWA	单胞开放谱、上下辐射 分配、光学阈值代理、 偏振趋势	交给 FDTD/FEM 构造有限阵列验证目标, 交给器件模型提供辐射 损耗输入
FDTD/FEM	有限器件近场/远场、 有限尺寸 Q 、场重叠 与边缘泄漏	交给增益、电学和热模型作为器件级输入
电热光模型	阈值电流、温升、工作 点漂移、稳定工作窗口	交给实验回写链做回验与设计更新

表 16.4: 工作流的继续/停机判据。若未通过当前层门槛, 不应升级到下一层最昂贵模型。

层级	继续前进前至少要确 认的事	若未通过应如何处理
PWEM	候选模频段、简并标签 和带图读取无歧义	回到单胞几何与归一化定义
CWT	参数变化方向与主导 耦合通道已识别	先缩小参数空间, 不要急于全波
RCWA	候选模开放谱与辐射 排序稳定, 功率守恒通 过	回查傅里叶截断、材料色散和边界设定
FDTD/FEM	阵列尺寸、PML、监视 面和网格收敛, 模式身 份稳定	回查有限尺寸与开放边界, 不要越级宣称器件优劣
电热光	注入、温升和工作点回 写后目标模仍占优	不能输出最终器件窗口, 应回到光学或注入分布层重设方案

- 再用 RCWA/FDTD/FEM 做物理验收（模式身份、边界条件、收敛与误差）；
- 最后用器件级电热模型验证工作窗口是否仍成立。

若缺少第二步和第三步，代理模型给出的“最优设计”只能算假设生成器，不应直接进入发布级结论。

16.9 互证不是礼仪，而是发现错误的主要手段

研究里常见的一种误区，是把多方法互证理解成“为了让论文更好看，多放几张图”。其实互证的主要价值是发现错误。原因很简单：不同方法对不同误差源敏感。若 RCWA 和 FDTD 对同一趋势意见一致，通常说明单胞开放物理和有限器件物理在该问题上相容；若它们结论冲突，就说明至少有一层物理被漏掉或被误设。

举例来说，若 CWT 预测某种椭圆孔扰动会增强偏振分裂，而 RCWA 也显示对应偏振通道的辐射损耗排序拉开，但 FEM 在真实有限外延层栈里却显示两支模式仍接近简并，那么问题很可能不在面内晶格，而在纵向层栈和边界。反之，若 RCWA 与 FEM 一致乐观，而器件级电热模型仍显示工作点下模式切换频繁，那么问题多半不在冷腔光学，而在注入与热反馈。

互证的作用不只是“确认”，还包括“定位”。

假设某方案在 RCWA 中显示目标模的光学阈值增益最低，于是看起来很理想；但 FEM 有限器件分析发现该模式在边缘附近场很强，导致边缘吸收一并增强；再进一步，热-电-光耦合分析又显示电流拥挤把增益集中在另一片区域。最终器件先起振的模式可能并非 RCWA 中最优的那支。这是它只负责单胞开放层级，后面两层信息把排序改写了，而非 RCWA 错。

这个例子在于多方法链存在的意义：每一层都在回答自己的问题，任何一层都不应擅自越权宣布最终器件结论。

16.10 什么时候该停在某一层，不再继续加方法

并非每个课题都需要把所有方法全部用满。若你的目标只是理解某种晶格扰动的基本耦合机制，CWT 加 RCWA 可能已经足够；若你的目标是论文级有限器件远场和偏振验证，通常还需要 FEM 或 FDTD；若你的目标是器件级阈值电流与高功率稳定工作区，那就必须继续进入后续的外延、电学和热学模型。

判断是否需要继续升级的一个实用标准是：当前层输出的信息，是否足以支持下一步设计决策。如果还不能，就必须升级；如果已经足够支持当前问题，不应为了“形式上更完整”而无条件加重模型。

易错点

“只要最后用了一次最重的全波仿真，前面的模型层级就都可以省略”是低效做法。这样会削弱结果的可解释性；一旦结果不理想，也很难判断应从哪个物理层级修改。对 PCSEL 这种跨层次器件，这类无分层路线尤其危险。

16.11 本章在后续器件章节前的作用

本章的任务，是在进入 第十七章至第二十章 之前，把光学方法层彻底整理清楚。因为后面的外延、增益、电注入和热效应章节都需要一个明确前提：哪些光学量已经可靠，哪些仍是近似代理量。如果这组输入边界不先说清楚，后面的器件模型就会拿错输入，进而把整个链条的结论层级拉乱。

因此，本章既是数值方法部分的总结，也是器件物理部分的入口。它提醒读者：从这里开始，任何“器件结论”都必须建立在多层方法分工明确的基础上。

回看本章

CWT、RCWA、FDTD 和 FEM 并非彼此竞争的替代品，而是 PCSEL 建模链中分工不同的四个层次。CWT 负责机理压缩，RCWA 负责开放周期单胞精修，FDTD/FEM 负责有限器件全波验证。可靠结论来自这些方法的有序接力与互证，而不是某一种方法单独承担全部任务。对 PCSEL 这样同时包含周期光学、开放辐射和器件多物理场的对象，多方法链是基本要求，而非奢侈配置。把这条分工原则说得再直白一些：第九章 中讨论的 BIC / quasi-BIC，并非脱离数值方法的抽象修辞，它需要 RCWA 或 FEM 去分辨 Γ 点附近辐射通道如何关闭或重新打开；而 第二十三章 中讨论的调制、模式竞争和线宽，也不是脱离光学方法的后处理，它们必须继承这里已经建立的模式排序、损耗分解和器件输入接口。

练习题

1. [基础] 解释为什么“哪种方法更准”这个问题在没有说明问题层级时往往没有意义。
2. [基础] 设计一条适用于大面积单模 PCSEL 的方法 workflow，并说明每一步的淘汰标准。
3. [进阶] 给出一个 RCWA 与 FDTD 结论可能冲突的场景，并说明冲突背后最可能缺失的物理是什么。
4. [综合] (思路提示：先定层级，再列关键式) 说明为什么互证时更应比较模式排序、对称性标签和参数趋势，而不是只比较单一绝对数值。

延伸阅读

- 下一章开始正式进入外延层的光学结构。后续章节将说明，前面光学方法的输出量如何被转换成外延设计中的纵向重叠、损耗重叠和上下辐射不对称。
- 若想补充方法基础，可回看 [28, 64] 中关于本征法、散射法与时域法的相关章节。
- 若想检验自己是否理解了“方法层输出如何喂给后续模块”，可带着本章的分工框架回读第九章、第二十三章：前者依赖你正确分辨开放辐射通道，后者依赖你正确继承模式损耗和阈值排序。

第四部分

器件物理与设计

第十七章 PCSEL 的半导体外延层与纵向光学结构

读完本章，你能

- 理解外延层为何与平面内晶格并列，成为 PCSEL 的第一性器件变量。
- 掌握分离限制异质结构（separate confinement heterostructure, SCH）、异质结和能带对齐怎样同时进入纵向光学约束与后续电注入设计。
- 能够把纵向波导模式、量子阱重叠、损耗层重叠、上下辐射不对称和材料台阶设计放到同一条器件链上理解。

读前准备 建议已掌握 第**六**章、第**七**章和第**十六**章 中关于面发射、阈值损耗、开放系统与数值方法分工的内容。

阅读主线 本章所处模型层级：**L2**

第**四**章至第**十六**章 把 PCSEL 当作周期开放电磁系统来理解，并已经得到一套关于候选模、辐射损耗、阈值增益和数值工作流的语言。但真实器件不会把这些结论“原封不动”交给实验，因为实际工作的模式并不生活在抽象二维平面里，而是嵌入在异质结外延层、量子阱、有损接触层和非对称边界条件共同构成的纵向结构中。本章的任务，就是把这条纵向主线搭起来。我们先回答为什么 PCSEL 必须采用分离限制异质结构，再说明导带、价带台阶如何进入设计；随后把典型层栈、纵向波导、本征模、有效折射率近似、重叠积分与损耗重叠放到一条链上，最后解释为什么这些外延细节会直接改写前面章节得到的冷腔排序与远场判断。

贯穿案例 B0：先交付纵向光学接口 从本章起，表**2.1**中的 B0 固定为 940 nm 附近的 GaAs-like 教学器件：上、下 SCH 各先取约 200 nm，有源区先放置 3–5 个 InGaAs 量子阱。这里的数字只用于建立透明的初始模型，并非 Noda–Yoshida 器件的外延复原。本章向下一章交付目标纵模的 n_{eff} 、量子阱重叠 Γ_{opt} 和损耗区重叠 Γ_{loss} ，而不止于一张层栈图；三者都应随层厚扫描并附带边界条件。

17.1 从二维光子晶体到真实器件：建立完整纵向物理链

如果只顺着前面的 Maxwell、Bloch、CWT 和全波仿真章节读下来，读者很容易形成一种危险印象：仿佛 PCSEL 的主要设计自由度都在平面内，外延层只是“把这个结构长出来”

的工艺载体。这在真实器件层面是错误的。原因很直接，前面得到的一切候选模、辐射耦合、Gamma 点排序和阈值增益需求，最终都要投影到一个实际三维模式上；而三维模式的纵向位置、与量子阱的重叠、与掺杂层和金属的重叠、对上/下辐射通道的耦合，全部由外延层决定。

换句话说，前面章节回答的是“如果给定一个纵向背景，平面内周期结构会如何组织反馈和辐射”；本章开始回答的是“这个纵向背景本身应该如何被搭建，才能让前面那些冷腔结论在真实器件里仍然有意义”。只要进入器件层面，纵向外延结构就和二维晶格一样，都是模式选择的一部分。

这条纵向链最容易被误读成“层栈画法”。更准确的读法是：外延层把一个冷腔光学模翻译成一组器件接口量，关键量是模场与有源区、损耗区各自的重叠，而非层名本身。图17.1 把这件事画成一张接口地图，后续章节中的量子阱增益、电注入和热回写都要沿着这张图继续展开。

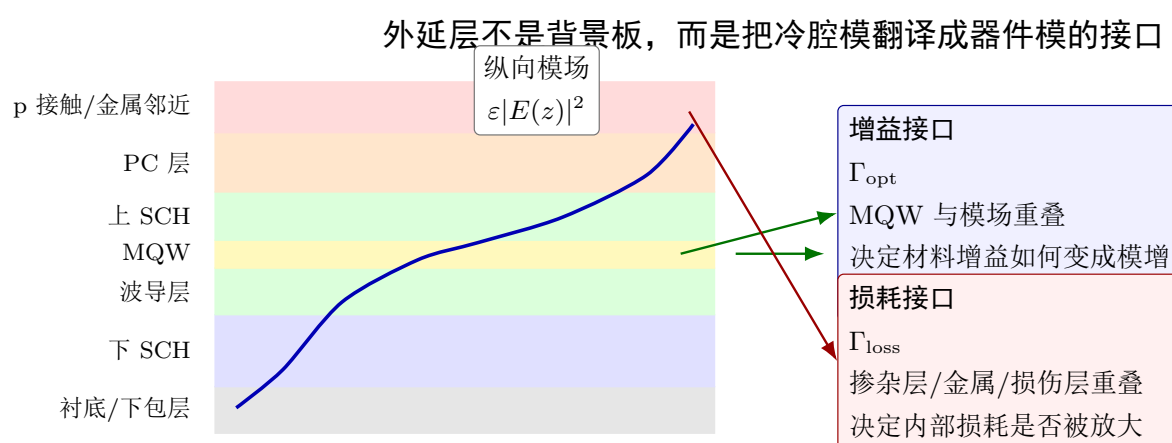


图 17.1: 外延层作为器件接口的图形化读法。平面内光子晶体给出冷腔候选模，但外延层通过纵向模场、量子阱重叠和损耗区重叠决定它能否成为真实器件模。

17.2 为什么 PCSEL 几乎总要落到 SCH：从教科书结构到器件可行性

分离限制异质结构（SCH）的核心思想，是把“强光学场区域”和“主要载流子受限区域”通过异质材料台阶组织在一起，但又不把两者机械地做成同一层。更具体地说，量子阱区负责提供受激辐射增益，周围较低折射率但带隙更大的波导/限制层负责把模式束缚在量子阱附近，同时提供载流子的势垒环境。这种“有源区很薄、光学模更宽、两者通过重叠而非完全重合来耦合”的结构，就是现代半导体激光器和 PCSEL 中最常见的纵向架构。

为什么 PCSEL 尤其需要 SCH？原因至少有三条。

第一，如果把增益材料本身直接做成唯一波导层，模式会过于贴近高吸收、高散射或高掺杂区域，阈值增益需求迅速恶化。第二，PCSEL 常常还需要在量子阱上方再放入光子晶体层、接触层乃至空气孔结构，这要求光学模既要与量子阱有足够重叠，又不能过度拉向表面损

耗区。第三，面发射器件中上下边界天然不对称，如果没有 SCH 提供的纵向缓冲和再分配空间，模式很容易被上表面工艺层或下方衬底通道拖偏。

因此，SCH 的器件作用不能归结为沿用传统边发射激光器模板。它把第七章的阈值条件落实到层栈中：模式既能在量子阱附近获得足够光学重叠，又不至于被表面与接触损耗吃掉。为了帮助读者直观理解 PCSEL 的典型纵向外延层堆叠结构，图17.2展示了一个完整的 GaAs 基 PCSEL 外延结构示意图。典型的 PCSEL 外延结构从下至上依次包括：GaAs 衬底、下 SCH 波导层（约 200 nm）、多量子阱有源区（3-5 个 InGaAs 阱）、上 SCH 波导层（约 200 nm）、光子晶体层、电流阻挡层和 p 型接触层。某些设计可能在底部包含 n 型掺杂层以降低串联电阻，但 PCSEL 的面内反馈机制使其通常不需要 DBR 反射镜——这与 VCSEL 的结构形成鲜明对比。图中清晰标注了各层的典型厚度范围、材料体系和掺杂类型，这些参数对于理解后续章节的电注入设计和热管理策略至关重要。

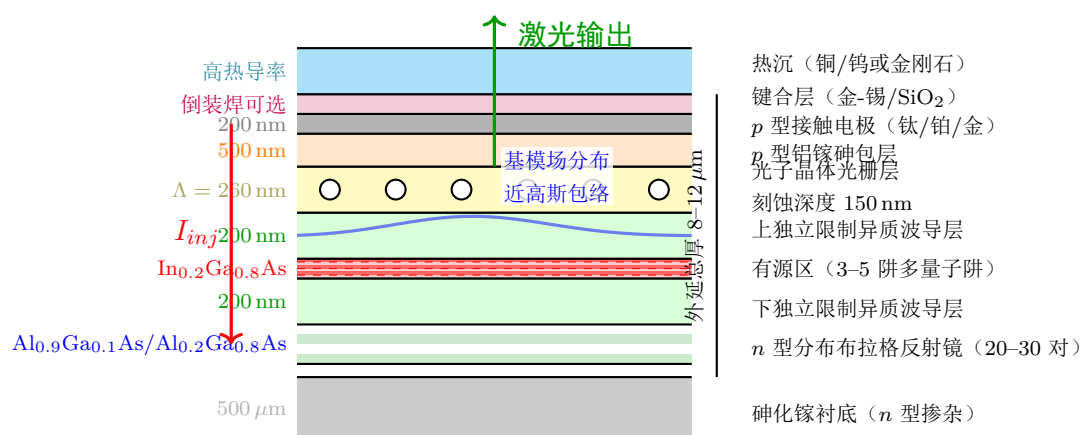


图 17.2: 典型光子晶体面发射激光器外延堆栈截面示意（未按比例）。图中给出衬底、下反射镜、上下限制层、有源量子阱、光子晶体层、包层与金属接触，并标出注入电流与垂直出光方向。

17.3 异质结和能带对齐为什么必须在本章就讲

如果只把本章理解成“光学层栈设计”，容易以为能带对齐属于后面电学章节。但这会把问题人为切断。异质结一方面通过折射率台阶形成光学波导，另一方面又通过带隙差形成载流子势垒；这两种作用来自同一组材料选择，因此必须在本章同时建立直觉。

设量子阱或波导层与限制层之间的带隙差为

$$\Delta E_g = E_{g,\text{barrier}} - E_{g,\text{well}}, \quad (17.1)$$

则工程上常把它分解为导带与价带台阶

$$\Delta E_c = Q_c \Delta E_g, \quad \Delta E_v = (1 - Q_c) \Delta E_g, \quad (17.2)$$

其中 Q_c 是导带偏移系数。式 (17.2) 属于器件材料与电学-光学接口模型。它的意义在于：材料层一旦确定，不仅折射率剖面确定了，电子与空穴在纵向上看到的势垒也同时确定了。

对常见 III-V 体系， Q_c 必须按平台给初值，而不是跨平台共用单个“标准值”。表17.1 给出本书采用的分平台起始区间，均应理解为**建模初值**：真实器件分析时仍需结合温度、应变和组分做二次标定。

表 17.1: Q_c 的平台化初值口径（仅作建模起点，不是跨平台通用常数；温度 $T = 300\text{ K}$ ）。			
材料平台	Q_c 起始区间	来源类型	说明
GaAs/AlGaAs（近红外）	0.60–0.70	[文献值]	教材与器件文献常见区间；室温附近常见初值约 0.67[10, 51]
InGaAsP/InP（1.3/1.55 μm ）	0.35–0.45	[文献值]	需随组分和应变重新标定，不可直接外推到 GaAs 平台 [10, 51]
InGaN/AlGaN（蓝光/深紫外）	0.60–0.80	[工程经验] 起点	极化场与应变耦合更强，建议先取宽区间再做平台内拟合 [7, 51]

这里应记住 Q_c 是材料平台相关的分配系数，无需背诵一个脱离材料体系的“标准值”。只要 Q_c 变化，电子与空穴各自的限制强弱、泄漏倾向以及对阻挡层的需求都会一起改变。

这对 PCSEL 的影响非常直接。若 ΔE_c 太小，电子在高注入或高温下容易越过量子阱区泄漏到上包层或接触层；若 ΔE_v 太小，则空穴更难被有效限制在有源区，因为空穴有效质量更大、迁移率更低、注入路径也常更差。于是，一个看似“光学上模重叠不错”的外延层，只要能带台阶设计不合理，到了电注入时就会表现出电子泄漏、空穴堆积或有源区载流子不足。这也是“冷腔结论不能直接变成器件结论”的第一层原因。

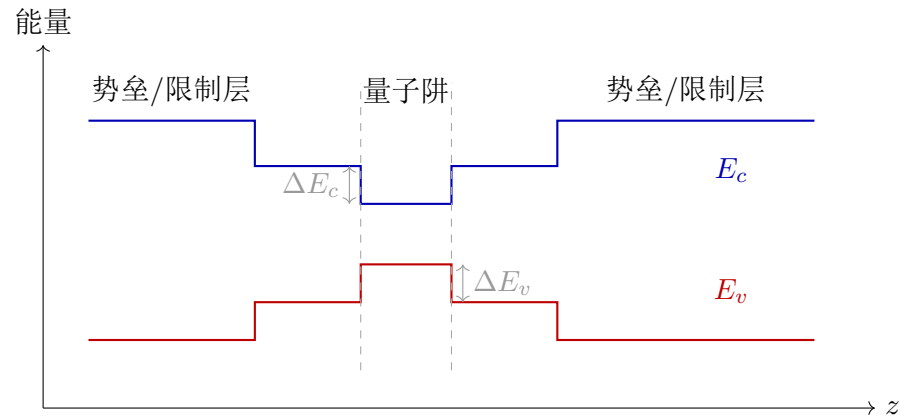


图 17.3: 量子阱与限制层的导带/价带台阶示意。对 PCSEL 来说，这些台阶既决定载流子限制，也通过材料选择间接决定纵向折射率结构。

对电注入 PCSEL 来说，还应把这张图读成一个位置关系示意：MQW 有源区通常埋在波导/SCH 区内，而光子晶体层多位于其上方的上波导或近表面区域。这样布置可避免直接刻蚀量子阱，同时让目标纵模的倏逝尾取样光子晶体散射，并保留对 MQW 的较高重叠与较低工艺损伤风险。

17.4 电子泄漏和空穴注入困难为什么会在 PCSEL 中显得更尖锐

在很多边发射激光器教材中，电子泄漏和空穴注入困难已经是标准话题；PCSEL 并没有逃离这些问题，反而因为大面积器件和上表面结构复杂性，常常更敏感。PCSEL 还要求较大横向区域内的有效模增益分布与目标模式匹配，局部获得增益远远不够。任何纵向泄漏机制都会破坏这种大面积匹配。

电子泄漏通常在两种情况下变严重：一是导带势垒不够高，二是有源区上方电场和温度抬高后，电子更容易跨越势垒进入上包层、接触层甚至表面附近高损耗区域。空穴注入困难则来自另一个方向：空穴迁移率较低、有效质量较大，且在很多材料体系中可用的价带势垒工程窗口更窄，于是空穴更容易在 p 侧波导区或限制层累积，而不是均匀进入所有量子阱。结果就是：从电流看来已经“注入完成”，从增益看来却没有使有源区获得均匀有效填充。

对于 PCSEL，这些机制之所以必须提前讲，是因为它们最终会把第十八章的材料增益和第十九章的注入分布改写成一个横向、纵向都不均匀的器件工作点。前面章节的冷腔模式排序，到了这里就开始遇到真正的器件反例：即便高 Q 、高重叠或远场理想，也不自动对应低阈值器件（红线 #2，见第2.10节）。

17.5 典型 PCSEL 外延层从下到上分别承担什么角色

现在再回到层栈本身。对于典型电注入 PCSEL，自下而上常见的层包括：衬底、下包层、下 SCH/波导层、量子阱区、上 SCH/波导层、光子晶体层、上包层、扩展/接触层与金属。不同路线还可能加入隧穿结、电流阻挡层、电子阻挡层或单独的 current spreading layer。它们的作用不能只按材料名记忆，而应按“对模式、载流子与损耗分别做了什么”来理解。

衬底与下包层主要提供机械支撑、散热通道和下方折射率边界；上下 SCH/波导层把模式缓慢引导到量子阱附近，同时给异质结台阶留出设计空间；量子阱区负责提供受激辐射增益；光子晶体层既参与平面内反馈，也会重新分配纵向模剖面；上包层、接触层与金属则在保证注入的同时引入自由载流子吸收、金属吸收和表面工艺约束。

这些层构成一个耦合系统。比如为了降低串联电阻而加厚上接触层，可能让模式尾部更靠近高损耗区；为了提高表面耦合而把光子晶体层放得更靠近量子阱，又可能让刻蚀损伤和表面复合风险上升。外延设计从一开始就是多目标平衡，而不是“先光学、后电学”。

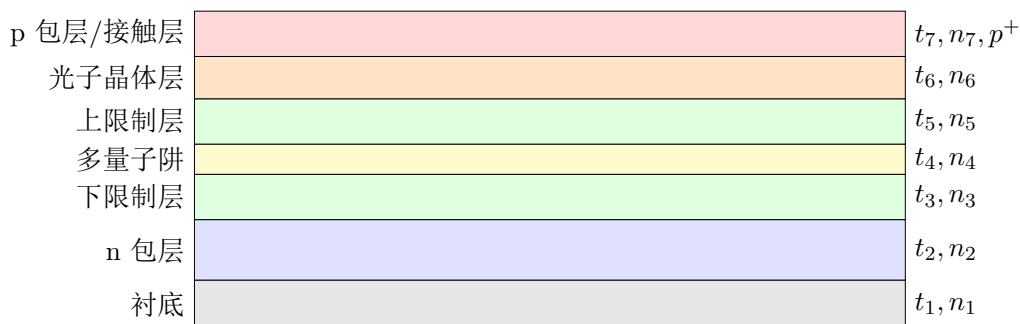


图 17.4: 典型 PCSEL 外延层纵向结构示意图。读图重点是每一层如何共同参与光学约束、载流子限制和损耗分布。

17.6 从 slab 波导方程开始理解纵向模式

一旦把平面内周期调制暂时冻结, PCSEL 的纵向问题首先就是一个层状波导问题。若只考虑沿 z 方向变化的折射率剖面 $n(z)$, 则在标量近似下可写

$$\frac{d^2\psi}{dz^2} + k_0^2 n^2(z)\psi - \beta_{\text{slab}}^2 \psi = 0, \quad (17.3)$$

其中 $\psi(z)$ 为纵向模剖面, β_{slab} 为面内传播常数, $k_0 = \omega/c$ 。这里统一用下标写成 β_{slab} , 避免与噪声章节的 β_{sp} 混淆。式 (17.3) 属于光学模型, 它回答的是: 在给定层栈下, 模式主峰位于何处、尾部穿透到哪些层、与量子阱和损耗层分别有多大重叠。

对 PCSEL 来说, 这一步告诉读者: 即便平面内晶格完全不变, 只要纵向层栈略有改变, 目标模式就可能不再是“同一支模式”。比如波导层变薄会让场更靠近表面, 接触层重掺杂则会让损耗尾部增大, 衬底折射率边界变化还会改写上下辐射比。也就是说, 冷腔中同一组平面参数, 不对应唯一的真实器件模。

从求解步骤看, 式 (17.3) 的逻辑非常朴素: 在“芯层折射率高于包层”的区域, 若 $\beta_{\text{slab}} < nk_0$, 解是振荡函数; 在包层中若 $\beta_{\text{slab}} > nk_0$, 解会变成倏逝衰减。某一支模式由各层解在界面处的连续拼接共同选出, 同时最外侧只允许衰减而非发散; 单独观察某一层的解无法完成选模。因此, 纵向模的身份从一开始就由整个层栈共同决定。

例子

对一个最简单的对称三层 slab, 若芯层折射率为 n_1 、包层折射率为 $n_2 < n_1$ 、芯层厚度为 d , 则偶对称 TE 模可写成

$$E_y(z) = \begin{cases} A \cos(\kappa z), & |z| \leq d/2, \\ B \exp[-\gamma(|z| - d/2)], & |z| > d/2, \end{cases}$$

其中

$$\kappa^2 = n_1^2 k_0^2 - \beta_{\text{slab}}^2, \quad \gamma^2 = \beta_{\text{slab}}^2 - n_2^2 k_0^2.$$

由边界连续条件得到色散方程

$$\kappa \tan\left(\frac{\kappa d}{2}\right) = \gamma. \quad (17.4)$$

这条式子虽然只是最简模型, 却非常适合训练直觉: 一旦芯层厚度、折射率对比或工作频率改变, β_{slab} 和纵向尾部衰减长度就会一起变化, 这也是 PCSEL 模场被拉向量子阱、表面或接触层的根源。

17.7 有效折射率近似在这里为什么既必要又危险

若进一步把纵向问题与平面问题做变量分离, 写成

$$\mathbf{E}(\rho, z) \approx u(\rho) \mathbf{e}_0(z), \quad (17.5)$$

并令 $\mathbf{e}_0(z)$ 先满足纵向本征方程，则平面内包络 $u(\rho)$ 看到的主导相位常数就是

$$\beta_{\text{slab}}^2 = n_{\text{eff}}^2 k_0^2. \quad (17.6)$$

有效折射率近似由此而来：它把“完整层栈如何支持某一纵向导模”这件事，压缩成一个供二维平面问题使用的标量背景。

实际设计中常先把纵向层栈投影成一个有效折射率

$$n_{\text{eff}} = \frac{\beta_{\text{slab}}}{k_0}, \quad (17.7)$$

再把三维问题降成二维平面模型。这一近似之所以必要，是因为前期扫描晶格常数、填充因子、孔径和对称性破缺时，若每次都带完整三维层栈，计算成本过高；它之所以危险，则是因为它把整条纵向物理链压缩成了一个标量。

只要问题开始依赖以下任一因素，有效折射率近似就会系统性变弱：不同候选模在纵向上的剖面明显不同；上下辐射不对称很强；掺杂层、金属或表面结构与模式尾部重叠显著；平面内周期本身能把模场重新拉向某一纵向层。这些恰好都是 PCSEL 常见的问题。因此，第十三章至第十六章中用二维与单胞模型得到的趋势，只能作为“在给定纵向背景下的候选筛选”，不能越权解释真实器件的最终阈值和远场。

对初学者来说，更实用的安全用法是三步：第一，用 n_{eff} 做前期参数扫描，只问候选频段和对称性趋势；第二，至少用一次包含真实层栈的单胞开放求解核对候选模是否仍保持相同排序；第三，任何阈值电流、最终远场和器件级偏振结论，都不能只靠有效折射率近似独立下结论。

17.8 光学约束因子、损耗重叠和“有源区已经在模场附近”不是一回事

一旦把纵向模剖面解出来，上一章和下一章之间最关键的接口量就是重叠积分。对量子阱区 V_{QW} ，最常见的光学约束因子定义为

$$\Gamma_{\text{opt}} = \frac{\int_{V_{\text{QW}}} \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}{\int_V \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}. \quad (17.8)$$

它属于光学模型到增益模型的接口量。但对 PCSEL 来说，只给出 Γ_{opt} 还远远不够，因为目标模式除了和量子阱区重叠，也会和重掺杂接触层、表面损伤层、金属邻近区重叠。于是还必须定义损耗重叠

$$\Gamma_{\text{loss}} = \frac{\int_{V_{\text{loss}}} \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}{\int_V \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}. \quad (17.9)$$

式 (17.8) 可以从“局域材料增益如何变成模增益”直接推出。若量子阱区存在近似均匀材料增益 g_{mat} ，则某一纵向模的功率增长率可写成

$$g_{\text{mod}} = \frac{\int_{V_{\text{QW}}} g_{\text{mat}} \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}{\int_V \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}. \quad (17.10)$$

当 g_{mat} 在量子阱区近似为常数时，式 (17.10) 立即退化为

$$g_{\text{mod}} = \Gamma_{\text{opt}} g_{\text{mat}}. \quad (17.11)$$

因此， Γ_{opt} 的真正身份是“把局域材料增益投影成该模式有效模增益的权重”；同理， Γ_{loss} 则是在回答“这个模式有多少能量密度被拖进高损耗区”。

本章必须把这两者并排写出来，原因就在于此。新手最常见的器件误解之一，是看到模场剖面“穿过了量子阱”就以为增益条件已经不错。真正决定阈值的是：在给定外延层里，目标模式是否能同时获得足够大的 Γ_{opt} 和足够小的 Γ_{loss} 。若只提高前者却把模式压向高损耗区，最终阈值仍可能变差。

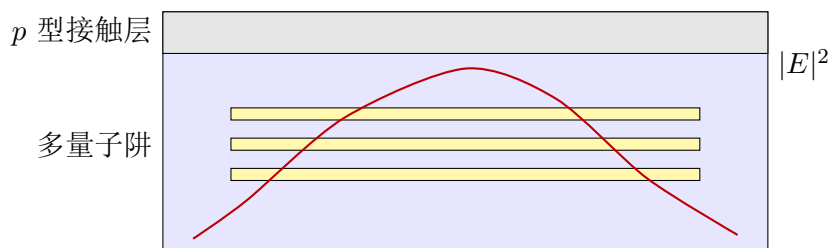


图 17.5: 量子阱重叠与损耗重叠示意。模式若被上表面和接触层拉得过近，即使量子阱重叠增加，也可能同时增加内部损耗。

17.9 掺杂和金属为什么在本章就已经是光学问题

电注入器件需要重掺杂层和金属接触，但这不意味着它们只属于 第十九章。只要模式场延伸到这些区域，它们立刻就以光学损耗的形式进入本章。工程上常用的自由载流子吸收近似可写为

$$\alpha_{\text{fc}} \approx \sigma_n n + \sigma_p p, \quad (17.12)$$

其中 n, p 是载流子浓度， σ_n, σ_p 是等效吸收系数。式 (17.12) 属于光学损耗代理模型，它的教学意义不在于给出最精确的数值，而在于提醒读者：掺杂既改变电阻，也改变阈值。

注记

自由载流子吸收的波长和温度依赖性。式 (17.12) 中的系数 σ_n, σ_p 并非常数，而是强烈依赖于波长和温度。对 GaAs 材料，经验公式给出 [62]:

$$\sigma_{n,p}(\lambda, T) \propto \lambda^q \cdot \left(\frac{T}{300 \text{ K}} \right)^s,$$

其中 $q \approx 2-3$ (波长指数), $s \approx 1-2$ (温度指数)。这意味着:

- 长波长器件 (如 $1.55 \mu\text{m}$ InP 基 PCSEL) 的 FCA 损耗显著高于近红外器件
- 高温工作时 FCA 会进一步增加, 形成正反馈的热 - 光耦合效应
- 在粗略估算中可取 $\sigma_n \approx 3 \times 10^{-18} \text{ cm}^2$, $\sigma_p \approx 7 \times 10^{-18} \text{ cm}^2$ (850 nm, 300 K) 作为参考值

因此, 在进行多波长或多温度仿真时, 必须更新 $\sigma_{n,p}$ 的值, 而不能简单套用室温近红外的数据。

金属的影响更不能简单理解成“再加一点吸收”。一旦模式被上拉到金属附近, 不仅损耗增加, 模式对称性、偏振选择与远场也会一起变化。因此, 在 PCSEL 里, 接触层与金属的设计不能后补, 而应从一开始就被看作光学边界条件的一部分。

17.10 外延层如何改写上下辐射不对称、偏振与远场

前面章节已经从平面内晶格的傅里叶分量和对称性角度解释了面发射、偏振与远场。但真实器件里, 模式向上看到的是空气、钝化层、表面结构和金属; 向下看到的是包层、衬底和更厚的半导体环境。即便平面内晶格完全对称, 这种纵向边界条件的不对称也会让上、下辐射通道不等价。

这会直接影响三类量。第一是出光方向与提取效率: 总辐射损耗相同, 不代表向上可用的面发射功率相同。第二是偏振选择: 不同矢量模与上下辐射通道的耦合权重不同, 因此外延层会参与 第八章 讨论的偏振选择。第三是远场角谱: 纵向相位积累和反射会改写法向附近的角分布。换句话说, 第八章 那些近场-远场映射, 在器件里必须与本章的纵向边界条件一起解读。

为了把这种“不对称”变成可比较的接口量, 常把向上提取比例定义为

$$\eta_{u/d} = \frac{P_{\text{up}}}{P_{\text{up}} + P_{\text{down}}}, \quad (17.13)$$

其中 P_{up} 与 P_{down} 分别是模式向上、向下辐射的功率。式 (17.13) 属于开放光学接口量。它不会单独决定阈值, 但会直接决定“同样一支冷腔候选模”在器件里究竟是更像一个向上可用的面发射模式, 还是更多地把能量泄向衬底和下包层。

17.11 层厚灵敏度与孔径工艺对比：从可算到可制造

到器件设计阶段，读者最常问的是“哪一层最值得优先优化”。对这个问题，最有效的回答是一张一阶灵敏度表。表17.2 给出本书 benchmark 层栈的教学型扫参示例（围绕基准值做 $\pm 10\%$ 改动，观察关键接口量变化）。它的价值在于排序优先级，而不是提供可直接外推到所有平台的“万能百分比”。

表 17.2: benchmark 外延层栈的一阶灵敏度示例（参数扰动 $\pm 10\%$ ）。数值用于设计优先级判断，不代表所有材料平台的统一结论。

扰动参数	关键量变化（代表趋势）	设计含义
上波导厚度 t_{core}	Γ_{opt} 约变化 $\pm(4\%-7\%)$ ； $\eta_{u/d}$ 约变化 $\pm(5\%-8\%)$	直接改写量子阱重叠与上下出光平衡，是优先级最高的纵向调参量之一。
光子晶体刻蚀深度 h_{etch}	α_{rad} 常变化 $\pm(10\%-20\%)$ ；选模排序对深刻蚀更敏感	主要影响辐射通道与模式分裂，过深刻蚀常带来辐射损耗和工艺粗糙度双重代价。
p 包层厚度 $t_{p\text{-clad}}$	与损耗层重叠变化可导致 α_{int} 约变化 $\pm(6\%-12\%)$	与接触层、掺杂层协同设计；只看电阻不看光学损耗容易把阈值推高。
MQW 中心位置偏移 $\Delta z_{\text{QW}} (\pm 10 \text{ nm})$	Γ_{opt} 常变化约 $\pm(3\%-6\%)$	在高 Q 、窄窗口设计中会直接改写净模增益排序，需与生长公差联动评估。

除了层厚，电流孔径工艺选择也会显著影响阈值与模式稳定窗口。以 VCSEL 体系的系统比较为例，指数引导氧化限制结构通常可获得更低阈值电流；而质子注入方案及增益引导氧化结构在横模选择性上可表现更优，同时具备更简化的工艺路径 [59]。这一权衡逻辑在 PCSEL 的电流孔径与注入工程中同样成立。

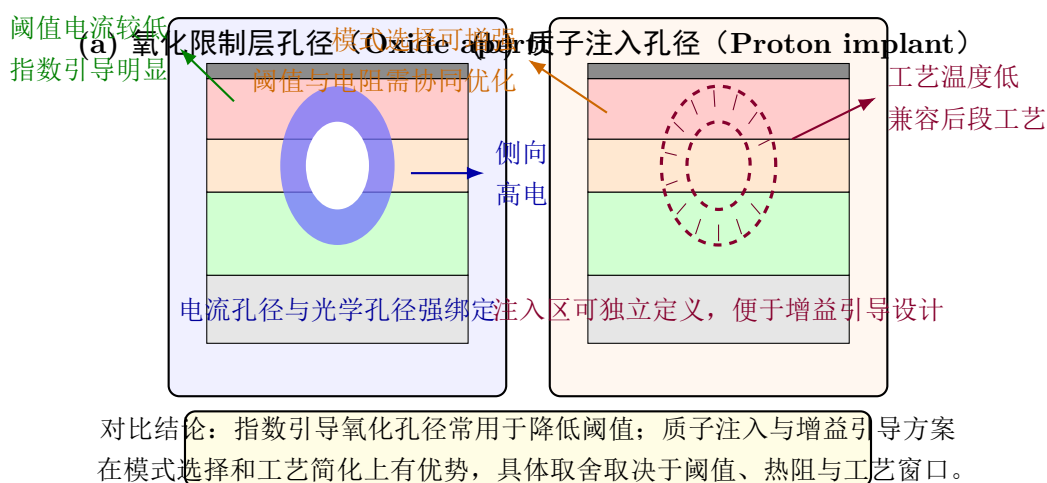


图 17.6: 氧化限制层与质子注入电流孔径方案对比。

17.12 本章输出给后续器件章节的量

到本章结束时，读者不应只得到一张层栈示意图，而应得到后面三章可直接调用的一组器件变量：目标模的纵向剖面、 Γ_{opt} 、 Γ_{loss} 、上下辐射不对称趋势，以及对层厚、折射率和异质结台阶变化的灵敏度。第十八章 会把这些量与量子阱增益连接起来，第十九章 会说明为什么这些外延层又决定了电流如何进入有源区，第二十章 则会展示温升如何反过来改写这些光学接口量。

这也说明了本章在器件主线中的位置：它是后续阈值电流、模式竞争和热稳定性讨论的层栈基础。

易错点

“外延层已经把量子阱放进去了，因此后续只需优化平面内晶格”是错误的。只要层厚、掺杂、金属位置或异质结台阶会改变模式的纵向位置和尾部重叠，外延层就一直在参与模式选择，而不是在几何确定后自动退出问题。

回看本章

PCSEL 的外延层首先是一个纵向光学结构，但它从一开始就带着器件物理信息进入设计。分离限制异质结构和异质结带边台阶决定了模式能否在量子阱附近取得合适重叠，同时把电子和空穴限制在可用区域；纵向层栈决定了光学约束、损耗重叠、上下辐射不对称与远场边界。有效折射率近似可以用于前期筛选，但不能取代三维器件验证。到本章结束时，读者应当已经清楚：前面章节得到的冷腔结论，只有在这里给出的外延层条件下才有机会变成真实器件结论。

练习题

1. [基础] 解释为什么 SCH 在 PCSEL 中是连接平面内光学设计与真实器件约束的基础，而非可省略的传统激光器结构。
2. [基础] 结合式 (17.2) 说明导带和价带台阶如何分别影响电子泄漏与空穴注入。
3. [进阶] 讨论为什么增大量子阱重叠并不自动降低阈值，回答中需同时引用 Γ_{opt} 和 Γ_{loss} 。
4. [综合] (思路提示：先定层级，再列关键式) 给出一个你认为 effective index 近似会明显失效的 PCSEL 场景，并说明失效来自哪些纵向物理。

延伸阅读

- 下一章将把本章建立的异质结、量子阱和重叠语言继续推进到有源区工程，解释单量子阱、多量子阱、应变和阻挡层为什么会改写材料增益到模增益的映射。阅读时应同时核对本章变量在后续模型中的定义和适用范围。

- 若想对照一般半导体激光器中的 SCH 与波导理论,可参考 [10] 中关于异质结波导、约束因子和损耗重叠的章节,但阅读时应始终记得:PCSEL 还要额外面对开放辐射与面内反馈耦合。阅读时应重点核对相关模型假设、参数范围与文献适用条件。
- 若想从器件实例反看外延层如何进入 PCSEL 主线,可参考 [24, 20, 80]。阅读时应重点核对相关模型假设、参数范围与文献适用条件。

第十八章 量子阱、有源区与增益模型

读完本章，你能

- 理解为什么 PCSEL 的有源区设计不能停留在“放入若干量子阱并赋予一个经验增益谱”这种层面。
- 掌握单量子阱 (SQW)、多量子阱 (MQW)、应变工程、TE/TM 增益差异以及阻挡层设计在 PCSEL 中各自回答什么器件问题。
- 建立“能带工程 → 材料增益 → 模增益 → 阈值载流子需求”的完整教材链，并明确它仍不等于阈值电流。

读前准备 建议已掌握 第七章、第十七章 中关于阈值增益、SCH、能带对齐与光学约束因子的内容。

阅读主线 本章所处模型层级：L2

上一章已经说明，量子阱并非孤立插入在层栈中的一层发光材料，而是与 SCH、异质结台阶和损耗重叠一起决定器件工作点的核心部件。本章进一步把这个结构转化为有源区工程语言。我们先从“为什么量子阱增益必须和能带工程一起讨论”开始，随后依次分析 SQW 与 MQW 的设计权衡、应变与 TE/TM 增益差异、限制层与电子/空穴阻挡层的作用，再把材料增益映射到模增益，最后明确指出：即便这一整条链已经写清，器件级阈值电流仍要等 第十九章、第二十章 才能闭合。

贯穿案例 B0：把层栈变成增益输入 本章沿用 B0 的 3-5 个 InGaAs 量子阱和上一章算得的 Γ_{opt} 。真正进入计算前，还必须用目标材料的阱宽、组分、应变和温度数据重新标定材料增益

$$g_{\text{mat}} = g_{\text{mat}}(\omega, N, T).$$

本书不把这些未指定参数伪装成已知的文献配方。本章的交付量是增益峰 $\lambda_g(N, T)$ 、透明载流子密度以及 $g_{\text{mod}} = \Gamma_{\text{opt}} g_{\text{mat}}$ ，它们随后与冷腔损耗共同给出 N_{th} ，但还不能直接给出 I_{th} 。

18.1 为什么本章不能写成一般半导体激光器的量子阱综述

量子阱和增益模型看上去像半导体激光器的一般话题，但对 PCSEL 来说，它们必须围绕一个更具体的问题来组织：前面章节已经给出一组候选模、辐射损耗和阈值增益需求，现在在有源区究竟能否在真实外延层里把这些候选模中的某一支推到阈值之上？

这句话的关键在于“某一支模式”。PCSEL 要问的是：哪一支面内反馈模式会在特定纵向重叠和横向分布下得到最大的有效模增益；知道材料“有增益”还不够。因此，有源区工程必须从一开始就和 第 7 章、第 8 章、第 17 章、第 19 章和第 20 章 同时对接。若把这一章写成一般性的 QW 增益综述，就会失去它在本书主线中的位置。

18.2 量子阱为什么是 PCSEL 有源区的自然选择

量子阱 (Quantum Well, QW) 之所以在 PCSEL 中几乎成为默认有源区，根本原因是它适合满足两个器件需求，而非笼统地“比体材料更先进”。第一，要在有限注入与有限损耗条件下提供足够高的材料增益；第二，要让增益谱、偏振选择和工作波长可以被材料与结构参数共同调节。

从能带观点看，量子阱把电子和空穴限制在窄势阱中，使纵向态量子化，改变了态密度与跃迁选择规则。这样做的结果，是在目标波段附近更容易形成较高的受激辐射增益，并允许通过阱宽、组分和应变工程调节增益峰位置和偏振特性。对 PCSEL 而言，这种可调谐性尤为重要，因为器件最终能否起振，不仅取决于模式损耗，还取决于增益峰是否与前面光学章节得到的候选模频率真正对齐。

18.3 SQW 与 MQW：不是多放几层阱这么简单

最直观的设计分支是单量子阱 (single quantum well, SQW) 和多量子阱 (multiple quantum wells, MQW)。如果只看材料增益的峰值，似乎多量子阱总更有利，因为总有源体积更大；但真实 PCSEL 设计必须同时看四个量：总可用增益、载流子在各阱之间的分配均匀性、应变累积以及与纵向模剖面的重叠。

为了帮助读者直观理解应变量子阱中的能带结构及其对偏振特性的影响，图 18.1 展示了一个典型的 InGaAs/GaAs 压应变量子阱示意。图中显示三个关键特征：第一，导带电子的抛物线型能带色散关系 $E_c(z)$ ；第二，价带中重空穴带 (HH, E_{HH}) 和轻空穴带 (LH, E_{LH}) 在应变作用下发生分裂 δE_{strain} ；第三，量子化能级 E_{c1} 、 E_{hh1} 、 E_{lh1} 及其包络波函数分布。对常见 (001) 压应变 InGaAs/GaAs 量子阱，HH 带上移后 C1-HH1 跃迁通常增强平面内偶极分量，因此更有利于 TE 增益而非 TM 增益；张应变情形才可能提升 TM 分量 [10, 2, 49]。图中的虚线框出了量子阱区域（宽度常见为数纳米到十纳米量级，具体取决于材料体系与应变预算），两侧为 AlGaAs 或 GaAsP 势垒层。

图 18.1 揭示的物理机制是 PCSEL 偏振控制的基石之一：通过精心设计的压应变，可以主动操纵价带结构，使特定偏振方向的增益占据主导地位。这对于需要稳定线偏振输出的 PCSEL 应用（如泵浦源、传感）至关重要。

SQW 的优点是结构简单、界面少、参数少，更容易把增益峰和模式重叠关系看清楚，因此非常适合作为理论推导和前期参数扫描的基准模型。它的缺点则是总有源体积有限，一旦目标模式损耗不够低，单阱可能难以提供足够的阈值增益裕量。

MQW 的优势在于可以在相同波导区内堆叠多个阱，从而提高总模增益容量，并在一定程度上分散单阱载流子密度需求。但 MQW 并非“阱越多越好”。随着阱数增加，首先会出现

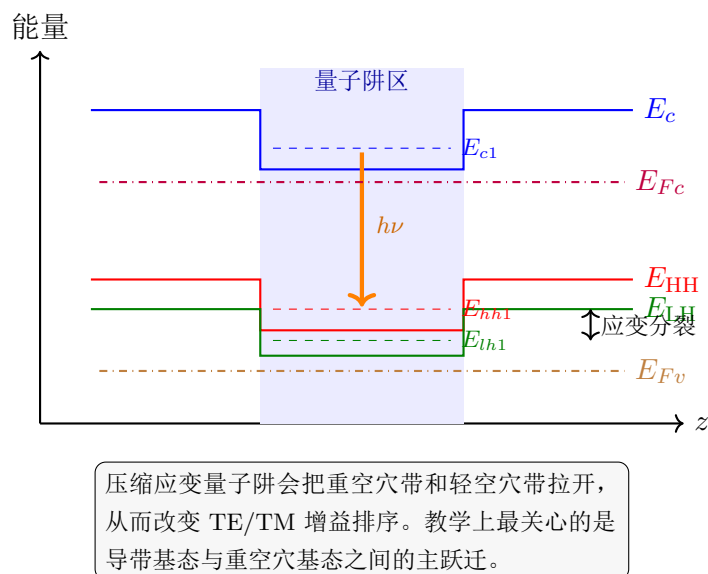


图 18.1: 压应变量子阱的能带结构及其偏振调控机制。(a) 未加应变时, GaAs 体材料的价带顶由重空穴带 (HH) 和轻空穴带 (LH) 近简并构成; (b) 施加双轴压应变后, HH 与 LH 分裂, HH 带上移; (c) 量子化后的能带结构: 导带形成电子子带 $C1$, 价带形成 HH1 与 LH1。对常见压应变 InGaAs/GaAs 量子阱, $C1$ -HH1 跃迁的偶极矩主要平行于量子阱平面, 通常对应 TE 主导; 若采用张应变设计, TM 分量可增强 [10, 2, 49]。

应变累积和生长质量压力; 其次, 注入后的电子与空穴未必能在各阱中均匀分布, 靠近注入一侧的阱可能先被过度填充, 远侧阱则利用不足; 再次, 若模式主峰没有覆盖所有阱, 额外增加的阱也不一定真正转化为模增益。

因此, SQW 与 MQW 的权衡不能只用一句“MQW 增益更高”结束, 而应当回答: 在给定的第十七章纵向模剖面下, 增加阱数到底是提高了 $\Gamma_{\text{opt}}g_{\text{mat}}$, 还是只是增加了生长、应变和注入不均匀性风险。

例子

若某个 PCSEL 模式的纵向场主要集中在波导中心, 而 MQW 堆栈中的最上两层阱已经接近上 SCH 或表面工艺层, 则这两层阱未必能像中间几层那样有效贡献模增益。此时“阱数增加”在器件里真正增加的可能更多是复杂度, 而不是阈值裕量。

18.4 有源区工程为什么必须回到能带图: 阱、势垒、限制层和阻挡层

量子阱从来不是单独设计的, 真正可工作的有源区总是由阱层、势垒层、SCH、限制层以及必要时的电子阻挡层 (electron blocking layer, EBL) 或空穴阻挡层 (hole blocking layer, HBL) 共同组成。它们分别回答不同问题。

阱层决定目标跃迁能量和增益峰位置; 势垒层把相邻阱分开, 控制载流子耦合和俘获; SCH 提供光学波导背景; 限制层和阻挡层则更偏向输运工程, 用来抑制电子向 p 侧泄漏、或在需要时抑制空穴向 n 侧回流。对于前面章节已经给出的 PCSEL 冷腔候选模而言, 这些层的作

用是把“有源区应该在哪一段纵向位置有效工作”这件事做成可设计对象。

工程上常把量子阱和势垒区域的材料增益写成

$$g_{\text{mat}} = g_{\text{mat}}(\omega, N, T, \epsilon_{\text{strain}}, \mathcal{B}), \quad (18.1)$$

其中 N 是局域载流子密度，决定反转是否足够； T 为温度，决定增益峰和透明载流子密度会不会漂移； ϵ_{strain} 是应变状态，决定价带分裂和 TE/TM 跃迁强度； $\mathcal{B}$ 代表阱宽、势垒厚度和组分等能带工程参数，决定量子化能级和势垒高度。式 (18.1) 属于**有源材料模型**。它强调了一件常被忽略的事实：材料增益是有源区工程的整体结果。

若把这一步再往量子力学层面退回半步，就能看见前一章的能带台阶为什么会直接进入增益谱。对有限深量子阱，阱内包络函数是振荡解，势垒中则转为指数衰减解；两者在界面上做波函数与流密度连续匹配时，会得到由 k_{well} 与 κ_{barrier} 共同决定的量子化条件。于是， ΔE_c 和 ΔE_v 并不只是“能带图上的台阶高度”，而是实际决定电子、空穴子带能级与束缚程度的势阱深度。对本章主线而言，最关键的因果链是：导带/价带台阶决定量子化能级，量子化能级决定主跃迁能量，主跃迁能量最终把增益峰 λ_g 锚定在某个波长附近。

若只想抓住进入阈值链的最小结构，可以先把目标模波长 λ_m 相对增益峰波长 λ_g 的失谐写成

$$\Delta\lambda = \lambda_m - \lambda_g(N, T), \quad (18.2)$$

并在峰值附近用一个最简二次近似表示

$$g_{\text{mat}}(\lambda_m, N, T) \approx g_{\text{pk}}(N, T) - a_\lambda(N, T) \Delta\lambda^2. \quad (18.3)$$

式 (18.3) 是教材化的最小代理模型（其中 a_λ 的量纲为 $\text{cm}^{-1}\text{nm}^{-2}$ ，确保量纲与增益一致），而非通用材料定律。其适用窗口必须绑定材料平台与工作波段：以 $1.55\text{ }\mu\text{m}$ InP 系 PCSEL 报道为例，增益谱半高宽常见于 $20\text{--}30\text{ nm}$ 量级，抛物线近似在峰值附近 $|\Delta\lambda| \lesssim 10\text{ nm}$ 常可用于一阶估算 [69, 46]；对 GaN/深紫外等平台，谱形与展宽机制差异更大，应重新标定窗口而非平移复用该区间 [10, 7]。它的作用是把“模式波长是否压在增益峰上”这件事，直接写成可计算的阈值链一环。

18.5 2024–2026 前沿补充：跨材料平台增益窗口不可直接平移

最新器件结果进一步强调了“平台绑定”的必要性：可见光 GaN-PCSEL、 $1.55\text{ }\mu\text{m}$ InP-PCSEL 与深紫外 PCSEL 的增益谱宽、温漂和可用工作窗口并不共享同一组经验常数 [65, 5, 7]。因此，本章所有失谐、增益窗口与阈值裕量参数，都应写成“平台 + 温区 + 注入条件”三元条件化形式。

对工程读者，一个实用准则是：

- 在同平台内，可把式 (18.3) 作为前端筛选代理；
- 跨平台比较时，不复用“同一数值窗口”，只复用“同一判断流程”；
- 进入器件级结论前，必须经由电热回写重新校准增益峰位置与有效窗口。

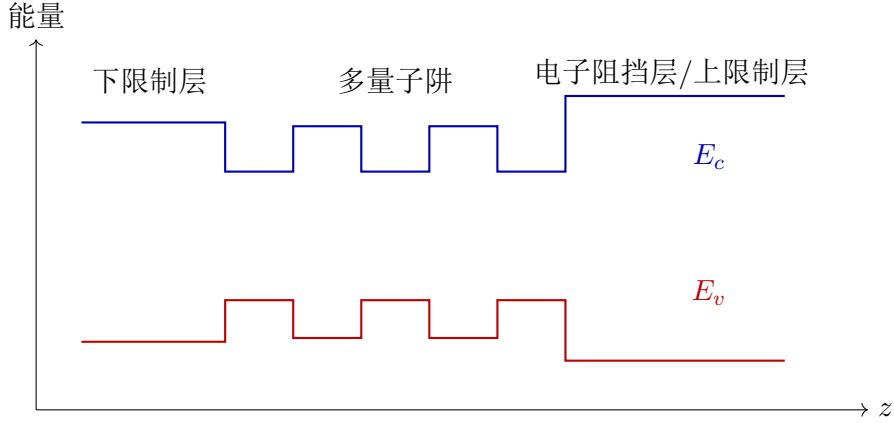


图 18.2: 多量子阱与阻挡层的能带示意。这里的每一道台阶都不仅影响载流子受限, 也会通过第十七章的重叠因子与损耗分布间接改写目标模式的有效增益。

18.6 应变工程和 TE/TM 增益差异：为什么它在 PCSEL 中尤其不能忽略

在普通边发射激光器里, 讨论 TE/TM 增益差异已经很常见; 在 PCSEL 中, 这个问题更关键, 因为前面第八章已经表明偏振选择和远场工程是器件设计的一部分。若有源区材料天然偏向某一偏振, 而光学腔又偏好另一支矢量模, 那么冷腔候选模与真正激射模之间的差距就会被进一步拉大。

应变工程的物理核心, 是通过晶格失配引入压应变或张应变, 改变价带顶的重空穴、轻空穴分裂和跃迁矩阵元, 从而改写 TE 与 TM 增益分量。压应变量子阱通常更有利于 TE 极化跃迁, 这对大多数以近 TE-like slab 模为目标的 PCSEL 是有利的; 而张应变则可能增强 TM 分量。工程上, 若你希望有源区与 PCSEL 的目标偏振对齐, 应变设计就必须与第八章的模式对称性和偏振选择一起考虑, 而不能在材料章节里单独处理。

在符号上, 可把材料增益拆成偏振分量

$$g_{\text{mat}} = g_{\text{mat,TE}} \hat{e}_{\text{TE}} + g_{\text{mat,TM}} \hat{e}_{\text{TM}}, \quad (18.4)$$

或者更常见地分别讨论两个标量谱。真正进入模阈值条件时, 对某一候选模 m 应写

$$g_{m,\text{mod}} \approx \Gamma_m g_{\text{mat},p(m)} - \alpha_{m,\text{int}}, \quad (18.5)$$

其中 $p(m)$ 表示该模主要采样的偏振分量。式 (18.5) 属于光学模与有源材料的接口公式。它清楚地说明: 同一组量子阱参数, 对不同偏振候选模给出的有效增益未必相同。PCSEL 的偏振选择, 并非纯冷腔对称性问题, 也是二者共同作用的结果, 而非纯材料问题。

若想把这种联合选择写得更定量, 可再引入模式对 TE/TM 分量的采样权重 $\eta_{m,\text{TE}}$ 与 $\eta_{m,\text{TM}}$, 满足

$$\eta_{m,\text{TE}} + \eta_{m,\text{TM}} = 1.$$

则更一般的接口式可写成

$$g_{m,\text{mod}} \approx \Gamma_m [\eta_{m,\text{TE}} g_{\text{mat,TE}} + \eta_{m,\text{TM}} g_{\text{mat,TM}}] - \alpha_{m,\text{int}}. \quad (18.6)$$

式 (18.6) 明确表明：即使冷腔更偏好某一支偏振家族，若有源区的 TE/TM 增益分裂朝着相反方向偏置，真正的激射模排序仍可能被重新洗牌。

18.7 电子阻挡层、空穴阻挡层和限制层到底在帮谁

从器件直觉看，EBL/HBL 似乎只是 第十九章 输运章节的内容，因为它们的直接作用是阻挡泄漏载流子。但就本章主线而言，它们其实在回答一个更精确的问题：你希望 第十七章 给出的量子阱重叠和纵向模式位置，对应的是“已被有效填充的有源区”，还是一个光场分布较理想、但载流子正在越层泄漏的伪有源区？

电子阻挡层最典型的任务，是在量子阱上方提高导带势垒，抑制电子在高注入和高温时越过有源区进入 p 侧包层。空穴阻挡层或优化的限制层则在需要时帮助改善空穴的空间分配，避免空穴过度停留在波导或某一侧阱区。对 PCSEL 来说，这些层的价值在于帮助模式真正“看见”一个更接近设计初衷的增益分布。

阻挡层的器件意义，在于避免本章的材料增益在映射到 第十九章 的注入分布时被严重稀释。

18.8 从材料增益到模增益：上一章的重叠因子在这里真正接上线

材料增益只有通过模式重叠，才会变成器件真的能用的模增益。最常用的近似写法是

$$g_{\text{mod}}(\omega, N, T) = \Gamma_{\text{opt}} g_{\text{mat}}(\omega, N, T), \quad (18.7)$$

其中 Γ_{opt} 来自 第十七章 的纵向重叠。注意内部损耗 α_{int} （包括自由载流子吸收、金属邻近损耗、表面损伤损耗等）不应从模增益中减去，而是出现在阈值条件中：

$$g_{\text{mod}}(N_{\text{th}}, T) = \alpha_{\text{rad}} + \alpha_{\text{int}}. \quad (18.8)$$

式 (18.7) 属于光学与有源材料的接口公式。它最重要的教学意义有两点：第一，材料增益高不自动等于模增益高；第二，即使材料增益已经对齐了目标波长，也仍可能被内部损耗吃掉。

把式 (18.1)、式 (18.3) 和式 (18.7) 连起来，本章真正要交给下一章的算式链就是

$$g_{\text{mat}}(\lambda_m, N, T) \rightarrow \Gamma_{\text{opt}} g_{\text{mat}} \rightarrow g_{\text{mod}} \rightarrow g_{\text{th}}. \quad (18.9)$$

只有沿着这条链逐层下去，读者才不会把“材料增益峰更高”误当成“目标模式阈值一定更低”。

若有源区填充在空间上并不均匀，则应写成加权积分形式

$$g_{\text{mod}} = \frac{\int_{V_{\text{QW}}} g_{\text{mat}}(\mathbf{r}) \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}{\int_V \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV} - \alpha_{\text{int}}. \quad (18.10)$$

这条式子已经提前把 第十九章 的注入非均匀性引进来了：同一个 MQW 结构，若电子泄漏或空穴分布不均导致各阱填充不同，最终参与模式竞争的并非某条理想均匀材料增益曲线，而是一个被器件输运重新加权后的局域增益场。

18.9 透明载流子密度、微分增益和工作点灵敏度

在器件教材里，透明载流子密度 N_{tr} 与阈值载流子密度 N_{th} 必须严格区分。 N_{tr} 对应材料净增益刚好为零的状态；而 PCSEL 真正达到阈值时，需要的是模增益补偿第七章已经定义的总损耗，因此 N_{th} 通常显著高于 N_{tr} 。对 GaAs/InGaAs 量子阱体系，若把它粗略换算成三维等效载流子密度， N_{tr} 常落在 10^{18} cm^{-3} 量级。混淆透明与阈值载流子密度，会系统性低估器件级注入需求。

另一个必须写出来的量是微分增益

$$g_N = \frac{\partial g_{\text{mat}}}{\partial N}. \quad (18.11)$$

式 (18.11) 属于**有源材料模型**。在 PCSEL 中，它不仅关系到调制响应，还关系到模式竞争的脆弱性。若某个模式虽然阈值附近占优，但对应的微分增益很低，则注入稍有不均匀或温度略有漂移时，另一模式就可能通过局域增益优势反超。于是，微分增益在 PCSEL 中不是只服务于高速分析，也服务于模式稳定性判断。

把上一章的阈值条件接回来，最小阈值载流子估计可写成

$$\Gamma_{\text{opt}} g_{\text{mat}}(\omega_m, N_{\text{th}}, T_{\text{th}}) - \alpha_{\text{int}}(N_{\text{th}}, T_{\text{th}}) = g_{\text{th},m}, \quad (18.12)$$

而在阈值附近做一阶展开时，

$$\delta g_{\text{mat}} \approx g_N \delta N + \frac{\partial g_{\text{mat}}}{\partial T} \delta T + \frac{\partial g_{\text{mat}}}{\partial \lambda} \delta \lambda. \quad (18.13)$$

式 (18.12) 和式 (18.13) 把“材料增益到模增益再到阈值载流子”的桥接真正写实了：阈值是由载流子、温度和模-增益峰失谐一起决定，而非由单个 N 决定。

18.10 简化增益模型、真实模型与它们各自的边界

在设计前期或参数不完整时，人们常用线性化模型

$$g_{\text{mat}} \approx g_N(N - N_{\text{tr}}) \quad (18.14)$$

或对数模型

$$g_{\text{mat}} = g_0 \ln\left(\frac{N}{N_{\text{tr}}}\right). \quad (18.15)$$

这两类模型都非常有用，但它们的正确身份必须被说清：它们是**真实器件问题的简化代理模型**。如果你的目标只是比较几个晶格参数和层栈方案在阈值附近的大致排序，这类模型足够高效；如果你的目标已经变成宽温区、强应变、多偏振耦合或高注入工作点的准确预测，就必须使用更真实的材料增益模型，把量子阱能级、谱线展宽、温度漂移和偏振分量显式纳入。

在 PCSEL 里，之所以要比一般教材更强调这点，是因为前面章节的冷腔候选模本身就很多。一旦增益模型过于粗糙，你很容易得到看似“模式 A 只比模式 B 高一点阈值裕量”的结论，但这点差别完全可能只是模型粗化误差，而非真实器件物理。因此，使用任何简化增益模型时都应写明参数来源、适用注入区间、温度范围和应变/偏振适用性。

18.11 为什么到这里仍然不能推出阈值电流

当本章写到这里，读者已经具备一条完整而严格的符号链：

$$\begin{aligned} [\text{阱宽、势垒、应变、阻挡层}] &\rightarrow g_{\text{mat}}(\omega, N, T) \\ &\rightarrow g_{\text{mod}}(\omega, N, T) \\ &\rightarrow g_{\text{th}} \rightarrow N_{\text{th}}. \end{aligned} \quad (18.16)$$

这条链已经足够说明“为什么同一个冷腔模式，在不同有源区工程下会有不同阈值载流子需求”。但它仍然没有告诉你：外加电流是否能在真实器件中把这些阱均匀填充、是否会发生电子泄漏、空穴注入不足、串联电阻抬高和发热回写。也就是说，式 (18.16) 的终点仍是一个光学与材料层级的阈值条件，还不是器件级阈值电流。

这也是本章最需要与 第 19 章 衔接的地方。只要读者在这里形成“有了 MQW 和一个增益模型，阈值电流就差不多出来了”的印象，后面的器件主线就会立刻被破坏。若要直接查看“从外加电流到阈值成立”的器件级方程链，请跳转到 第 19 章和式 (19.7)、式 (19.16) 和式 (19.17)；若要查看显式载流子-光子速率方程形式，请见 第二十三章和式 (23.1) 和式 (23.2)。

易错点

“量子阱设计好了、模增益公式也写出来了，所以阈值电流只剩下一个换算因子”是错误的。只要载流子在各阱之间分配不均、阻挡层设计不完整或注入受热反馈影响，本章给出的只是阈值载流子需求，而不是外加电流的器件答案。

回看本章

PCSEL 的有源区工程，必须围绕真实模式而不是抽象材料谱线来组织。SQW 与 MQW 的选择、应变工程、TE/TM 增益差异、势垒与阻挡层设计，共同决定材料增益能否在真实层栈中转化为目标模式的有效模增益。第 17 章 给出的 SCH 与能带台阶在这里继续发挥作用：它们既决定载流子是否被真正限制在有源区，也决定有源区与模式的纵向重叠。到本章结束时，读者应当已经清楚：从材料增益到模增益的链条已经建立，但它仍不能直接越级给出阈值电流，因为电流路径、接触和热反馈还没有进入求解。

从冷腔到增益补偿：连接第 3–8 章的模式分析与本章的有源区设计

回顾本书的理论架构，我们可以清晰地看到两条主线的汇合：

主线一（第 3–8 章）：冷腔模式分析

- 第 3 章：从 Maxwell 方程导出周期介质中的本征问题
- 第 4–5 章：Bloch 定理与能带结构，识别 Γ 点候选模式
- 第 6–8 章：面内反馈机制、偏振选择、远场对称性
- 核心产出：被动结构的复频率 $\tilde{\omega}_n = \omega_n - i\gamma_n$ ，其中 γ_n 包含辐射损耗 $+Q_{\text{out}}$

主线二（第 17–19 章）：有源区增益与注入

- 第 17 章：量子阱材料增益 $g_{\text{mat}}(\omega, N, T) \rightarrow$ 模增益 g_{mod}
- 第 18 章：电注入、输运、泄漏与非均匀填充
- 第 19 章：热效应对增益谱和折射率的回写
- **核心产出**：给定注入条件下的净增益能力 $g_{\text{net}}(N, T, I)$

两条主线的交汇点：阈值条件

$$\underbrace{g_{\text{mod}}(N_{\text{th}}, T)}_{\text{第 17 章}} = \underbrace{\alpha_{\text{i}} + \alpha_{\text{rad}}}_{\text{第 3-8 章}} = \frac{\omega_0}{2Q v_{g,m}} = \frac{n_{g,m} \omega_0}{2Q c}$$

这个等式的深刻含义是：**阈值载流子浓度 N_{th} 是冷腔品质因数 Q 的函数**，而非一个孤立的材料参数。高 Q 模式（如 BIC 或 quasi-BIC）之所以重要，正是因为它们降低了 α_{rad} ，从而降低了对 g_{mod} 的需求，进而降低了 N_{th} 。

这也解释了为什么本书反复强调“先做冷腔分析，再做有源区优化”的顺序：如果连被动结构的 Q_{out} 都不知道，就无法判断某个增益峰值偏移是否会致命；如果不清楚模式的近场分布，就无法准确计算 Γ_{QW} 。因此，本章的所有公式都必须建立在第 3-8 章的模式信息之上，而不能脱离具体模式空谈“材料增益大小”。

练习题

1. **[基础]** 比较 SQW 与 MQW 在 PCSEL 中的优缺点，回答中必须同时讨论模增益容量、注入均匀性和应变累积。
2. **[基础]** 解释为什么压应变量子阱往往更有利于 TE-like PCSEL 模式，并说明这与第八章的偏振选择机制如何耦合。
3. **[进阶]** 结合式 (18.7) 和式 (18.10)，说明为什么阻挡层设计不良会通过“局域增益稀释”影响模式竞争。
4. **[综合]**（思路提示：先定层级，再列关键式）说明为何电子阻挡层的作用不能被简单表述为“减少漏电”，而应放到模增益与阈值链条中理解。

延伸阅读

- 下一章将从本章的 N_{th} 与局域增益分布出发，解释接触、电流扩展层、孔径和漂移-扩散输运为什么会把这些理想条件重新改写成真实器件电流分布。阅读时应同时核对本章变量在后续模型中的定义和适用范围。
- 若想补充量子阱增益与应变工程的经典背景，可对照 [10] 中相关章节阅读，但阅读时要始终把问题放回 PCSEL 的模式选择与面发射器件约束中。阅读时应重点核对量子阱、有源区与模增益映射的标准假设和文献参数范围。

- 若想对照不同材料平台中有源区工程如何进入真实 PCSEL，可参考 [43, 74, 80]。阅读时应重点核对量子阱、有源区与模增益映射的标准假设和文献参数范围。

第十九章 电注入、载流子输运与阈值电流

读完本章，你能

- 理解为什么阈值电流是一个由接触、输运、复合、孔径和温升共同决定的器件级量，而不是阈值增益的换单位版本。
- 掌握电注入 PCSEL 中最基础的器件方程组：Poisson 方程、漂移-扩散方程、连续性方程，以及接触和界面条件如何进入求解。
- 能够解释 current spreading layer、电流孔径、注入非均匀性与模式竞争之间的直接联系。

读前准备 建议已掌握 第七章、第十七章和第十八章 中关于阈值增益、纵向重叠与有源区工程的内容。

阅读主线 本章所处模型层级：L3

前两章已经把“哪一种有源区和纵向层栈更有希望满足目标模阈值增益”讲清楚，但真实器件还缺最后一条关键链：外部电流到底如何进入器件、沿哪些层扩展、在哪些区域变成量子阱载流子、又在哪些区域被串联电阻、漏电、复合和不均匀注入吃掉。本章就沿这条链展开。我们先重新定义阈值电流的器件含义，再讨论欧姆接触、肖特基接触、接触电阻与接触几何；随后引入 current spreading layer 和电流孔径，给出横向电流扩展的最小模型，再写出 Poisson+ 漂移-扩散 + 连续性方程组，最后把局域载流子分布与 第十八章 的模增益公式重新接上，明确解释为什么纯冷腔结论和均匀增益模型在这里开始系统性失效。

贯穿案例 B0/R1：从载流子阈值到可扩展注入 B0 在这里接收 N_{th} ，并把问题改写为：给定接触与 spreading layer 后，能否在目标孔径内得到与目标模匹配的 $J(x, y)$ 和 $N(x, y, z)$ 。当口径从 B0 的 $200\mu m$ 教学尺度走向 R1 的 $500\mu m-3mm$ 公开器件尺度时，均匀注入不再是可以默认的边界条件。本章因此只在明确给出接触电阻、复合参数和注入效率后报告 I_{th} ；Noda-Yoshida 文献未在本书引用范围内给出的阈值电流，不由输出功率反推或补造。

19.1 阈值电流回答的器件问题

第七章 给出的阈值增益 g_{th} 回答的是一个光学问题：目标模式要补偿多少总损耗。第十八章 进一步把这个条件转成了材料增益与阈值载流子需求。但真实器件的阈值电流 I_{th} 回答的是更长的一句话：在给定接触、层栈、阻挡层、电阻、复合和温度条件下，需要施加多大的

外部电流，才能在空间上形成足够接近设计目标的载流子分布，使目标模式的局域模增益达到阈值。

这一定义里最重要的词是“空间上”。PCSEL 追求大面积面发射和单模工作，而大面积器件几乎不可能自动均匀注入。因此，阈值电流既取决于总载流子数是否足够，也取决于载流子在横向和纵向上的分布能否匹配目标模式。后一条件一旦失效，前面章节中的冷腔排序就会被重新改写。

直觉框

阈值增益给出单位长度损耗需要由多少模增益补偿；阈值电流则要求在接触电阻、分流、局部注入受限和发热存在时，外部电流仍能在目标有源区形成足够的局域载流子分布。对大面积 PCSEL 而言，后一问题才是器件级阈值分析的核心。

19.2 欧姆接触和肖特基接触如何进入边界条件

接触是器件求解里最容易被“口头默认”的部分，但在电注入 PCSEL 中，它必须以边界条件的形式显式进入模型。欧姆接触（ohmic contact）的理想极限，是金属和半导体之间几乎不形成整流势垒，外加电压可以高效转化为载流子注入。工程上，欧姆接触常通过重掺杂接触层和合适金属体系实现，其最小模型通常写成接触电势已知，或写成表面法向电流与接触比电阻（specific contact resistivity）相关：

$$J_{\perp} = \frac{V_{\text{metal}} - V_s}{\rho_c}, \quad (19.1)$$

其中 V_s 是半导体表面电势， ρ_c 为接触比电阻。这里的单位应是 $\Omega \cdot \text{m}^2$ ，而不是常见误写的 Ω/m^2 。教材级量纲感可先记成：[文献值] 在 GaAs 器件中， p 侧欧姆接触的 ρ_c 常见于 10^{-6} – $10^{-5} \Omega \cdot \text{cm}^2$ ， n 侧常在 $10^{-6} \Omega \cdot \text{cm}^2$ 量级 [10]。式 (19.1) 属于电学边界模型。它提醒我们：即便“欧姆”接触已经成立，接触仍然会通过有限的 ρ_c 引入额外压降和非均匀注入。

肖特基接触（Schottky contact）则对应明显的势垒和整流行为。其更贴近物理的边界模型通常写成热发射形式

$$J_{\perp} \approx A^{**} T^2 \exp\left(-\frac{q\Phi_B}{k_B T}\right) \left[\exp\left(\frac{qV}{k_B T}\right) - 1\right], \quad (19.2)$$

其中 Φ_B 为势垒高度， A^{**} 为有效 Richardson 常数。式 (19.2) 也是电学边界模型。它在 PCSEL 中的意义不在于鼓励你把每个接触都建成复杂的热发射边界，而在于提醒：如果接触没有真正做成低阻欧姆注入，却在模型里被理想欧姆边界替代，那么求得的阈值电流很可能过于乐观。

19.3 接触几何和接触电阻为什么会直接进入模式问题

接触不仅有材料问题，还有几何问题。PCSEL 常见的接触布局包括顶部环形接触、部分开窗接触、局部注入窗口以及带透明导电层的面接触方案。几何一旦有限，电流就不再能“从整个器件上表面同时垂直打进来”，而必须先接触附近注入，再通过横向导电层扩展到其他区域。于是，接触几何本身就成为模式选择的一部分。

如果接触只覆盖中心区域，中心附近更容易首先达到阈值载流子密度；如果接触为环形，边缘区域可能先获益；若透明 spreading layer 的片电阻过大，离接触越远的区域掉压越多，量子阱填充就越弱。这样一来，不同横向模因场分布不同，看到的有效模增益也不同。换句话说，接触几何会把 第十八章 的局域增益模型投影到大面积模式上，相当于第一道空间滤波器；它不能只按电路细节处理。

19.4 什么是 current spreading layer，它在 PCSEL 里到底帮了什么忙

current spreading layer 的器件意义，可以用一句更精确的话概括：它要让电流分布的空间尺度尽可能接近目标模式所需的增益分布尺度，“把电流摊开”只是表面描述。对于大面积 PCSEL，这种空间尺度匹配具有决定性作用。因为模式可以在整个孔阵上扩展，而金属接触和注入窗口通常远比这个模式范围更局部。若没有横向低阻 spreading layer，电流几乎一定在接触附近拥挤，远处区域难以注入。

最小的二维 spreading 模型可写成薄层内的片电流连续方程

$$\nabla_{\perp} \cdot (\sigma_s t_s \nabla_{\perp} V_s) = J_{\perp}, \quad (19.3)$$

其中 σ_s 是 spreading layer 电导率， t_s 是其厚度， V_s 为该层的横向电势， $J_{\perp}$ 是注入到下方结区的法向电流密度。式 (19.3) 属于电学简化模型。若再用局部传输关系

$$J_{\perp} \approx \frac{V_s - V_j}{\rho_c}, \quad (19.4)$$

则可以得到一个常见的工程估计：电流扩展长度约为

$$L_s \sim \sqrt{\frac{\rho_c}{R_{\square}}}, \quad R_{\square} = \frac{1}{\sigma_s t_s}, \quad (19.5)$$

其中 $R_{\square}$ 是 spreading layer 的片电阻。式 (19.5) 是工程估计公式，不应当作严格器件定理。它的教学价值在于给出一个极清晰的判断：如果器件横向尺寸远大于 L_s ，那就不应该再幻想“注入大体均匀”。

这也是大面积 PCSEL 器件和许多小面积半导体激光器的关键差别之一：在小面积器件里，注入非均匀也许只是二阶修正；在大面积 PCSEL 里，它经常是模式竞争和阈值电流的主导变量。

19.5 为什么大面积器件不可能自动均匀注入

从上面的扩展长度公式已经可以看出，只要接触比电阻有限、spreading layer 片电阻有限、器件面积继续增大，横向电势就一定存在梯度，法向注入也一定存在空间起伏。这还只是最乐观的一层。真实 PCSEL 中还叠加了以下机制：

- p 侧导电常常比 n 侧更差，因此横向掉压尤其容易发生在上侧；

- 接触窗口、孔阵列蚀区与绝缘区往往交替出现，导致横向导电路径本身并不均匀；
- 异质结和阻挡层使不同区域的纵向注入阻抗不同；
- 温升又会进一步改写局部电阻和注入效率。

因此，对大面积 PCSEL 来说，“均匀注入”不应被当作默认现实，而应被当作需要专门证明的特例。只要没有接触设计、spreading layer、孔径和热管理的联合论证，就不能把均匀增益模型视作器件模型的起点。

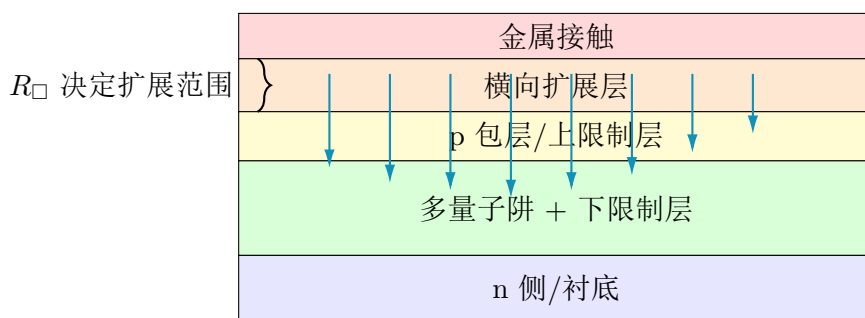


图 19.1: 横向扩展层与注入非均匀性的关系示意。即使总电流足够，只要横向扩展能力不足，量子阱中的有效载流子分布也不会自动均匀。

19.6 电流孔径、注入窗口与模式竞争为什么直接耦合

“电流孔径”在这里不应被狭义理解成某一种特定工艺结构，而应理解为：最终允许电流高效进入有源区的有效横向区域。它可能由金属开窗、绝缘层开口、局部氧化、隧穿结布局或 spreading layer 电势分布共同决定。对 PCSEL 来说，这个有效孔径很少恰好与目标模式的横向场分布完全一致。

如果电流孔径比目标模式小得多，那么模式外围区域无法获得足够增益，结果会偏向那些更局域的模式；如果孔径形状本身不对称，则本该简并的模式可能被人为分裂，偏振和远场也可能一起偏移；如果孔径位置与接触几何共同导致中心与边缘注入差异，则模式竞争会从“谁损耗更低”转变为“谁更会吃到局域增益”。因此在 PCSEL 中，注入孔径和模式选择不能分属两个独立章节，它们实际上在器件里是同一件事的两个面。

若只想先得到一个最小横向扩展模型，可把 current spreading layer 近似成一张片电阻为 $R_{\square}$ 的二维导电片，其片上电位 $V_s(x, y)$ 满足

$$\nabla_{\parallel}^2 V_s - \frac{V_s - V_{\text{cont}}}{L_s^2} = 0, \quad (19.6)$$

其中 V_{cont} 是接触电位， L_s 是电流扩展长度。[工程经验] 对 GaN 基面发射结构，p-GaN 片电阻常见在 10^3 – $10^4 \Omega/\square$ 量级，对应 L_s 可落在“几微米到十几微米”的窗口；但该窗口强依赖掺杂、透明导电层和接触工艺，不应跨平台直接套用。式 (19.6) 不能替代完整漂移-扩散求解，但它在教学上有价值：只要 L_s 明显小于器件横向尺寸（例如器件直径约 $100 \mu\text{m}$ ），均匀注入就不再是默认可信假设。

19.7 漂移-扩散方程组：器件模型的骨架

把上面的直觉写成可求解模型，基础框架仍是 Poisson 方程、电子和空穴的漂移-扩散与连续性方程。首先是 Poisson 方程

$$-\nabla \cdot (\epsilon \nabla \phi) = q(p - n + N_D^+ - N_A^-), \quad (19.7)$$

其中 ϕ 为电势， n, p 为电子与空穴浓度， N_D^+, N_A^- 为离化施主和受主浓度。式 (19.7) 属于电学模型。

电子和空穴电流密度写为

$$\mathbf{J}_n = q\mu_n n(-\nabla \phi) + qD_n \nabla n, \quad (19.8)$$

$$\mathbf{J}_p = q\mu_p p(-\nabla \phi) - qD_p \nabla p, \quad (19.9)$$

连续性方程写为

$$\nabla \cdot \mathbf{J}_n = q(R - G), \quad (19.10)$$

$$\nabla \cdot \mathbf{J}_p = -q(R - G), \quad (19.11)$$

其中 R 表示总复合率， G 表示产生项。对电注入 PCSEL 来说， G 主要体现在边界与接触的注入通量中。式 (19.8) 至式 (19.11) 都属于电学模型。这里还要把符号约定明确：本书统一取 $\mathbf{E} = -\nabla \phi$ 。因此电子漂移项写成 $q\mu_n n(-\nabla \phi)$ ，空穴漂移项写成 $q\mu_p p(-\nabla \phi)$ ，这样才与第二十九章中的附录方程和半导体器件教材里的静电势定义保持一致。

对第一次接触器件方程的读者，更稳妥的读法是把这三组方程分工记清。Poisson 方程决定电势和电场如何由空间电荷重新分布；漂移-扩散式决定给定电场和浓度梯度后，电子与空穴怎样流动；连续性方程决定这些流动是否真的在某一区域积累或耗尽载流子。只有三者联立， $n(\mathbf{r})$ 、 $p(\mathbf{r})$ 才不只是“看起来合理的分布图”，而是一个与边界条件和复合过程自洽的器件工作点。

边界条件决定这组方程如何进入器件几何。最常见的最低层写法是：欧姆接触近似固定电势或准费米能级；绝缘边界满足

$$\mathbf{J}_n \cdot \hat{\mathbf{n}} = 0, \quad \mathbf{J}_p \cdot \hat{\mathbf{n}} = 0;$$

而肖特基或有限复合速度边界则通过表面通量关系进入。也正因此，接触从一开始就是漂移-扩散问题定义的一部分，不能等求解完成后再补成工程细节。

欧姆接触 vs 肖特基接触的最小物理图像：欧姆接触可理解为“金属与半导体在该边界几乎不形成注入势垒”，所以数值上常用固定电势或固定准费米能级；肖特基接触则存在势垒高度，注入通量受势垒与局域电场共同控制，边界条件通常写成温度相关的热发射/隧穿通量模型，而不是简单 Dirichlet 条件。

直觉框

准费米能级为何必须单独定义：平衡态只有一个费米能级 E_F 。注入工作时电子与空穴处于“各自局域近平衡、彼此不再全局平衡”的状态，因此要分别引入电子准费米能级 E_{F_c} 与空穴准费米能级 E_{F_v} ，并用

$$n \approx N_C \exp\left(\frac{E_{F_c} - E_C}{k_B T}\right), \quad p \approx N_V \exp\left(\frac{E_V - E_{F_v}}{k_B T}\right)$$

把电学求解得到的载流子密度和能带图连接起来。两条准费米能级之间的分离，正是有源区可提供净增益的直接能带判据。

若载流子仍可近似看作处于局域热平衡，则扩散系数和迁移率之间满足 Einstein 关系

$$D_n = \mu_n \frac{k_B T}{q}, \quad D_p = \mu_p \frac{k_B T}{q}. \quad (19.12)$$

式 (19.12) 的物理含义是：热涨落导致的扩散和电场驱动的漂移，并非两套互不相干的机制，而是同一热平衡统计的两种表现。对 PCSEL 器件建模来说，它给出了一个最基本的量纲与数量级一致性检查。

这组方程本身并不新鲜，但在 PCSEL 中它们的求解目标与普通小面积二极管不同：我们并不只关心总电流和总压降，而更关心 $n(\mathbf{r})$ 、 $p(\mathbf{r})$ 在有源区的空间分布，因为后续它们会通过 第十八章 的增益模型进入模式竞争。

19.8 复合、漏电与局域注入效率：为什么达到同一阈值载流子需求时，所需电流仍可能完全不同

即使器件最终都需要相近的阈值载流子密度，所需的外加电流也可能相差很大，因为注入电流并不会全部转化为量子阱中的有效载流子。最常用的复合代理模型为

$$R(N) = AN + BN^2 + CN^3, \quad (19.13)$$

其中 A 、 B 、 C 分别对应 SRH、辐射和 Auger 项。式 (19.13) 属于电学与有源区动力学接口模型。其重要之处不在于三项的形式本身，而在于它强迫我们承认：阈值电流是在与多种损耗通道竞争后得到的结果。

更具体地说， AN 对应 Shockley–Read–Hall 非辐射复合，它通过深能级缺陷把电子和空穴重新吃掉，因此对材料缺陷和界面质量极其敏感； BN^2 对应辐射复合，因为需要一个电子和一个空穴同时参与，故在近似单载流子密度表述下呈二次项； CN^3 对应 Auger 复合，能量被传给第三个载流子而不是以光子形式放出，因此在高载流子密度和某些长波材料平台中尤其致命。把这三项分别看清，读者才不会把式 (19.13) 误当成只是“拟合方便三项式”。

对 PCSEL 来说，还要在式 (19.13) 之外额外警惕漏电与旁路电流。若电子越过量子阱泄漏到上层，或空穴未能真正进入所有目标阱层，则从外部看来“流进器件”的电流，并不都对目标模式有效。更准确的器件语言应聚焦“进入并有效填充有源区的局域注入效率”，而非只报外部注入电流。这也是 第十八章 里阻挡层设计必须与本章联读的原因。

这件事可以用一条最小阈值电流估计式写得更硬。若某一近似模型已经给出量子阱中的阈值载流子需求 $N_{\text{th}}(\mathbf{r})$ ，则对应的外加阈值电流至少要满足

$$I_{\text{th}} \approx \frac{q}{\eta_{\text{inj}}} \int_{V_{\text{QW}}} [AN_{\text{th}} + BN_{\text{th}}^2 + CN_{\text{th}}^3] dV + I_{\text{leak}}, \quad (19.14)$$

其中 η_{inj} 是有效注入效率, I_{leak} 汇总漏电和旁路电流。对典型 PCSEL 结构, η_{inj} 常先按 0.6–0.9 做 [本书算例] 敏感度扫描, 再由具体结深、阻挡层和 spreading layer 设计收紧。式 (19.14) 的核心意义是把“同一个 N_{th} 为何会对应完全不同的 I_{th} ”写成公式: 只要复合通道更强、注入效率更低或漏电更大, 阈值电流就会明显抬升, 而非给出单一高保真数值。

19.9 把局域载流子分布重新接回模增益: 模式竞争在这里被重写

当前面所有电学求解完成后, 本章向光学模型输出量子阱中的局域载流子分布 $N(x, y, z)$; 总电流值本身不足以完成耦合。对于候选模式 m , 更一般的有效模增益应写成

$$g_{m,\text{eff}} = \frac{\int_{V_{\text{QW}}} g_{\text{mat}}(\omega_m, N(\mathbf{r}), T(\mathbf{r})) \varepsilon(\mathbf{r}) |\mathbf{E}_m(\mathbf{r})|^2 dV}{\int_V \varepsilon(\mathbf{r}) |\mathbf{E}_m(\mathbf{r})|^2 dV} - \alpha_{m,\text{int}}. \quad (19.15)$$

式 (19.15) 属于电学-光学耦合模型。它说明了为什么注入分布会直接反馈到模式竞争: 不同模式采样的是不同的空间权重, 因而对同一组 $N(\mathbf{r})$ 的“感受”不同。

这条式子也清楚地解释了为什么大面积 PCSEL 绝不可能把均匀增益当作默认现实。只要接触有限、片电阻有限、孔径有限或热效应存在, $N(\mathbf{r})$ 就必然不是常数, 模式竞争排序也就不会等同于冷腔排序。图19.2展示了方形 p-接触电极下方的二维电流密度分布。边缘处电流密度高于中心是常见现象; “可达中心的 3–5 倍”在本书中按 [教学示意] 处理, 用于提示电流拥挤风险而非给出普适系数。

设两个候选模式在冷腔中损耗非常接近。模式 A 的横向场更集中在器件中心, 模式 B 更偏向边缘。如果接触与 spreading layer 设计导致中心注入远强于边缘, 那么式 (19.15) 会使模式 A 获得更大有效模增益, 即便模式 B 在被动结构里原本拥有略高的 Q 。这也是“高 Q 不自动等于最低阈值”的器件版解释。

19.10 阈值电流的变量链条

到这里, 可以把器件级阈值链条写得足够清楚:

$$I \rightarrow (V_s, \phi, n, p, \mathbf{J}) \rightarrow N(\mathbf{r}) \rightarrow g_{\text{mat}}(\mathbf{r}) \rightarrow g_{m,\text{eff}} \rightarrow g_{\text{th},m}. \quad (19.16)$$

阈值电流 I_{th} 的定义, 是使某一目标模式第一次满足

$$g_{m,\text{eff}} = g_{\text{th},m} \quad (19.17)$$

的最小外加电流。式 (19.16) 和式 (19.17) 的意义在于, 它们把“冷腔结论不能直接变成器件结论”这句话精确化了: 冷腔分析只能给出链条最右端的 $g_{\text{th},m}$; 而 I 如何经过接触、扩展、复合、漏电和空间分布变成 $g_{m,\text{eff}}$, 是器件求解的主体。

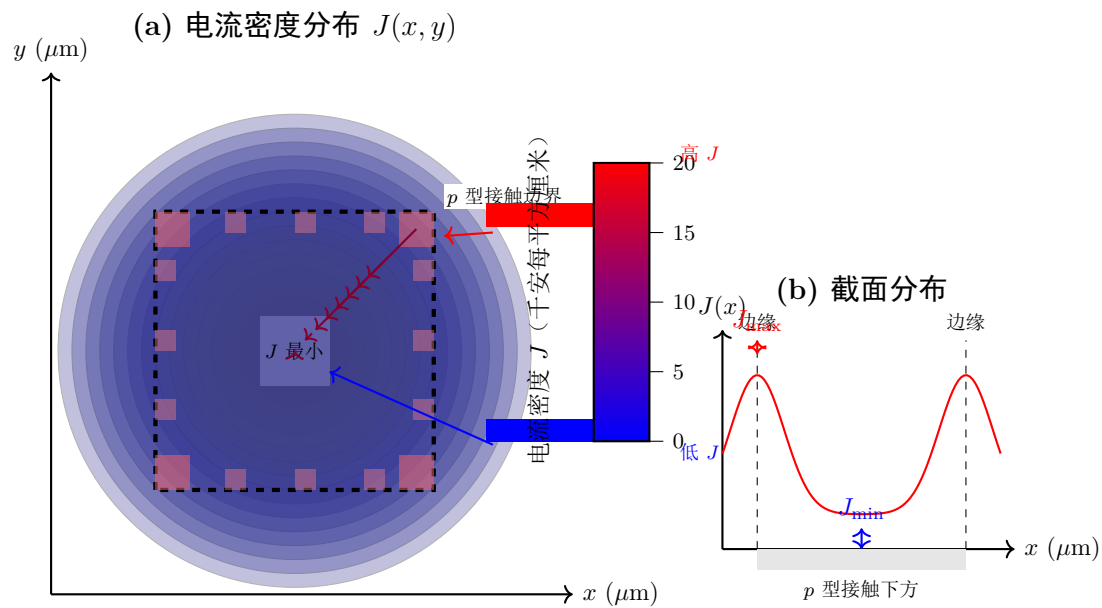


图 19.2: PCSEL 中的电流拥挤效应。(a) 方形 p 接触电极下方的二维电流密度分布 $J(x, y)$ 热力图: 红色表示高电流密度, 蓝色表示低电流密度。四个角点和边缘处的电流密度最高, 因为横向电流倾向于沿电阻最小的路径流动。(b) 沿中心截面的电流密度分布曲线 $J(x)$, 清楚显示边缘峰值和中心凹陷的特征。电流拥挤会增强边缘处的俄歇复合、局部焦耳热并引入增益分布非均匀。

19.11 为什么本章天然要求下一章的热反馈进入

严格地说，只要器件是电注入的，本章的等温结果就只是一阶近似。因为接触电阻、spreading layer 片电阻和局域复合都会发热，而温升会反过来改写迁移率、接触特性、复合系数和增益谱。也就是说，第二十章 将闭合本章尚未完成的器件回路，远不止在末端增加一项热修正。

因此，最稳妥的理解方式是：本章建立的是电注入与空间载流子分布的骨架，第二十章 会把这个骨架放到自洽工作点里。只要读者守住等温近似的边界，就不会再把本章的等温阈值电流估计误当作最终器件结论。

易错点

把“理想欧姆接触 + 均匀注入 + 固定温度”下得到的阈值电流，当作真实 PCSEL 的器件阈值，是一种系统性的过度乐观。对大面积器件尤其如此，因为接触几何、spreading layer 和注入孔径本身就会把增益分布空间化。

回看本章

电注入 PCSEL 的阈值电流问题，本质上是在问：有限接触和有限横向导电能力下，外部电流怎样通过接触边界、current spreading layer、异质结、复合与漏电，转化为与目标模式匹配的局域载流子分布。欧姆/肖特基接触通过边界条件进入模型，接触电阻和接触几何重塑横向掉压，电流孔径与 spreading length 决定注入是否接近均匀；Poisson+ 漂移-扩散 + 连续性方程给出了器件电学骨架，而有效模增益必须通过局域载流子分布重新加权。到本章结束时，读者应当已经完全清楚：阈值电流是一个空间分辨的器件工作点量，不能由阈值增益简单换算单位得到。

练习题

1. [基础] 解释为什么电流扩展长度 L_s 与器件横向尺寸的比较，可以作为判断“均匀注入是否可信”的第一道工程标准。
2. [基础] 比较欧姆接触和肖特基接触在模型中如何进入，并说明为什么把非理想接触理想化为欧姆边界会导致阈值电流偏低。
3. [进阶] 结合式 (19.15)，说明电流孔径如何通过局域增益分布改变模式竞争，而不仅仅改变总阈值电流。
4. [综合] (思路提示：先定层级，再列关键式) 讨论为什么在大面积 PCSEL 中，current spreading layer 的设计不能被视为“只影响电阻，不影响光学”的附属问题。

延伸阅读

- 下一章将把本章的电流分布、Joule 热和非辐射复合热接成热-电-光闭环，并进一步比较不同材料平台如何改写这条链。阅读时应同时核对本章变量在后续模型中的定义和适用范围。
- 若想补充漂移-扩散、接触模型和电流扩展的经典背景，可参考 [10] 中相应章节，但阅读时应始终把关注点放回 PCSEL 的大面积模式与非均匀注入问题。阅读时应重点核对相关模型假设、参数范围与文献适用条件。
- 若想对照真实电注入 PCSEL 的器件案例，可参考 [20, 69, 80]。阅读时应重点核对相关模型假设、参数范围与文献适用条件。

第二十章 热效应、非均匀性与热-电-光耦合

读完本章，你应能

- 理解热效应为何会主动重下载流子分布、增益谱、损耗和模式排序，因而不能作为冷腔分析的末端修正。
- 掌握 Joule 热、复合热、温度依赖材料参数和多物理场迭代之间的关系，并能判断什么时候简化热模型已经不够。
- 对 GaAs、InP、GaN 等主要材料平台在折射率、增益谱、热导、掺杂损耗和工艺难点上的差异形成器件级认识。

读前准备 建议已掌握 第七章、第十八章和第十九章 中关于阈值条件、局域增益分布和电注入输运的内容。

阅读主线 本章所处模型层级：L4

到上一章为止，我们已经知道阈值电流为什么是一个空间分辨的器件问题，但那仍是在等温框架下建立的。真实 PCSEL 一旦开始导通，接触、电流扩展层、阻挡层和有源区就会发热，而温度又会反过来改写迁移率、复合、折射率、材料增益和吸收。于是，第十七章至第十九章 中看似分别处理的外延层、有源区和输运，会在本章重新闭合成一个热-电-光回路。本章先从热源开始，再讨论温升如何同时回写到光学和电学；随后解释热斑、空间烧孔和模式竞争的关系，最后用一个平台差异小节说明：同一套 PCSEL 理论框架在 GaAs、InP、GaN 上会怎样“变形”。

贯穿案例 B0/R1：热模型只报告算得出的量 B0 先以 940 nm、直径 200 μm 的教学热源建立量级感，输出完整的 $T(x, y, z)$ 、最大温升位置以及由 $\Delta n(T)$ 和 $\lambda_g(T)$ 导致的失谐。R1 的 3 mm、50 W 连续输出则提供放大后的现实边界 [74, 50]：它说明热补偿必须进入设计闭环，却不能在缺少封装热阻、耗散功率和电流分布时推出一个“文献器件温升”。本章后续所有数值都据此标明是 B0 教学估算还是 R1 文献指标。

20.1 为什么热效应在 PCSEL 中比在许多小面积激光器里更具结构性

并非所有半导体激光器都对热效应同样敏感。PCSEL 的特殊性在于：它一方面追求较大的发光面积和面内相干反馈，另一方面又要求单模、窄远场和稳定偏振。前者让横向热扩散尺度变长、局部热斑更容易出现；后者则意味着任何小范围的折射率、增益或损耗扰动，都可能被模式选择机制放大成宏观输出差异。

因此，热效应在 PCSEL 中最危险之处，是它会改写前面章节建立的排序关系；平均温升只描述了其中一部分。某个在冷腔中看起来损耗最低的模式，到了真实工作点下可能因为热斑位置、局域折射率抬升或注入退化而不再最优。

20.2 热源必须写成三维分布

热学建模应先明确热源的空间结构，再写热方程。对电注入 PCSEL 来说，最常见的热源至少包括三类。

第一类是 Joule 热：

$$Q_{\text{Joule}} = \mathbf{J} \cdot \mathbf{E}_{\text{elec}}, \quad (20.1)$$

它主要沿高电阻接触区、spreading layer 和局部电流拥挤通道集中。式 (20.1) 属于耦合模型中的热源项，因为它依赖 第十九章 的电学求解结果。

第二类是非辐射复合热。若以局域非辐射复合率 R_{nr} 表示，则可写成

$$Q_{\text{nr}} \propto R_{\text{nr}} \Delta E, \quad (20.2)$$

其中 ΔE 表示相关能量释放尺度。它同样属于耦合模型中的热源项。

非辐射复合热的微观来源 式 (20.2) 中的 R_{nr} 包含多种复合机制：

- **Shockley-Read-Hall (SRH) 复合**：通过缺陷态的非辐射复合， $R_{\text{SRH}} \propto np - n_i^2$
- **Auger 复合**：三粒子过程， $R_{\text{Auger}} = C_n n^2 p + C_p n p^2$ ，在高载流子浓度下主导
- **表面复合**：在刻蚀侧壁和界面处，先写面复合通量

$$U_{\text{surf}} = S_{\text{surf}} \frac{np - n_i^2}{n + p + 2n_i}, \quad [U_{\text{surf}}] = \text{cm}^{-2}\text{s}^{-1},$$

再通过有效厚度 t_{eff} 换算成体复合率

$$R_{\text{surf,vol}} = \frac{U_{\text{surf}}}{t_{\text{eff}}}, \quad [R_{\text{surf,vol}}] = \text{cm}^{-3}\text{s}^{-1}.$$

每次非辐射复合事件释放的能量约为带隙 E_g ，这部分能量最终转化为晶格热能。对 GaAs 基 InGaAs 量子阱，Auger 与 SRH 复合的相对贡献取决于载流子密度、温度、阱结构和缺陷密度；在高载流子浓度或高温区域，Auger 复合的占比可能显著上升。

第三类常被低估的热源，是光吸收相关发热，包括自由载流子吸收、金属邻近吸收和向非目标辐射通道泄漏后再被材料吸收的能量。工程上可以把它们合并写成一个有效光学热源 $Q_{\text{opt,abs}}$ ，并在多物理场求解中由光学模块回写给热学模块。

于是总热源最小可写成

$$Q_{\text{tot}} = Q_{\text{Joule}} + Q_{\text{nr}} + Q_{\text{opt,abs}}. \quad (20.3)$$

式 (20.3) 的作用是把“哪些功率最终变成热”这件事从口头描述变成热方程右端的输入。

这三类热源的分布通常并不重合。Joule 热可能集中在接触附近，复合热主要集中在量子阱附近，光吸收热又可能靠近金属或表面工艺层。

20.3 热方程求解的是温度场

最基础的稳态热学模型最好写成保守形式

$$-\nabla \cdot [\kappa_T(\mathbf{r}) \nabla T(\mathbf{r})] = Q_{\text{tot}}(\mathbf{r}), \quad (20.4)$$

其中 κ_T 为热导率， $T(\mathbf{r})$ 为温度场，而 Q_{tot} 正是式 (20.3) 中定义的总热源。式 (20.4) 属于**热学模型**。这一方程的教学意义在于：温度升高几乎一定具有空间结构。

这时边界条件成为热模型本身的一部分，不能降为实现细节。最常见的两类写法是

$$T|_{\partial\Omega_D} = T_0, \quad (20.5)$$

表示某一边界与理想热沉紧耦合，温度被固定；以及

$$-\kappa_T \partial_n T = h(T - T_{\text{amb}}), \quad (20.6)$$

表示边界通过有限换热系数 h 与环境交换热量。前者更像强制热沉，后者更像封装、空气侧或有限导热胶层的等效模型。

完整的热传导方程。若考虑瞬态过程，三维非稳态热传导方程为

$$\rho C_p \frac{\partial T}{\partial t} - \nabla \cdot (\kappa_T \nabla T) = Q_{\text{tot}}(\mathbf{r}, t), \quad (20.7)$$

其中 ρ 是材料密度， C_p 是定压比热容。稳态近似下该方程即退化为前面的式 (20.4)。

瞬态热效应的适用条件

稳态近似式 (20.4) 在以下情况可能失效：

- **脉冲工作模式：**当脉冲宽度 $\tau_{\text{pulse}} \lesssim \tau_{\text{th}} = R_{\text{th}} C_{\text{th}}$ 时，必须使用瞬态方程20.7
- **高速调制：**调制频率 $f_{\text{mod}} > 1/(2\pi\tau_{\text{th}})$ 时，热场跟不上电信号变化
- **启动瞬态：**器件开启后 $t < 3\tau_{\text{th}}$ 期间，温度场处于建立过程

对 CW 工作的 PCSEL，若热时间常数 $\tau_{\text{th}} \ll$ 观测时间尺度，稳态近似通常足够准确。典型 GaAs 基 PCSEL 的 τ_{th} 在 $1\ \mu\text{s}$ to $10\ \mu\text{s}$ 量级 (芯片级快速分量) 至 $0.1\ \text{ms}$ to $1\ \text{ms}$

(含封装慢速分量)。

若只想先建立热学量纲感，还可以把空间分布压缩成一个集总模型：

$$C_{th} \frac{d\Delta T}{dt} + \frac{\Delta T}{R_{th}} = P_{heat}(t), \quad (20.8)$$

其中 R_{th} 是热阻， C_{th} 是等效热容， P_{heat} 是总发热功率。稳态下

$$\Delta T_{ss} = R_{th} P_{heat}, \quad \tau_{th} = R_{th} C_{th}. \quad (20.9)$$

式 (20.8) 和式 (20.9) 当然不能取代三维热场，但它们给了读者一把最基本的尺子：到底是“平均热负载已经过大”，还是“平均温升还行，但空间热斑已经毁掉模式稳定性”。对直径约 200 μm to 300 μm 、衬底厚度约 200 μm 的 GaAs 基 PCSEL, R_{th} 常见于 10 K/W to 100 K/W 的工程窗口；这里按 [工程经验] 处理，因为该值强依赖键合方式、散热器材料、热源定义和封装路径。实作时必须用结构化热仿真或实测热阻回标替代“默认常数”。

热时间常数的最小估算例子（双时间尺度） 很多器件会同时出现“快热常数 + 慢热常数”。文献中，半导体激光器已报道约 5.6 μs 与 7.2 μs 的快速热时间常数 [12]，VCSEL 的瞬态热成像也测得约 $\tau \approx 9.7 \mu\text{s}$ 的快速分量 [15]。这里的 5.6 μs to 9.7 μs 应理解为芯片级快速热分量（有源区/芯片内部热扩散），不包含封装与热沉路径的慢时间常数。

对本书 benchmark 量级，可用两级集总模型做估算。热阻估算需注意几何形状：对方形芯片上的局部热源，应使用热扩展电阻公式 $R_{th} \approx 1/(2\kappa d)$ （半球形扩展）或 $R_{th} \approx L/(\kappa A)$ （一维近似）。对 GaAs ($\kappa \approx 55 \text{ W/mK}$) 芯片，若热源直径约 200 μm 、衬底厚度 200 μm ：

- 一维近似： $R_{th,1D} \approx L/(\kappa A) \approx 200 \mu\text{m} / (55 \text{ W/mK} \cdot \pi \cdot 100^2 \mu\text{m}^2) \approx 115 \text{ K/W}$
- 半球形扩展： $R_{th,spread} \approx 1/(2\kappa d) \approx 1/(2 \cdot 55 \text{ W/mK} \cdot 200 \mu\text{m}) \approx 45 \text{ K/W}$

实际值介于两者之间，取决于热源分布和边界条件。取中间值 $R_{th,chip} \approx 70 \text{ K/W}$ ，热容 $C_{th,chip} \approx 1.2 \times 10^{-6} \text{ J/K}$ ，则：

$$\tau_{chip} = R_{th,chip} C_{th,chip} \approx (70 \text{ K/W}) \times (1.2 \times 10^{-6} \text{ J/K}) \approx 84 \mu\text{s}.$$

封装级： $R_{th,pkg} \approx 35 \text{ K/W}$ ， $C_{th,pkg} \approx 50 \times 10^{-6} \text{ J/K}$ ，则：

$$\tau_{pkg} = R_{th,pkg} C_{th,pkg} \approx (35 \text{ K/W}) \times (50 \times 10^{-6} \text{ J/K}) \approx 1.75 \text{ ms}.$$

这说明”本征有源区升温”和”封装/热沉缓慢回温”通常不在同一时间尺度。若只报告一个热时间常数，往往会把动态漂移和慢热滚降机制混在一起。本例的集总参数仅用于说明双时间尺度概念，不可直接用于器件设计。

对 PCSEL 而言，平均温升只适合回答非常粗的问题，例如“某个方案是否在总体热负载上明显更差”。但只要你关心阈值附近模式排序、高功率单模稳定性、远场主瓣是否漂移、偏振是否翻转，平均温升就不够了。因为这些现象敏感的是 $T(x, y, z)$ 的梯度和局域峰值，而不是整体平均值。一个局部热斑就足以把模式中心拉偏、把局部增益峰推离目标波长，进而改变模式竞争。

20.4 温升如何回写到光学：折射率、增益谱、吸收和冷腔频率全都会动

热反馈对光学的第一条路径，是折射率变化。最常见的局部近似写法是

$$n(T) \approx n(T_0) + \frac{\partial n}{\partial T}(T - T_0), \quad (20.10)$$

式 (20.10) 属于光学模型的热修正项。一旦 T 在横向上不均匀，器件内部就会出现热透镜；一旦 T 在纵向上不均匀，模场也会被拖向不同层。于是，第十七章建立的纵向重叠与上下辐射不对称，到了工作点就不再保持冷腔值。

体材料 dn/dT vs. 波导有效折射率 dn_{eff}/dT

需特别注意：式 (20.10) 中的 $\partial n/\partial T$ 在不同语境下指代不同对象：

- 体材料系数 $(dn/dT)_{\text{bulk}}$ ：描述块体半导体材料（如 GaAs、InP）的折射率温度依赖性，典型值 $\sim 2 \times 10^{-4} \text{ K}^{-1}$
- 波导有效系数 $(dn_{\text{eff}}/dT)_{\text{wg}}$ ：描述波导模式有效折射率的温度依赖性，需考虑模式 confinement 和多层结构影响

对典型的 PCSEL 波导结构，由于模式部分分布在低折射率包层中，有效热光系数通常略低于体材料值：

$$\frac{dn_{\text{eff}}}{dT} \approx \Gamma_{\text{wg}} \left(\frac{dn}{dT} \right)_{\text{active}} + (1 - \Gamma_{\text{wg}}) \left(\frac{dn}{dT} \right)_{\text{clad}}, \quad (20.11)$$

其中 Γ_{wg} 为模式在有源区的 confinement 因子。对 GaAs 基 PCSEL， $(dn_{\text{eff}}/dT)_{\text{wg}} \approx (1.5\text{--}2.0) \times 10^{-4} \text{ K}^{-1}$ ，略低于体材料 GaAs 的 $(dn/dT)_{\text{bulk}} \approx 2.5 \times 10^{-4} \text{ K}^{-1}$ 。

第二条路径是材料增益谱漂移。第十八章已经把材料增益写成

$$g_{\text{mat}}(\omega, N, T), \quad (20.12)$$

这意味着温度一升高，透明载流子密度、增益峰位置和微分增益都可能变化。于是，同一个外加电流在不同温度下未必对应同样的阈值裕量。

第三条路径是内部损耗的热增强。自由载流子吸收、接触邻近损耗和某些材料吸收过程都可能随温度增强。于是，温升会同时平移工作波长、提高损耗并降低增益匹配。对 PCSEL 来说，这三条路径一旦叠加，就足以把冷腔中的高 Q 候选模淘汰掉。

把这些热回写写成一组最小线性化关系，就是

$$\Delta n \approx \frac{\partial n}{\partial T} \Delta T, \quad (20.13)$$

$$\Delta g_{\text{mat}} \approx \frac{\partial g_{\text{mat}}}{\partial T} \Delta T, \quad (20.14)$$

$$\Delta \alpha \approx \frac{\partial \alpha}{\partial T} \Delta T, \quad (20.15)$$

$$\Delta \mu \approx \frac{\partial \mu}{\partial T} \Delta T. \quad (20.16)$$

式 (20.13) 至式 (20.16) 的意义, 在于把“热会改写光 and 电”这句口头话, 变成可进入迭代求解的接口: 光学模块最关心 $\Delta n, \Delta g, \Delta \alpha$, 电学模块最关心 $\Delta \mu$ 与相关接触/复合参数的变化。为避免跨平台误读, 本章把定量句统一写成“范围 + 条件 + 引文”: [文献值] 在近红外 GaAs/InP 相关平台, $\partial n / \partial T$ 常见于 10^{-4} K^{-1} 量级 (典型约 $2 \times 10^{-4} \text{ K}^{-1}$ 到 $3 \times 10^{-4} \text{ K}^{-1}$) [11]; 模增益温度斜率通常为负且依赖工作点, 在阈值附近可近似为

$$\left. \frac{\partial g_{\text{mod}}}{\partial T} \right|_{n=n_{\text{th}}} \approx \Gamma_{\text{opt}} \left. \frac{\partial g_{\text{mat}}}{\partial T} \right|_{n=n_{\text{th}}}, \quad (20.17)$$

其中 n_{th} 是阈值载流子浓度。作为 [本书算例] 或 [工程经验] 参数, 可在 -0.1 到 $-0.2 \text{ cm}^{-1} \text{ K}^{-1}$ 范围内做敏感度扫描, 而非固定常数。需注意该斜率强烈依赖于量子阱结构、应变状态和工作载流子浓度。

dn/dT 回写的数量级估算 (benchmark 量级)

取 $n_{\text{eff}} \approx 3.5$ 、 $\partial n / \partial T \approx 2.5 \times 10^{-4} \text{ K}^{-1}$ 、 $\Delta T = 10 \text{ K}$, 则

$$\Delta n_{\text{eff}} \approx \frac{\partial n}{\partial T} \Delta T \approx 2.5 \times 10^{-3}, \quad \frac{\Delta \omega}{\omega} \approx -\frac{\Delta n_{\text{eff}}}{n_{\text{eff}}} \approx -7.1 \times 10^{-4}.$$

若中心波长约 $\lambda_0 \approx 940 \text{ nm}$, 对应

$$\Delta \lambda \approx \lambda_0 \frac{\Delta n_{\text{eff}}}{n_{\text{eff}}} \approx 0.67 \text{ nm}.$$

这说明即使只有 10 K 的温升, 热致频移也已达到亚纳米到纳米量级; 当候选模间隔、远场最优或偏振稳定窗口本身就很窄时, 这个漂移足以改写模式排序。

这些多物理场耦合路径总结于图 20.1。如图所示, 温度升高 ΔT 通过四个主要通道影响器件性能: 光学通道 (折射率和带隙变化)、电学通道 (迁移率和电阻变化)、力学通道 (热应力累积), 以及它们之间可能形成的正反馈回路。理解这些耦合机制是设计高效散热方案和避免热失控的前提。

图 20.2 将热路与电路进行类比。该类比的作用, 是把复杂的三维热传导问题在给定假设下约化为等效热阻网络。

20.5 温升如何回写到电学: 热又在重写注入分布

热反馈的危险不只在光学侧。温度上升通常会降低迁移率、改变接触电阻、改写复合系数, 甚至加重局部电流拥挤。于是, 第十九章里的 $n(\mathbf{r})$ 、 $p(\mathbf{r})$ 与 $\mathbf{J}(\mathbf{r})$ 其实是温度相关的。用更直白的话说, 热场不是求出之后“交给光学后处理”的终点, 它会反过来改变电流路径本身。

这会形成一个典型的正反馈: 电流拥挤区更热, 温度升高后该区域的增益可能下降, 但接触电阻和局域导电路径又可能进一步不均匀, 最终导致载流子与模式分布一起重排。很多高功率 PCSEL 中观测到的模式切换、主瓣拉宽和输出滚降, 都可以沿着这条热-电-光正反馈去理解。

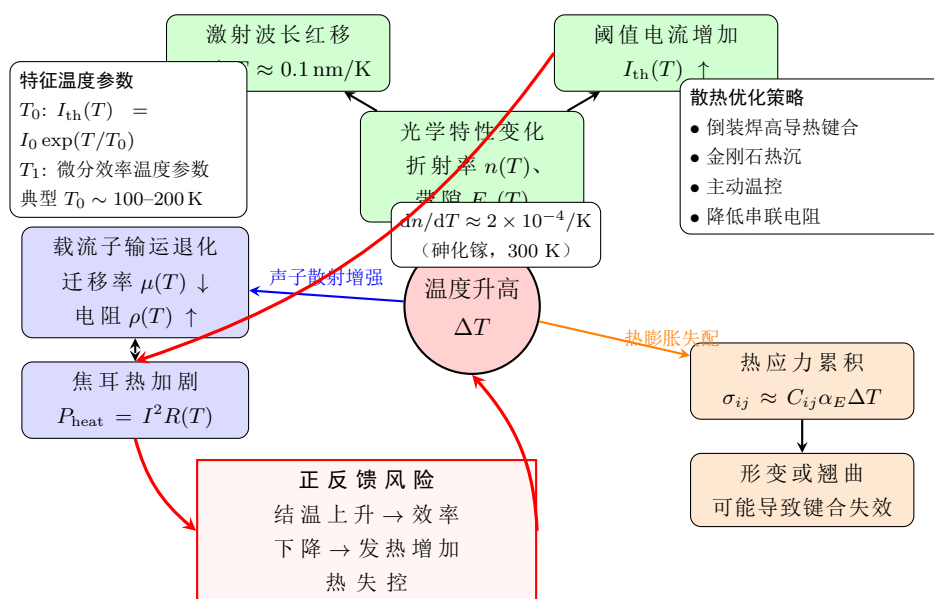
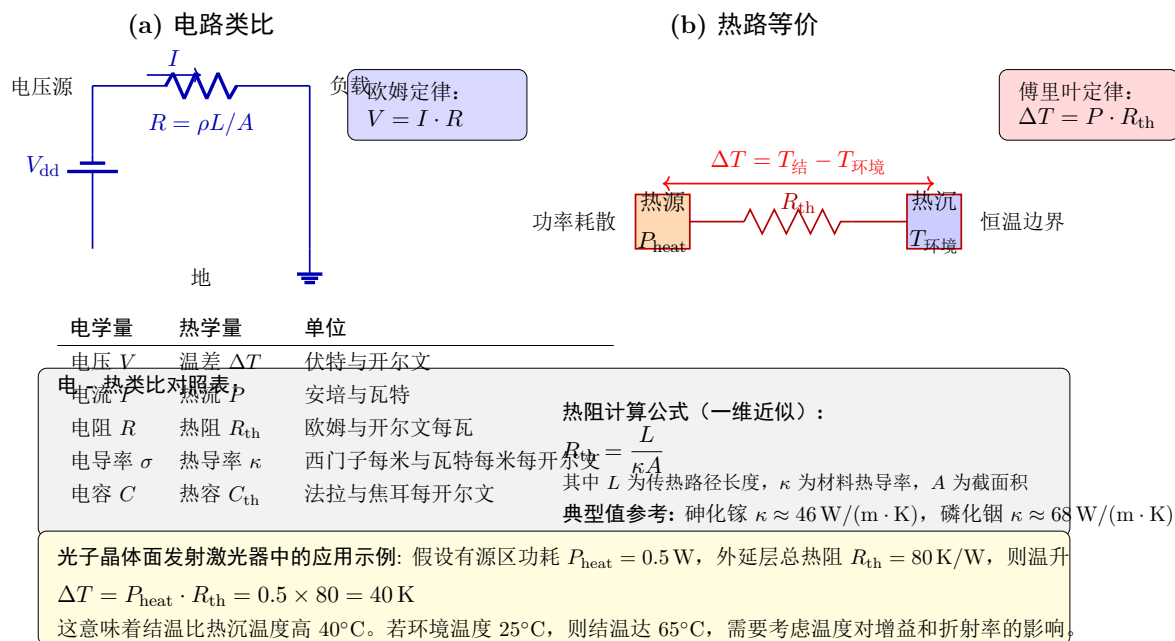


图 20.1: PCSEL 中热效应的多物理场耦合路径图

图 20.2: 热路等效电路类比图。(a) 熟悉的直流电路：电压源驱动电流流过电阻，遵循欧姆定律 $V = IR$ ；(b) 热路系统：热源驱动热流流过热阻，遵循傅里叶定律 $\Delta T = P \cdot R_{th}$ 。这种类比可以把复杂的三维热传导问题压缩为等效热阻网络，用电路分析的方法快速估算关键节点的温升。

20.6 热斑、空间烧孔和模式竞争：为什么高功率问题不能只看平均参数

当前面三章合并起来时，最值得警惕的工作点现象有三个：热斑、空间烧孔和局域损耗增强。热斑意味着局部折射率更高、材料增益峰更红移；空间烧孔意味着不同区域的载流子被模式消耗得不一样；局域损耗增强则意味着某些区域虽然有模式场，却因温度和掺杂共同作用而变成净负贡献。它们叠加后的结果是：模式竞争不再由单一冷腔损耗或单一模增益决定，而由一个同时随空间和工作点变化的分布系统决定。

因此，进入真实电注入和高功率工作区后，模式稳定性本质上是一个热-电-光分布稳定性问题，而不是带结构和冷腔 Q 能单独回答的问题（红线 #3，见第2.10节）。

例子

设某 PCSEL 在低电流下显示为法向单峰远场，但随驱动升高出现主瓣分裂。最直接的排查思路不应只盯着第六章至第九章的平面内耦合系数，而应优先检查：是否中心区域热斑抬高了折射率并拉偏模式，是否边缘区域因注入不足而形成局域负增益，以及这种空间重排是否使原本本次优的高阶模式开始获得更大局域模增益。

20.7 把第十七章至第十九章 闭成热-电-光回路

现在可以把完整的器件回路写成

$$\begin{aligned} I \rightarrow (\phi, n, p, \mathbf{J}) &\rightarrow (Q_{\text{Joule}}, Q_{\text{nr}}, Q_{\text{opt,abs}}) \rightarrow T(\mathbf{r}) \\ &\rightarrow (n(\mathbf{r}), g_{\text{mat}}(\mathbf{r}), \alpha(\mathbf{r}), \mu(\mathbf{r})) \rightarrow \text{模式、阈值、远场、偏振.} \end{aligned} \quad (20.18)$$

式 (20.18) 表示本章最核心的器件逻辑链，不能按单个方程孤立阅读。它告诉读者：第十七章的纵向模、第十八章的有源区工程和第十九章的注入分布，从这里开始由顺次单向传递转为迭代闭环。

因此，可信的器件工作点通常需要分步迭代：先在初始温度下求电学和载流子分布，再由电学和非辐射复合产生热源求热场，再更新折射率、增益、损耗和迁移率，重新解电学和光学，直到阈值、温度峰值和模式分布都收敛。只做一次“冷腔 + 固定均匀增益 + 平均温升修正”的计算，严格地说还不能叫自洽器件求解（红线 #4，见第2.10节）。

典型的收敛判据

热 - 电-光迭代可在以下任一判据满足时停止（建议同时监控三个量，并要求连续两轮都满足）：

1. 温度判据: $\max |T^{(k+1)} - T^{(k)}| < 0.5 \text{ K}$
2. 光学判据: $|\lambda_{\text{peak}}^{(k+1)} - \lambda_{\text{peak}}^{(k)}| < 0.2 \text{ nm}$
3. 器件判据: $|I_{\text{th}}^{(k+1)} - I_{\text{th}}^{(k)}| / I_{\text{th}}^{(k)} < 1\%$

实践中，通常需要 5-10 轮迭代才能达到稳定收敛。若超过 15 轮仍未收敛，应检查网

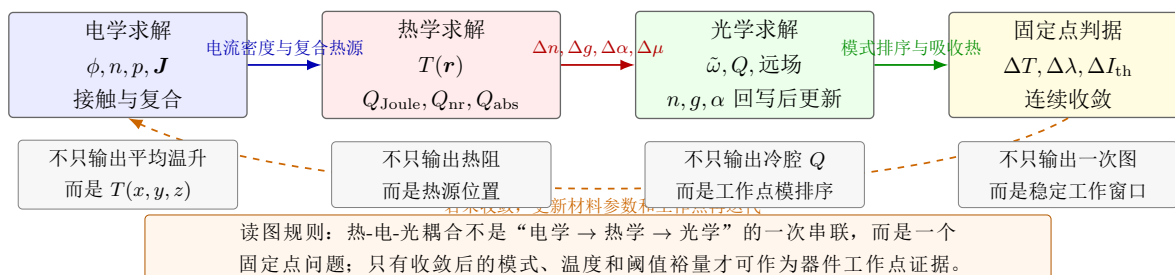


图 20.3: 热-电-光耦合的固定点迭代图像。电学给出电流、载流子和热源，热学给出空间温度场，光学在回写后的材料参数中重新评估模式和阈值；若温度、波长或阈值裕量未收敛，就必须回到电学与材料参数继续迭代。

格分辨率、材料参数的温度依赖性是否过强、或是否存在多稳态竞争。

三轮耦合的计算成本估算 以一个 $100\ \mu\text{m} \times 100\ \mu\text{m}$ 的 PCSEL 为例：

- 电学 (CHARGE): 5 分钟 (稳态漂移 - 扩散方程)
- 热学 (HEAT): 2 分钟 (稳态热传导)
- 光学 (FDTD 有限阵列): 4 小时 (若用 RCWA 单胞则 10 分钟)

若需要 6 轮迭代达到收敛，总时长可能是 24 小时量级。因此 workflow 章节建议采用分层策略：先用单胞 RCWA 等简化模型排除明显不合理的设计，再对少数最终候选方案投入全耦合计算资源。

20.8 哪些问题还可以用简化热模型，哪些已经必须做多物理场闭环

务实的器件建模不意味着一开始就做最复杂的全三维耦合，而是要根据问题目标选择模型层级。如果你的目标只是比较几个几何方案在低功率、阈值附近的相对热敏感性，那么集中热阻或一维稳态热模型往往足以筛掉明显不合理的方案。如果目标是估算阈值附近温升对材料增益峰和折射率的影响，则二维或简化三维热模型配合 第十九章 的注入分布，已经是合理的中间层。

但如果研究目标变成下列任一种，就不应该再停留在简化热模型上：高功率单模稳定性、模式切换边界、远场随驱动的系统漂移、偏振翻转、上下出光比变化、明显不对称接触布局、非均匀孔径或大尺寸器件中心/边缘性能分裂。因为这些问题敏感的是温度和注入的空间重排，而不是平均热阻本身。

20.9 不同材料平台如何让同一个 PCSEL 框架发生变形

到这里，读者已经具备理解平台差异的正确角度了：GaAs、InP、GaN 会同时改写 第十七章至第二十章 整条器件链中的几乎每一个接口量，远不止替换一套材料参数。

平台	GaAs 系	InP 系	GaN 系
典型波段与增益谱	近红外，围绕 850 nm to 980 nm 等窗口较成熟；量子阱与 AlGaAs/InGaAs 体系结合紧密	通信波段，围绕 1.3 μm to 1.55 μm ；长波长增益与温度漂移更敏感	可见到近紫外；InGaN/GaN 量子阱中极化场和应变效应更突出
折射率与纵向光学	折射率较高，纵向波导与表面耦合工程空间较大	折射率也高，但长波长下层栈厚度与吸收窗口管理更敏感	折射率相对较低，模式尺寸、表面耦合和空气孔设计尺度都会变化
热学与散热	热导中等，工艺成熟，但大面积高功率下仍要警惕热斑	热导通常较弱，长波器件对温升更敏感，热滚降问题往往更早出现	材料本征热导潜力较好，但实际器件受 p 侧导电与接触、基底和工艺路径强烈制约
掺杂损耗与注入难点	p 侧自由载流子吸收和接触层损耗需要平衡；总体工艺路线较成熟	p 侧接触和长波吸收层管理往往更苛刻，内部损耗更容易抬高阈值	p-GaN 导电与低阻接触长期是难点，spreading layer 与透明接触设计常成为核心问题
工艺与结构难点	高质量外延、刻蚀与再生长经验丰富，适合系统优化	长波材料组合、热管理与低损接触需要更谨慎协同	干法刻蚀、表面损伤、极化场、应变和接触工程往往同时限制设计窗口

表 20.1: GaAs、InP、GaN 三类 PCSEL 材料平台的器件级差异。表中不给出工艺手册式数字，而强调它们如何改写 第十七章至第二十章 建立的同一套器件框架。

这张表强调的是平台差异：同一套 PCSEL 理论框架在不同材料体系中的主要限制并不相同。GaAs 平台外延、刻蚀和接触路线较成熟，适合系统化优化；InP 平台更早受到长波热敏感和内部损耗限制；GaN 平台则常把 spreading layer、p 侧接触和极化场推到设计前台。因此，平台选择不能只看目标波长，还要判断主要复杂度将落在光学、电学、热学或工艺中的哪一段。

相同的平面内晶格参数、相似的冷腔带结构，放在不同平台上时，可能因为折射率背景、热导、p 侧导电和掺杂损耗不同，而对应完全不同的真实器件工作点（红线 #3，见第2.10节）。器件级设计的提问方式因此是：“这个平面内方案在某一平台上，是否还能通过 第十七章至第二十章 这条器件链，维持所需的模式重叠、注入均匀性和热稳定性？”

回看本章

热效应在 PCSEL 中是把 第十七章至第十九章 重新耦合成器件工作点的主导机制，而非补充项。Joule 热、复合热和吸收热共同生成三维热源，热方程给出空间温度场，而温度场又通过折射率、增益谱、吸收、迁移率和接触特性回写到模式与注入分布中。对大面积、单模、高功率 PCSEL 来说，需要分析的是热斑、空间烧孔和模式竞争的耦合，而不是单一平均温升。GaAs、InP、GaN 等材料平台则进一步表明：同一套 PCSEL 光学框架在不同器件材料上会因折射率、热导、掺杂损耗和工艺路径不同而表现出完全不同的工作约束。至此，第十七章至第二十章 构成了一套器件物理模块：外延层定义纵向模式背景，有源区定义可用增益，注入工程定义局域载流子分布，热反馈决定这个分布和模式是否能在真实工作点下稳定存在。

练习题

1. [基础] 解释为什么在大面积 PCSEL 中，平均温升通常不足以判断单模稳定性，回答中必须同时讨论热斑和模式重排。
2. [基础] 结合 式 (20.18) 说明温升如何同时通过折射率和迁移率两条路径改写模式竞争。
3. [进阶] 对比 GaAs 与 GaN 平台，说明为什么 current spreading layer 和接触工程在两者中的设计压力不同。
4. [综合]（思路提示：先定层级，再列关键式）给出一个你认为必须采用空间分辨多物理场求解的 PCSEL 问题，并说明简化热模型为什么在该问题上不够。

延伸阅读

- 下一章将把这四章建立的器件链与前面的开放电磁和数值方法重新汇总，进入大面积单模 PCSEL 的综合设计原则。阅读时应同时核对本章变量在后续模型中的定义和适用范围。

- 若想补充半导体激光器热阻、热滚降与工作点漂移的经典背景，可对照 [10] 相关章节阅读；但在 PCSEL 中，应始终额外关注热场如何通过空间分布重写模式选择。阅读时应重点核对相关模型假设、参数范围与文献适用条件。
- 若想对照高功率与连续波 PCSEL 中热-电-光约束怎样真正显形，可参考 [20, 75, 74, 69]。阅读时应重点核对相关模型假设、参数范围与文献适用条件。

第二十一章 大面积单模设计原则

读完本章，你应能

- 理解为什么“大面积”和“单模”在 PCSEL 中天然存在张力，以及这种张力为什么不能只靠单胞带结构解决。
- 掌握模鉴别必须同时发生在单胞、有限孔阵和器件工作点三个层次上的原因。
- 明确哪些设计问题主要由光学仿真回答，哪些问题必须进入电学与热学模型后才能判断。

读前准备 建议已掌握 第十四章、第十五章、第十七章、第十九章和第二十章 中关于开放结构、有限尺寸、外延层、电注入和热反馈的内容。

阅读主线 本章所处模型层级：L2→L4

前面三章已经把一个关键事实说清楚了：真实 PCSEL 的目标并非在被动冷腔里找到一个形状理想的 Gamma 点模式，而是在给定电流、温度和外延约束下，使目标模式在大面积孔阵中相对于竞争模式保持正的净裕量。这个目标决定了本章的写法。我们先解释为什么大面积单模从来不是“把面积做大而模态不变”这么简单，再建立模鉴别裕量的概念，随后把设计原则分成三个尺度来讨论：单胞尺度的辐射与对称性、有限孔阵尺度的横向包络与边缘损耗、器件尺度的注入与热分布。最后再把这些原则翻译成可执行的仿真任务，并说明在 Lumerical 与 COMSOL 风格环境中通常各自承担哪些计算。

贯穿案例 R1：用公开放大路线检验模裕量 本章从 B0 的接口模型切换到 表2.1 的 R1 实验锚点。后文把 500 μm 、1 mm 和 3 mm 三个尺度放在同一条放大链上，分别检查边缘损耗、Hermitian/非 Hermitian 耦合平衡以及热致折射率预补偿。这样，模裕量不再只是符号定义，而能与已报道的功率、发散角和线宽逐项对照。

21.1 为什么大面积单模是 PCSEL 最难也最值钱的问题

PCSEL 之所以受到持续关注，一个核心原因就是它有潜力同时获得较大出光面积和较好光束质量。这个说法的潜台词是：目标模式必须在较大横向面积上保持相干，局域振荡区域的松散拼接达不到这一要求。器件面积一旦增大，可利用的横向自由度也随之增多，竞争模数量、边缘效应、注入不均匀性和热非均匀性都会变得更显著。“大面积”因此会主动增加模式竞争的维数，不能按被动几何放大的理解。实现大面积相干的关键在于建立有效的横向耦合物

理机制。图21.1总结了三种主要的同步途径：倏逝波隧穿耦合（短程最近邻作用）、辐射场耦合（长程远场干涉）以及共享全局腔模（集体约束条件）。实际设计中往往需要综合运用这三种机制，通过适度密集的排布增强倏逝耦合，同时在边缘设置分布式反馈结构提供额外的模式选择性。

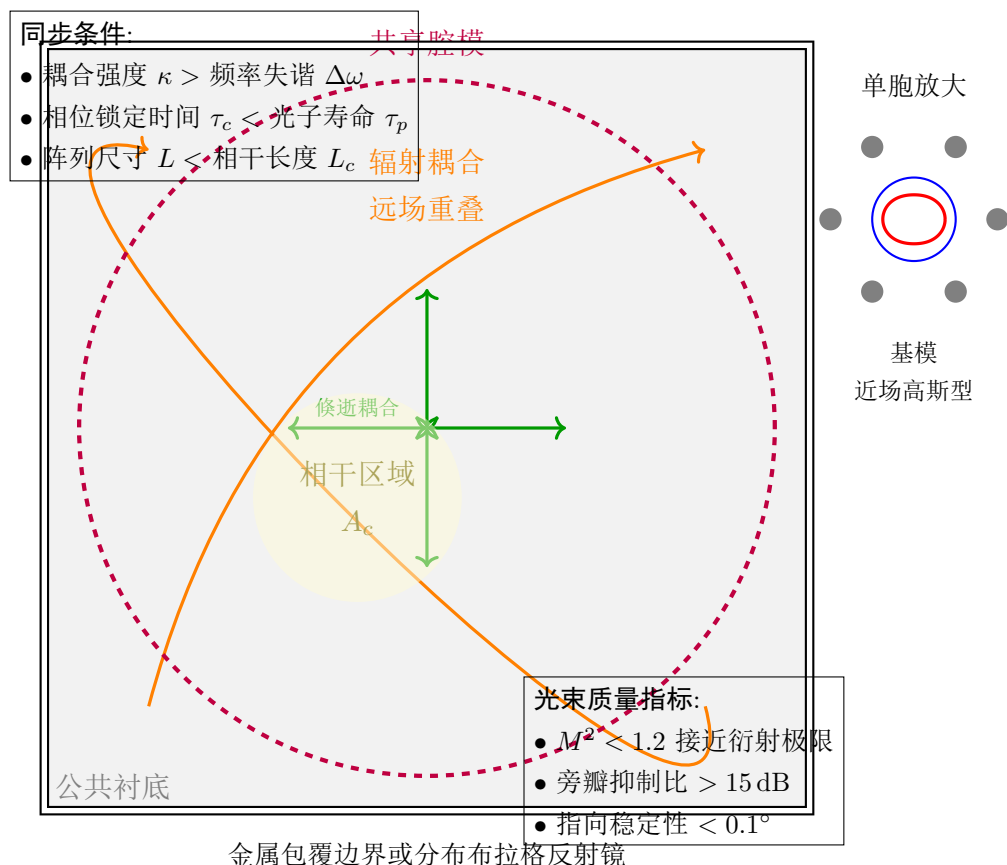


图 21.1: 大面积 PCSEL 阵列中相干锁定的三种主要横向耦合机制：倏逝近邻耦合、辐射远场耦合以及共享全局腔模。右上角给出单胞局域模式示意，右下角列出评价大面积单模器件时常用的光束质量指标。

如果只从二维光子晶体能带的角度看问题，很容易以为只要选中了 Gamma 点附近一支低损耗模，单模问题就已经基本解决。事实上，这最多只解决了第一道门槛：说明在无限周期、冷腔、小信号框架下，目标模有希望成为候选模。真正的器件级单模要求远高于此。目标模既要首先起振，还要对有限孔阵边界、局部增益起伏、温度漂移和偏振扰动具有足够鲁棒性，从而在工作窗口内持续压住其它模。

21.2 模鉴别要写成裕量

为了让后面的讨论有共同语言，可以把目标模 m_* 相对于竞争模 m_j 的优势写成净模裕量

$$\Delta G_{*j}(I, T) = [g_{\text{mod},*}(I, T) - \alpha_*(I, T)] - [g_{\text{mod},j}(I, T) - \alpha_j(I, T)], \quad (21.1)$$

其中 $g_{\text{mod},i}$ 是第 i 个模在给定工作点下的模增益， α_i 是其总损耗，包含辐射损耗、内部吸收、边缘损耗以及工作状态下可能出现的附加损耗。式 (21.1) 不是新的基本定律，它是一个组织设计问题的记账式写法。它的价值在于提醒我们：

第一，单模判据是相对的，不是绝对的。目标模本身阈值低还不够，还必须比竞争模更有优势。

第二，这个优势依赖工作点 (I, T) 。也就是说，单模描述的是工作窗口中的稳定区间，不能作为结构固有的单一标签。

第三，式 (21.1) 中的每一项可能来自不同物理模块。 $g_{\text{mod},i}$ 需要量子阱增益与注入分布， α_i 则既依赖开放光学，也依赖掺杂吸收、金属邻近损耗、边缘损耗和热漂移。

模增益的空间积分定义

模增益是空间积分量，定义为

$$g_{\text{mod},i} \approx \Gamma_{\text{opt},i} g_{\text{mat}}, \quad \Gamma_{\text{opt},i} = \frac{\int_{V_{\text{act}}} \varepsilon_r(\mathbf{r}) |\mathbf{E}_i(\mathbf{r})|^2 dV}{\int_{V_{\text{total}}} \varepsilon_r(\mathbf{r}) |\mathbf{E}_i(\mathbf{r})|^2 dV}, \quad (21.2)$$

其中第一式对应“有源区内材料增益近似均匀”的情形。若材料增益随局域载流子浓度与温度变化，即 $g_{\text{mat}} = g_{\text{mat}}(n(\mathbf{r}), T(\mathbf{r}))$ ，则不能提出积分号，必须逐点加权：

$$g_{\text{mod},i} = \frac{\int_{V_{\text{act}}} g_{\text{mat}}(n(\mathbf{r}), T(\mathbf{r})) \cdot \varepsilon_r(\mathbf{r}) |\mathbf{E}_i(\mathbf{r})|^2 dV}{\int_{V_{\text{total}}} \varepsilon_r(\mathbf{r}) |\mathbf{E}_i(\mathbf{r})|^2 dV}. \quad (21.3)$$

这是为什么电学 - 光学耦合求解对大面积器件至关重要的原因：只有自洽获得 $n(\mathbf{r})$ 和 $T(\mathbf{r})$ 的空间分布，才能准确计算 $g_{\text{mod},i}$ 。

因此，大面积单模设计需要在目标工作窗口内尽可能让

$$\Delta G_{*j}(I, T) > 0, \quad \forall j \neq *, \quad (21.4)$$

并且这个不等式对工艺误差和热漂移不至于过分敏感（红线 #2，见第2.10节）。

表21.1 把模裕量拆成了不同求解模块，图21.2 则把同一件事放回工作窗口中看：目标模需要在一片对电流、温升和工艺扰动仍然保持正裕量的区域内取胜，仅赢得某一个名义点并不够。

21.3 第一尺度：单胞与 Gamma 点物理决定的是候选资格

设计的第一尺度是单胞尺度。这里关心的问题是：在无限周期近似下，哪类 Gamma 点模具有较低被动辐射损耗，哪类近场对称性更容易给出期望偏振和远场，哪类对称性破缺或双

表 21.1: 模裕量各项与求解模块的交接关系。

裕量中的量	主要来源层级	更自然的求解接口
$g_{\text{mod},i}$	有源区、注入分布、温度回写后的模增益	量子阱增益模型 + 电学/热学回写 + 模场重叠
$\alpha_{\text{rad},i}$	开放辐射损耗	RCWA，或有限阵列 FDTD/FEM
$\alpha_{\text{int},i}$	自由载流子吸收、金属损耗、材料吸收	外延层与器件损耗模型
$\alpha_{\text{edge},i}$	有限孔阵边缘泄漏与侧向逃逸	有限阵列全波，或带包络的有限尺寸 CWT
热回写项	$\Delta T \rightarrow \Delta n, \Delta g, \Delta \alpha$	热-电-光耦合模块

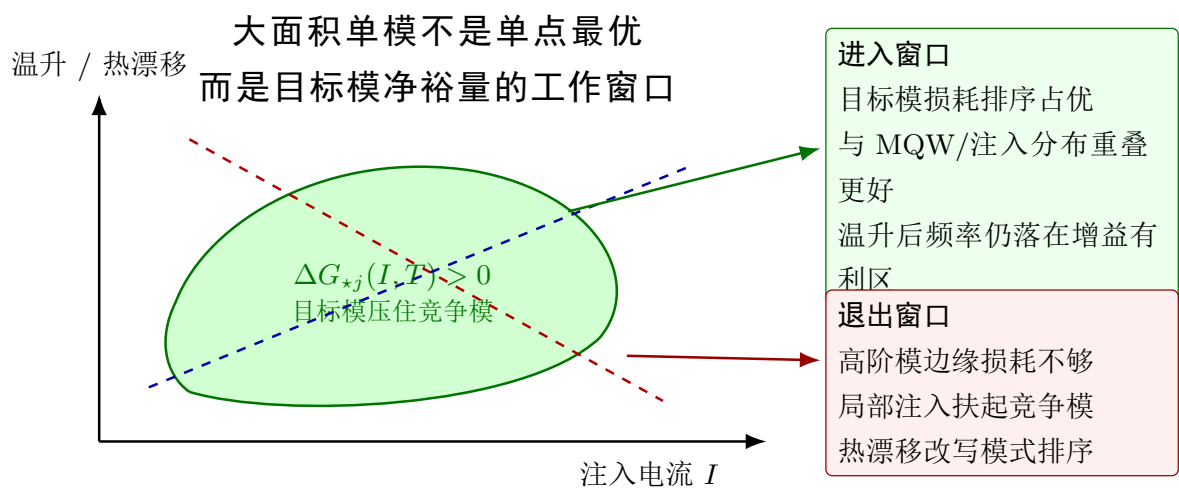


图 21.2: 大面积单模设计的净模裕量地图。绿色区域表示目标模相对竞争模保持正净增益裕量的工作窗口；边界可能来自热漂移、竞争模增益追上、边缘损耗不足或注入非均匀。

晶格扰动能把目标模与竞争模分开。这一层上的典型设计变量包括晶格常数 a 、孔半径 r 、占空比、刻蚀深度、双晶格位移、单胞形状微扰和上下层折射率对比等。

这一层可以回答的，是“目标模是否值得继续研究”。如果在单胞尺度上目标模就已经没有明显辐射优势，或者偏振选择完全不稳定，那么继续投入有限器件和电热建模通常是浪费时间。但必须认识到，单胞尺度只给出候选资格，不给出最终单模结论。因为单胞尺度默认了无限周期、相同局域环境和均匀增益代理，这些假设一旦进入大面积器件就都会受到挑战。

从求解器角度看，这一层最自然的工具是 第十三章、第十四章 中的 PWE、CWT 和 RCWA。如果用商业软件来承载这一层，Lumerical 中通常优先使用 MODE/FDE 做纵向等效折射率与波导模检查，再用 RCWA 或周期 FDTD 做单胞散射与辐射通道分析；COMSOL 中通常使用 Wave Optics 的频域周期单胞模型，通过 Floquet 型相位周期边界和端口或入射波设置读取反射、透射、衍射效率与局域场分布。

21.4 第二尺度：有限孔阵与边缘损耗决定的是横向可实现性

一旦走出无限周期近似，问题立刻变了。真实 PCSEL 只有有限个周期，边界会引入额外泄漏，目标模也会从严格 Bloch 模转化为带有横向包络的有限腔模式。此时设计的第二尺度就出现了：横向包络如何被组织，边缘处是增强还是抑制目标模，有限孔阵尺寸是否已经足以支撑目标模的相位一致性。

所谓边缘损耗，本质上就是原本在无限周期里被周期散射网络反复折返的能量，到了有限边界处不再被完整送回，而是沿侧向或斜向辐射通道逃逸出去。因此它是 Bragg 反馈网络被截断后的直接物理后果，不能当作外加的“附加罚项”。

边缘损耗的标度律

对一维有限周期结构，边缘损耗系数随周期数 N 呈指数衰减：

$$\alpha_{\text{edge}} \propto \exp(-N/N_{\text{char}}), \quad (21.5)$$

其中 N_{char} 是特征衰减长度，依赖于带隙深度和模式在禁带中的位置。对二维 PCSEL 阵列，若阵列尺寸为 $L \times L$ 且辐射衰减常数为 α_{rad} ，则有效边缘损耗满足 [25, 32]：

$$\alpha_{\text{edge}} \approx \alpha_0 \exp(-L/L_{\text{loc}}), \quad (21.6)$$

其中 L_{loc} 是模式的局域化长度，与带隙宽度成反比。对位于 Γ 点的模式， $L_{\text{loc}} \sim a/\sqrt{\Delta n/n}$ ，其中 a 是晶格常数， $\Delta n/n$ 是折射率对比度。这说明：

- 高折射率对比结构 ($\Delta n/n$ 大) 的边缘损耗衰减更快，所需周期数更少
- 深带隙模式比浅带隙模式更能抵抗边缘泄漏
- 对大面积器件 ($L \gg L_{\text{loc}}$)，边缘损耗不再是主导因素，内部损耗和辐射损耗成为主要考量

这一层最重要的直觉是，大面积并不自动有利于单模。面积增大确实可以减小衍射发散、提高功率潜力，但同时也会给许多横向模提供生存空间。为了让目标模在大面积下仍然占优，阵列放大必须与横向损耗分布设计同步进行。例如：

- 通过边缘区扰动增加竞争模的横向泄漏；
- 通过双晶格或对称性工程让目标模在近法向辐射上更占优势；
- 通过局部参数渐变或包络控制，让目标模横向分布与注入区域更匹配；
- 在保持单模性的前提下控制上下辐射和侧向泄漏的权衡。

这里应检查目标模与竞争模的损耗排序是否朝正确方向移动，而不能只追求“损耗越小越好”。有时为了压制竞争模，会故意让整体损耗略有增加；只要目标模的相对优势扩大，器件级单模性就可能改善。

这一层通常已经超出周期单胞 RCWA 的默认边界，需要转向有限尺寸全波仿真。Lumerical 中最常见的是 FDTD；COMSOL 中则更常见的是 Wave Optics 的三维频域或本征频率求解并辅以 PML。这里的输出应当包括有限阵列的近场、远场、复频率或等效 Q 、边缘敏感性，以及目标模与竞争模的排序是否随阵列大小改变。

21.5 第三尺度：注入与温升决定的是工作点下的真实胜负

即使前两个尺度都看起来很好，器件仍然可能在实际工作中丢掉单模性。原因在于第三尺度，也就是器件工作尺度。到了这里，模式竞争不再仅仅由被动辐射损耗决定，而是由载流子分布、局部模增益、串联电阻、Joule 热、非辐射复合热以及温升引起的折射率和增益谱漂移共同决定。

这也是第十九章、第二十章反复强调的地方。目标模光学上“形状直观”，并不意味着它在电流注入下得到的净增益最大；目标模冷腔频率“居中”，也不意味着温升后它仍然处在增益谱最有利的位置。大面积器件尤其容易出现这类反转，因为横向注入不均匀和热非均匀会随面积扩大而加剧，导致目标模原本依赖的对称性被工作状态主动破坏。

因此，真正的大面积单模设计必须把外延层光学、注入路径和热扩散一起作为设计变量。常见的协同思路包括：

- 让量子阱重叠最大的区域尽量与电流高密度区域对齐，而不是错位；
- 通过电极形状和限制层设计，减少边缘过强注入导致的竞争模局部起振；
- 通过层栈和散热路径设计，降低温升诱发的模排序翻转；
- 让目标模对小幅折射率漂移和增益峰位漂移具有更好的鲁棒性。

这部分若采用 Lumerical 工作流，通常需要 DEVICE 中的 CHARGE / HEAT 与光学求解器之间做脚本化数据交换；若采用 COMSOL 工作流，则更常见的是在 Wave Optics、Semiconductor 与 Heat Transfer 模块之间做统一几何下的多物理场耦合或分阶段迭代。两种方式的取舍属于工程权衡：Lumerical 的光学链往往更高效，COMSOL 的自定义耦合 PDE 链往往更灵活。

21.6 边缘损耗与表面提取共享同一预算

很多新手在谈单模设计时，会把“提高表面提取效率”和“抑制边缘泄漏”看成两个互不相干的目标。实际上，这二者共享同一份辐射预算。PCSEL 需要垂直辐射才有面发射输出，但又不希望这种辐射机制让竞争模也轻易起振；同时，器件还不希望能量在横向边缘无序泄漏，因为那会破坏单模相干面积。设计时应把有限的辐射耦合尽可能分配给目标模需要的垂直输出通道，并减少竞争模或边缘泄漏通道的份额；单独追求最小辐射损耗无法完成这种分配。

因此一些看起来“很高 Q ”的结构在大面积器件里并不一定表现最好。若过分压低所有辐射损耗，虽然被动 Q 可能很高，但目标模未必最容易提取，竞争模也可能被一并保护起来。相反，适度引入受控垂直辐射并选择性增加竞争模边缘泄漏，反而更有利于稳定单模工作。

例子

设两种方案 A 与 B 在单胞层面都能支持目标 Gamma 点模。方案 A 的被动 Q 更高，但有限阵列中竞争模的边缘泄漏也同样很小；方案 B 的目标模被动 Q 稍低，却能通过边缘扰动显著增大竞争模的侧向泄漏。在器件级上，方案 B 反而可能更容易保持大面积单模，因为它扩大的是目标模相对于竞争模的净裕量，而不是盲目提高所有模的寿命。

21.7 从 2023 年 Noda tutorial 抽出的放大设计规则

Noda 等人的 tutorial 清楚展示了 PCSEL 从 200 μm 级、500 μm 级推进到 3 mm 级时，哪些设计旋钮会失效、哪些旋钮必须升级；这比孤立的性能数字更适合作为本章主线 [50]。把这篇论文的路线压缩成本书工作流，可以得到三条规则。

第一，高对称单胞适合解释 Γ 点矢量光束和 BIC / quasi-BIC 物理，但不自动适合高亮度输出。圆孔等高对称结构会让若干带边模的法向辐射振幅相消，辐射常数过小；若目标是高斜率效率和单瓣远场，就必须通过三角孔、椭圆孔或双晶格等方式有控制地打破相消，而不是把“高 Q ”本身当作设计终点。

第二，双晶格通过两组晶格点之间约 $\lambda/(4n_{\text{eff}})$ 的相对位移重排 180° 反向散射；“单胞里孔更多”只是几何表象。这样可以削弱某些面内反馈通道，使基模横向包络变宽，同时让高阶横向模在有限孔阵边缘承担更大的相对泄漏损耗。对 500 μm 量级器件，这一机制已经足以支持 10 W 级脉冲单模和窄发散输出 [75, 50]。

第三，进入 3 mm 甚至更大口径后，单靠“增加高阶模边缘损耗”不够。此时基模和高阶模的面内损耗都可能趋近很小，模鉴别需要转向辐射常数随面内波矢偏离 Γ 点的曲率。Noda 等人把面内 $180^\circ/90^\circ$ 衍射视为 Hermitian 耦合，把经由辐射波的耦合视为非 Hermitian 耦合，并强调两者都要降低、且要保持相互平衡。若把 Hermitian 耦合 R 盲目压到接近零，虽然可能增大辐射损耗差的曲率，却会压低基模与高阶模的频率间隔，使载流子或热致折射率扰动更容易把它们重新耦合起来。稳定的大面积单模设计因此是“ R 与非 Hermitian 耦合 μ 同时足够小、彼此量级匹配，并且工作时的热折射率漂移已被预补偿或纳入验证”，而非“某个耦合越小越好”。

这三条规则正好落回式 (21.1)：高亮度 PCSEL 需要在目标电流和温度窗口内，同时保

持可提取的目标模辐射、足够大的竞争模净损耗、足够大的频率隔离，以及对热致折射率分布的鲁棒性。单独最小化 α_{rad} 、最大化 Q 或最大化某一个带边频率差，都不足以替代这组联合条件。

为了避免把这篇 tutorial 读成“高功率结果列表”，表21.2 把其中最重要的路线压缩成一个案例矩阵。阅读这张表时要注意，左列给出的是历史上出现的器件或设计阶段，右列给出的是本书真正希望读者带走的设计语言。

表 21.2: Noda 2023 tutorial 在本书中的案例化读法：从器件里程碑抽取大面积单模设计规则。

阶段	关键结构动作	tutorial 中的证据角色	本书中的设计解释
$\Gamma^{(2)}$ 点初始演示	三角晶格或方格晶格 支撑区中心带边共振	证明二维光子晶体可作为面发射 激光腔，并给出偏振与远场的对 称性证据 [43, 50]	单胞与能带只给候选资格，说明 二维相干反馈可形成，但尚未解 决高亮度输出窗口
非对称单胞	用三角孔、直角三角 孔等破坏高对称相消	瓦级单模与单瓣远场显示，辐射 通道可被孔形主动打开 [20]	BIC/quasi-BIC 语言必须服务于 可提取辐射；目标不是最高 Q ， 而是目标模辐射与竞争模损耗的 相对排序
双晶格	两组晶格点相对位移 约 $\lambda/(4n_{\text{eff}})$	500 μm 级器件获得 10 W 级脉冲 单模和窄发散输出 [75, 50]	通过重排 180° 衍射相位，让基模 包络展开，并把高阶模惩罚转移 到有限边缘损耗上
Hermitian/非 Hermitian 联合调 控	同时调节面内保守耦 合和辐射耗散耦合	3 mm 级连续波 50 W、极窄发散 和约 $1 \text{ GW cm}^{-2} \text{ sr}^{-1}$ 亮度成为 毫米级放大证据 [74]	大面积设计的主目标升级为复频 率排序、热工作窗口和模式裕量 共同稳定，而不是单独压低某一 项损耗

这张表也给出了本章后续仿真任务的优先级。若目标仍停留在“证明 Γ 点候选模存在”，PWE、RCWA 或周期单胞模型就已经能回答许多问题；若目标是复现双晶格的模式裕量，则必须转向有限尺寸 CWT 或有限阵列全波；若目标是判断 3 mm 级连续波工作是否稳定，则必须把热、电和增益空间分布接进式 (21.1)。同一篇 tutorial 因此在不同章节中承担不同角色：第 六 章读它的散射网络，第 十二 章读它的复耦合参数，本章读它的尺度放大规则，下一章则读它的性能报告边界。

21.8 把设计原则翻译成仿真任务：应该在哪类工具里算什么

到了这里，设计原则必须落回可计算任务。最实用的做法，是按“物理问题”而不是“软件名称”来分配求解器，然后再把这些求解器映射到 Lumerical 或 COMSOL 中。表21.3 给出的是一种实用的、但并非唯一的映射方式。

如果进一步问“简单建模过程是什么”，可以先给出最小可执行版本：

在 Lumerical 路径中，先用 MODE/FDE 检查纵向模与重叠，再用 RCWA 扫单胞参数和法向附近角谱，得到候选结构；随后把候选结构放入 FDTD 的有限阵列模型里计算近场、远场和边缘敏感性；最后在 CHARGE 中解注入分布、在 HEAT 中解温升，并把得到的 $N(\mathbf{r})$ 与 $T(\mathbf{r})$ 以代理方式回写到光学模型中做阈值和模竞争再评估。

表 21.3: 大面积单模设计中常见问题与 Lumerical / COMSOL 的典型实现映射。这里给出的是工程上常见分工，而不是唯一正确方案。

待判断性质	首选物理层级	Lumerical 中常见实现	COMSOL 中常见实现
纵向波导模、 n_{eff} 、约束因子	2D / 2.5D 光学	MODE/FDE 计算层栈模场、有效折射率与损耗	2D 横截面 Wave Optics 本征/模态分析
单胞辐射、偏振、上下出光比例	周期开放光学	RCWA 或周期 FDTD	Wave Optics 频域单胞，配合相位周期边界与端口/入射波
有限孔阵模场、远场、边缘敏感性	有限尺寸全波	FDTD，配合 PML 与近远场变换	3D Wave Optics 频域或本征频率研究，配合 PML
电流扩展、载流子分布、J-V	漂移-扩散电学	CHARGE	Semiconductor
温升、热源、热漂移	热传导	HEAT	Heat Transfer in Solids
热-电-光迭代后工作窗口	多物理场耦合	脚本串联 RCWA/FDTD 与 CHARGE/HEAT	Wave Optics、Semiconductor 与 Heat Transfer 分阶段耦合或统一模型

在 COMSOL 路径中，先在二维横截面或简化三维模型中做纵向模分析；随后在单胞三维频域模型中用相位周期边界和端口求出反射、透射与辐射分配；再切换到有限阵列三维模型，用 PML 包围开放区域求近场和远场；最后在同一几何下加入 Semiconductor 与 Heat Transfer 物理接口，解出电流密度、载流子密度和温度场，再将这些结果通过折射率和增益代理回写到光学接口中迭代。

21.9 真正的设计目标是工作窗口，而不是最佳单点

教材最后必须把一句话明确：大面积单模设计追求可工作的参数窗口，不能停在一个“最优参数点”。若某个结构只有在极窄的孔半径、公差和电流点上保持单模，它在研究原型中可能还能看到理想图样，但在真实器件中通常缺乏可复现性。相反，哪怕某个方案的理论最优值略差，只要它对工艺偏差、热漂移和注入不均匀有更强鲁棒性，它就更接近可用的设计。

这也是为什么本章始终围绕式 (21.1) 展开，而不是围绕某个单独的 Q 、某个单独的远场主瓣角或某个单独的阈值增益。对大面积 PCSEL 来说，单模性本质上是一个多约束交集问题，只有把这些约束压进同一工作窗口，结果才有器件意义。

易错点

“把被动 Q 最高的方案选出来，再交给电学和热学去补救”通常不是高效流程。真正高效的做法，是在单胞和有限阵列阶段就开始追踪目标模相对于竞争模的排序，而不是等器件级耦合完成后才发现一开始选错了候选结构。

21.10 2024–2026 设计导向补充：高功率、调制与系统链路

近年的大面积 PCSEL 路线正在把“单模 + 高亮度”与“可调制 + 可系统集成”同时作为目标：一方面，器件级高亮度与可扩展输出继续推进 [74]；另一方面，频率调制与长距离相干自由空间链路指标被直接引入设计目标函数 [26]。2024–2025 的另一条关键信号是“规模化与有限尺寸效应”被显式纳入证据链：横向耦合阵列路线给出了大面积相干扩展的器件证据 [16]，而有限尺寸研究直接指出了“无限周期结论向器件外推”的失效边界 [52]。因此，本章的“工作窗口”概念需要再向系统层扩展：不仅要有模裕量窗口，还要有调制窗口、热稳定窗口和链路容差窗口。对工程实现来说，这比追求单个最优点更关键。与全书统一前沿矩阵可对照表 1.6。

同时，面向超大面积器件的理论与动态研究也在补全本章“尺度跃迁”链条：全半导体大面积结构设计给出了毫米级尺度下的模式鉴别路径 [30]，而时域动力学仿真进一步展示了高泵浦条件下单模保持与高阶模抑制的条件窗口 [55]。这两类证据共同指向一个更严格的设计要求：大面积单模结论必须同时通过静态模式排序与动态稳定性两道验收。

报告规范

跨平台外推三元注释

- **外推来源：**来源于 PCSEL 直接证据（大面积耦合阵列、有限尺寸效应）并辅以系统应用文献 [16, 52, 26]
- **边界条件：**结论成立需同时满足：给定阵列尺寸、明确边缘截断与热边界、并给出模式排序与远场随尺寸的收敛证据
- **不可迁移项：**不可把无限周期单胞中的高 Q /高选择性直接迁移为器件级鲁棒单模结论；也不可把某一平台工艺窗口直接迁移到异材料体系

回看本章

大面积单模设计的核心，是在单胞、有限孔阵和器件工作点三个尺度上同时扩大目标模相对于竞争模的净裕量，而非追求一个孤立的低损耗目标模。单胞尺度决定候选资格，有限阵列尺度决定横向可实现性，器件尺度决定真实工作点下的胜负。Lumerical 与 COMSOL 都可以承载这条链，但各自更擅长的部分不同：前者通常更高效地承担 RCWA/FDTD/模式分析，后者通常更灵活地承担多物理场 PDE 耦合。无论使用哪条路径，最终目标都是稳定、可复现的工作窗口，而非最佳单点。

练习题

1. [基础] 结合式 (21.1) 解释为什么“目标模阈值低”并不自动等价于“器件单模稳定”。
2. [基础] 说明为什么大面积器件中的边缘工程应当被视为模鉴别设计，而不是几何收尾工

作。

3. [进阶] 选择一种你熟悉的结构扰动方式，分析它更可能改变的是单胞辐射排序、有限阵列边缘损耗，还是电热工作点下的净模增益排序。
4. [综合]（思路提示：先定层级，再列关键式）分别写出一条以 Lumerical 为主和一条以 COMSOL 为主的大面积单模验证路径，并指出各自最关键的接口变量是什么。

延伸阅读

- 下一章将进一步讨论名义值设计为何仍然不够，并系统引入有限尺寸、无序与工艺误差鲁棒性的问题，把“单模设计原则”升级成“可制造器件设计原则”。阅读时应同时核对本章变量在后续模型中的定义和适用范围。
- 若想继续深化模式竞争与单模稳定性，可结合 第二十二章、第二十五章 一起阅读，把本章的设计原则放入完整验证与容差分析工作流中。阅读时应同时核对本章变量在后续模型中的定义和适用范围。

第二十二章 有限尺寸、无序与工艺误差鲁棒性

读完本章，你能

- 理解为什么器件设计不能只优化名义值，以及“最优单点设计”和“鲁棒性设计”在目标函数上有何根本区别。
- 掌握孔半径无序、晶格常数漂移、刻蚀深度误差、侧壁倾角、表面粗糙、金属错位、电流孔径错位和阵列边缘截断如何进入 PCSEL 模型。
- 学会把几何误差、电学误差和热学误差统一组织成 tolerance analysis，并知道何时可用微扰观点，何时必须重做全波或多物理场计算。

读前准备 建议已掌握 第十五章和十九至二十一 中关于有限尺寸、注入非均匀性、热回写和
大面积单模设计的内容。

阅读主线 本章所处模型层级：L2→L4

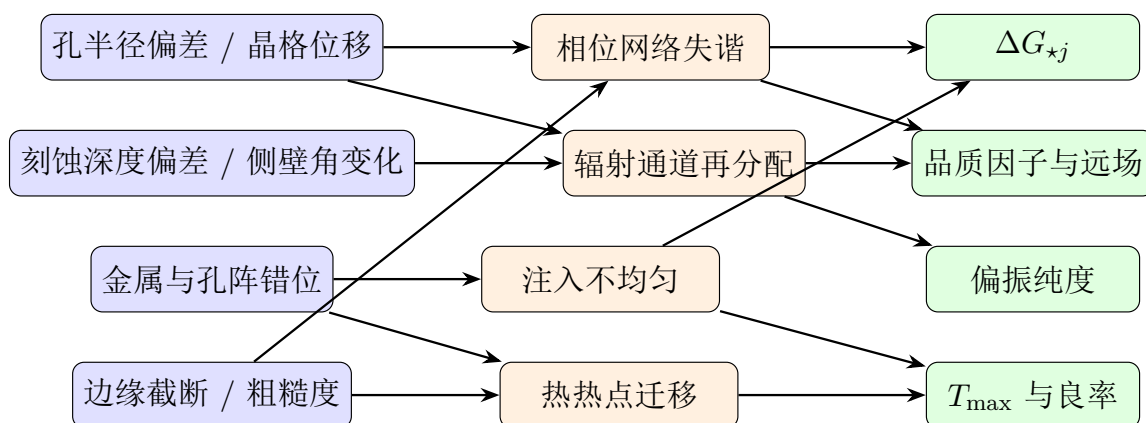


图 22.1: 容差分析中的误差传播图：工艺偏差先映射为物理机制变化，再传递到可观测器件指标。

先看 图22.1，可以把“工艺误差 → 物理机制 → 性能指标”的因果链一次看全，避免把本章读成分散清单。

上一章已经把“大面积单模设计”写成了三层问题：单胞候选资格、有限孔阵横向可实现性、以及器件工作点下的真实胜负。现在必须进一步承认一个更现实的事实：真实器件从来不是名义 CAD 几何。孔半径会偏、晶格常数会漂、刻蚀会浅或过深、侧壁不是理想竖直、金属

和电流孔径也未必和设计中心完全对齐。若书稿停留在名义值设计，那么它仍然更像一份研究草图，而不是一份器件教材。本章因此不把误差写成零散 checklist，而是从“误差如何进入模型”开始，随后讨论它们如何改变模式排序、辐射损耗、偏振、远场、注入均匀性和热分布，最后把这些物理变化组织成可执行的 tolerance analysis 与 robustness window 报告规范。

22.1 为什么名义值最优不等于器件最优

在很多纯光学优化问题里，设计者会先寻找一个名义参数点 $\mathbf{p}_0$ ，使某个性能指标在这个点上达到最好。例如，让被动辐射损耗最低、远场主瓣最窄、或者目标模与竞争模的频率分裂最大。这样的做法本身没有错，但它只回答了一个单点问题：若器件精确等于这一个参数点，它会怎样表现。真实器件设计要回答的是另一个问题：当参数在加工、公差、封装和工作过程中发生不可避免的偏移时，目标模式是否仍然赢得足够稳定。

于是，名义值最优和鲁棒性最优的目标函数天然不同。名义值优化追求

$$\max_{\mathbf{p}} F(\mathbf{p}), \quad (22.1)$$

而鲁棒性设计更接近追求

$$\max_{\mathbf{p}} \mathbb{E}[F(\mathbf{p} + \delta\mathbf{p})] \quad \text{或} \quad \max_{\mathbf{p}} \min_{\delta\mathbf{p} \in \mathcal{U}} F(\mathbf{p} + \delta\mathbf{p}), \quad (22.2)$$

其中 $\delta\mathbf{p}$ 表示工艺与装配引入的偏差， $\mathcal{U}$ 是容差集合。式 (22.1) 对应的是“找到理论最好点”；式 (22.2) 对应的是“找到在扰动下仍然可用的工作窗口”。二者可能重合，但在 PCSEL 中经常并不重合，因为目标模的优势常常建立在 Gamma 点简并分裂、有限阵列边缘控制、注入均匀性和热平衡这几个本来就敏感的机制之上。

直觉框

对 PCSEL 来说，最优单点设计关心的是“峰值有多高”，鲁棒性设计关心的是“平台有多宽”。器件导向研究更需要后者，因为加工出来的是一个有公差的结构，而不是一个数学点。

22.2 误差如何进入模型：从参数向量到性能向量

为了把各种误差放进同一框架，先把器件的名义参数写成

$$\mathbf{p} = (r, a, h_{\text{etch}}, \theta_{\text{sw}}, \sigma_{\text{rough}}, \Delta x_{\text{metal}}, \Delta x_{\text{ap}}, \mathbf{p}_{\text{edge}}, \dots), \quad (22.3)$$

其中 r 是孔半径， a 是晶格常数， h_{etch} 是刻蚀深度， θ_{sw} 是侧壁角， σ_{rough} 是表面粗糙度代理量， Δx_{metal} 表示金属图形相对孔阵的错位， Δx_{ap} 表示电流孔径相对目标模式中心的错位， $\mathbf{p}_{\text{edge}}$ 描述阵列边缘截断方式。

与之对应，把关心的器件输出组织成性能向量

$$\mathbf{y} = (\omega_m, \alpha_{\text{rad},m}, \Delta G_{*j}, \Pi, \mathcal{F}_{\text{FF}}, \eta_{\text{inj}}, T_{\text{max}}, \dots), \quad (22.4)$$

其中 ω_m 是模式频率, $\alpha_{\text{rad},m}$ 是辐射损耗, ΔG_{*j} 是上一章定义的目标模相对竞争模的净裕量, Π 可用来代表偏振纯度, $\mathcal{F}_{\text{FF}}$ 代表远场特征量, η_{inj} 表示注入效率, T_{max} 表示峰值温升。鲁棒性问题的核心不是盯住某一个标量, 而是研究

$$\delta \mathbf{p} \longrightarrow \delta \mathbf{y} \quad (22.5)$$

这条映射。

若扰动足够小、模式身份不变且开放损耗变化不太强烈, 可以先用一阶灵敏度近似

$$\delta y_i \approx \sum_j \left. \frac{\partial y_i}{\partial p_j} \right|_{\mathbf{p}_0} \delta p_j. \quad (22.6)$$

这一步的意义是先快速识别“谁最敏感”, 而非用线性模型替代真实器件。一旦出现模式排序改变、远场拓扑变化、注入路径重组或热热点迁移, 式 (22.6) 往往就不再够用, 必须回到全波或多物理场重算。

式 (22.6) 是工程上的局部近似。它成立的前提至少包括: 扰动幅度小、模式标签可连续追踪、目标量对参数是局部光滑的。对强辐射开放模、模式简并附近、以及电热反馈导致的阈值翻转问题, 这个近似必须谨慎使用。

22.3 几何无序的第一组: 孔半径无序与晶格常数漂移

孔半径无序 (hole radius disorder) 是最典型、也最容易在早期模型中被忽略的误差。对单胞来说, 孔半径直接改变介电常数的平面傅里叶分量, 因此会同时影响 Gamma 点附近耦合强度、频率位置和辐射通道强弱。若无序是全局一致的, 例如整片晶圆上孔半径系统性偏小, 那么它更像参数漂移; 若无序是从孔到孔随机变化, 那么它会打破原本依赖周期性的相干反馈, 并可能把原本受对称性保护的模混合起来。

对有限器件来说, 孔半径无序最先改变的通常是模式排序和远场纯净度, 而非“有没有模”。因为目标模与竞争模对单胞微扰的敏感方向并不相同, 有些模更依赖某个特定 Fourier 分量维持低辐射与高相干, 一旦这些分量在横向上随机起伏, 目标模的辐射损耗就可能增加得更快。于是本来在名义值设计中占优的模式, 可能在有限阵列里被别的模反超。

晶格常数变化 (lattice constant variation) 看起来只是比例尺变化, 实际上影响更深。全局的 a 漂移会整体移动归一化频率与辐射条件, 从而改变工作波长相对于增益峰位和 light cone 的位置; 局部的 $a(\mathbf{r})$ 起伏则会造成 Bloch 相位积累失配, 相当于在面内反馈网络中引入相位噪声。对大面积 PCSEL 来说, 这种相位失配尤其关键, 因为它会先破坏大面积相干, 再表现为远场主瓣变宽、旁瓣上升或偏振不稳定。

从建模层级看, 小幅全局 r 或 a 漂移通常可先用 PWE、CWT 或单胞 RCWA 做局部灵敏度; 但只要问题变成孔到孔随机无序, 或者设计本身工作在简并/近简并点附近, 就应尽快转向有限尺寸全波模型来检查模排序和远场是否保持。若同时还关心阈值电流变化, 则必须把这一步与后续的注入和热回写结合。

22.4 几何无序的第二组：刻蚀深度、侧壁倾角与表面粗糙

刻蚀深度误差 (etch depth error) 之所以危险, 是因为它不只是改变平面内散射强度, 还直接改变纵向层栈中的有效折射率剖面。换句话说, 它同时碰了 第 6 章、第 17 章 的两条主线: 既动了纵向模背景, 也动了面内耦合与垂直辐射机制。若设计恰好依赖浅刻蚀或部分刻蚀来平衡辐射与导模约束, 那么刻蚀深度误差对辐射损耗、上下辐射不对称和偏振都可能十分敏感。

侧壁角或锥度 (sidewall angle / taper) 进一步说明真实工艺不是理想二维截面。理想模型通常默认孔壁竖直, 而真实刻蚀常导致上宽下窄或上窄下宽的三维轮廓。这会改变孔洞在不同高度处的有效横向尺寸, 从而影响纵向不同场分量看到的等效散射强度。对近似 TE/TM 分类已经不够严格的 slab PCSEL 而言, 这种三维几何变化还可能增强矢量模混合, 使原本清晰的偏振选择变得更脆弱。

表面粗糙 (surface roughness) 则更像分布式弱散射源。若粗糙幅度较小、相关长度远小于模式横向尺度, 工程上常把它先当成附加散射损耗或附加吸收代理, 并研究其对 Q 、阈值增益和线宽趋势的影响。但这只是第一层近似。因为粗糙不只改变总损耗, 还可能打散原本规则的远场相位关系。对于大面积单模器件, 这种影响有时会首先表现为远场旁瓣抬升、偏振纯度下降和批次间重复性变差。

这一组误差的共同特点是: 它们都带有明显的三维性和开放辐射敏感性。因此, 若只是想筛查灵敏度大小规律, 可先在单胞 RCWA 或周期频域模型中扫 h_{etch} 与 θ_{sw} ; 若要判断远场、偏振和有限器件模排序是否真正稳健, 通常必须进入三维有限阵列 FDTD 或有限元频域模型。表面粗糙若只作为附加损耗代理使用, 应明确写出这是简化模型, 而不是对真实散射过程的完整求解。

设某设计名义上依赖浅刻蚀来实现“目标模足够辐射、竞争模仍受抑制”的平衡。若刻蚀实际偏浅, 目标模可能重新被过度束缚, 导致出光下降; 若刻蚀偏深, 竞争模和目标模都可能打开更多辐射通道, 净排序反而变差。这里最关键的是目标模相对于竞争模的损耗排序是否翻转, 而非总损耗单调上升还是下降。

22.5 装配与器件误差：金属错位、电流孔径错位和边缘截断

到了器件层面, 误差不再只发生在光子晶体孔阵本身。金属图形相对孔阵的错位 (metal misalignment) 会改变两个看似分离、实际上紧密耦合的问题。第一, 它改变金属邻近吸收和上下边界的局部光学条件; 第二, 它改变电流进入有源区的路径和局部串联电阻。于是, 同样的金属错位既可能改写辐射损耗, 也可能改写注入均匀性和温升分布。

电流孔径错位 (current aperture misalignment) 更直接地触碰 第 19 章 建立的器件主线。对大面积 PCSEL 来说, 目标模能否保持优势, 很大程度上取决于高注入区域是否与目标模的主要场强区域重合。若孔径中心偏离目标模中心, 即使冷腔模场本身较理想, 也可能因为局部模增益分布失配而让竞争模在边缘或局部热点处先获利。这正说明了为什么“光学模场分布合理”并不等于“器件注入均匀”。

阵列边缘截断方式 (array edge truncation) 则是有限尺寸问题最具体的落地形式。很多

名义值设计默认孔阵是一个完美的矩形或圆形窗口，但实际加工中边缘往往会因为版图裁切、刻蚀邻近效应和保护环设置而形成不同的终止单元。边缘是有限阵列泄漏、散射和模式包络反射的交汇处，因此截断方式不只是版图收尾，而是有限器件边界条件的一部分。不同边缘截断可能导致完全不同的侧向泄漏排序，进而改变大面积单模性。

这一组误差几乎总是要求比单胞更高层级的模型。金属错位和孔径错位若只想看局部一阶趋势，可以先用简化的横向注入代理或损耗代理做 screening；但若判断阈值电流、热点位置、模式竞争和远场稳定性，则必须进入电热耦合后的有限尺寸模型。边缘截断方式则天然属于有限器件全波问题，几乎不能仅靠无限周期模型回答。

22.6 误差为什么会改变模排序、辐射、偏振、远场、注入与热

到这里可以把各种误差背后的共同机制总结出来。第一类机制是**相位网络变化**。孔半径无序、晶格常数起伏和边缘截断会改变面内反馈网络的相位匹配，因此它们首先威胁的是模式排序和远场相干。第二类机制是**辐射通道重分配**。刻蚀深度、侧壁角和表面粗糙会改变目标模和竞争模向上、向下以及横向泄漏的比例，因此它们首先威胁的是辐射损耗、偏振和远场纯净度。第三类机制是**工作点重定位**。金属错位和电流孔径错位会改变载流子和温度的空间分布，因此它们首先威胁的是注入均匀性、热点位置和阈值电流。

这三类机制在真实器件里往往不是分开的。例如，金属错位先造成电流分布偏斜，进而产生温升不对称；温升不对称再通过折射率和增益谱漂移改变模式排序；模式排序变化又进一步改变近场和远场。也正因为这条链条存在，鲁棒性问题若只停留在冷腔名义值，就会系统性低估工艺误差的真实后果。

为了在设计中跟踪这些变化，一个实用写法是继续使用上一章的模裕量

$$\Delta G_{\star j}(\mathbf{p}) = [g_{\text{mod},\star}(\mathbf{p}) - \alpha_{\star}(\mathbf{p})] - [g_{\text{mod},j}(\mathbf{p}) - \alpha_j(\mathbf{p})]. \quad (22.7)$$

鲁棒性分析的核心不是只问 $\Delta G_{\star j}(\mathbf{p}_0)$ 是否为正，而是问在容差集合内

$$\Delta G_{\star j}(\mathbf{p}_0 + \delta \mathbf{p}) > 0 \quad (22.8)$$

是否仍对足够多的扰动样本成立。若答案是否定的，那么名义设计指标再高，也不该被称为鲁棒设计。

22.7 哪些误差可以先用微扰观点，哪些必须重算

这一节最容易被写成表格清单，但重要的是建立判断标准。能否先用微扰观点，不取决于误差名称本身，而取决于它是否满足三个条件：第一，扰动幅度小到不会改变拓扑和模式身份；第二，目标输出对参数局部光滑；第三，误差主要作用于已经分离好的单一物理模块，而不是同时重组光学、电学和热学链。

满足这些条件时，适合先做微扰或局部灵敏度分析的情形通常包括：小幅全局孔半径偏移、小幅全局晶格常数偏移、浅层刻蚀深度偏差、以及可被等效为附加损耗的弱粗糙。这些问

题可先用 PWE/CWT/RCWA 或局部频域模型给出敏感方向，再挑最敏感的样本进入全波验证。

一旦出现下列情形，就应直接升级到全波或多物理场重算：孔到孔随机无序导致模式局域化趋势；刻蚀深度与侧壁角同时变化并显著改变上下辐射；金属错位和孔径错位导致注入路径重组；边缘截断方式改变；温升和注入反馈已足以改变模式排序。在这些问题里，局部一阶导数往往不再是决定性信息，因为真实器件响应可能由模翻转、热点迁移或远场拓扑变化主导。

表 22.1: 常见误差的首选分析层级与最终验证层级。

误差类型	首先关注的物理量	初筛模型	最终验证模型
孔半径全局偏移	频率漂移、辐射排序、偏振趋势	PWE / CWT / 单胞 RCWA	有限阵列全波
孔到孔随机半径无序	模排序、远场纯度、局域散射	有限阵列简化扫描	有限阵列全波，多样本统计
晶格常数全局漂移	归一化频率、增益峰失配	PWE / 单胞频域	单胞开放模型加有限阵列抽查
刻蚀深度与侧壁角	辐射通道、上下不对称、偏振	单胞 RCWA / 频域 FEM	三维有限阵列全波
表面粗糙	附加散射损耗、远场退化	损耗代理或局部频域	需要时做校准后的全波抽查
金属错位	金属吸收、注入路径、热点	简化损耗/注入代理	电-热-光耦合有限器件模型
电流孔径错位	模增益分布、竞争模获益	注入代理 + 冷腔模场叠加	漂移-扩散 + 热回写 + 光学重算
阵列边缘截断	边缘泄漏、横向包络、远场	简化边缘参数扫	有限阵列全波

22.8 如何做 tolerance analysis: 不是盲扫，而是分层抽样

可用的 tolerance analysis 很少是“所有参数都从最小到最大全组合扫一遍”，因为 PCSEL 的参数维度高、每次全波或多物理场求解成本也高。更有效的流程通常分四步。

第一步，先做参数分层。把误差按物理来源分成几何、装配、电学和热学四组，并区分系统误差与随机误差。系统误差更适合做 corner analysis，随机误差更适合做 Monte Carlo 或拉丁超立方采样。

第二步，先做局部灵敏度筛查。通过式 (22.6) 或有限差分灵敏度，识别对 ΔG_{*j} 、远场主瓣、偏振纯度和峰值温升最敏感的少数参数。没有必要把所有低敏感度参数都立即带入最昂贵模型。

第三步，对高敏感参数做分层验证。若某参数主要影响单胞辐射，可先在 RCWA 或周期频域里扫；若某参数主要影响有限尺寸边缘泄漏，应直接在有限阵列全波里扫；若某参数影响注入和温升，则必须进入 CHARGE/HEAT 或 Semiconductor/Heat Transfer 一类模型。

第四步，才做联合统计。此时不是所有参数都等权，而是对关键参数赋予来自工艺的统计区间，构造样本集 $\{\mathbf{p}^{(s)}\}_{s=1}^{N_s}$ ，计算通过率

$$P_{\text{pass}} = \frac{1}{N_s} \sum_{s=1}^{N_s} \mathbf{1} \left[\Delta G_{*j}^{(s)} > 0, \Pi^{(s)} > \Pi_{\min}, \mathcal{F}_{\text{FF}}^{(s)} \in \mathcal{A}_{\text{FF}}, T_{\max}^{(s)} < T_{\max}^{\text{allow}} \right], \quad (22.9)$$

其中 $1[\cdot]$ 是指示函数。式 (22.9) 的意义很直接：是问“在一批可能加工出来的器件里，有多大比例仍满足设计要求”，而非问“名义器件好不好”。

若采用 Lumerical 路径，通常会用脚本串联 MODE/RCWA/FDTD 与 CHARGE/HEAT，对样本进行分阶段求解；若采用 COMSOL 路径，则更适合在统一几何参数化下做 parametric sweep，并在必要时把最贵的多物理场样本数量压缩到关键角点和代表性随机样本。

表 22.2: 器件导向 PCSEL 设计中可直接起用的工艺容差定量判据。它们是“超出此量级就必须高度警惕模式身份和器件窗口被改写”的经验起点，而非普适常数。

参数	起始容差窗口	首先监视的量	设计判据
晶格常数 a	$\pm 2\%$	频率偏移、增益峰失配	目标模排序不翻转，工作波长仍留在目标增益窗口内
孔半径 r	$\pm 5\%$	辐射损耗、偏振趋势、远场旁瓣	Q 与远场主瓣保持在允许窗口内
刻蚀深度 h_{etch}	$\pm 10 \text{ nm}$	上下辐射比、辐射损耗、偏振	目标模仍保持最小光学阈值或最大模裕量
波导层厚度	$\pm 5\%$	有效折射率、纵模位置、重叠因子	纵向模身份不翻转， Γ_{opt} 变化仍可接受
量子阱位置	$\pm 20 \text{ nm}$	Γ_{QW} 、阈值增益	阈值链变化不超过设计裕量

若希望把“鲁棒”写成更硬的工程语言，至少应给出一条推荐设计裕量：在名义设计点下，目标模相对首个竞争模的净裕量建议满足 $\Delta G_{*j} > 5 \text{ cm}^{-1}$ ；在容差边界或最坏样本下，仍应保持 $\Delta G_{*j} > 1 \text{ cm}^{-1}$ 。这两条门槛在本书中明确标记为 [本书设计规则]，用于教学与工程筛选，不应被当作跨平台统一标准。只要这条裕量在统计样本中频繁翻到零以下，就不应再把该结构称为鲁棒设计。

表 22.3: 鲁棒性报告时建议明确给出的量化判据。表中符号应由具体项目自行替换成明确门槛。

判据类别	推荐量化写法	用途
模身份保持	目标模在全部样本中与同一模族标签持续对应	防止把模式翻转误写成“参数敏感但可接受”
模裕量下界	$\min_s \Delta G_{*j}^{(s)} > 0$ 或给出最坏样本下界	防止只汇报均值而不说明最差情形
远场窗口	$\mathcal{F}_{\text{FF}}^{(s)} \in \mathcal{A}_{\text{FF}}$ 的通过率或最差偏离量	把远场要求从主观“看起来差不多”变成窗口判据
热安全边界	$\max_s T_{\text{max}}^{(s)} < T_{\text{max}}^{\text{allow}}$	防止只从光学视角宣布鲁棒
良率 / 通过率	P_{pass} 或角点最差样本通过情况	把“鲁棒”写成明确概率或最坏情形语言

22.9 如何报告 robustness window

鲁棒性报告最容易犯的错误，是只汇报“最优点”或者只给“某个参数可容忍 $\pm x$ ”而不说明其它参数是否同时固定。真正有用的 robustness window 至少应包含四层信息。

第一，名义设计点。必须给出 p_0 及其对应的名义性能。

第二，单参数窗口。说明在其它参数固定时，每个关键参数可接受的变化区间，例如孔半径、刻蚀深度、孔径错位等。

第三，联合窗口或通过率。说明当多个关键误差同时存在时，器件通过率是多少，或者最坏情况下的性能下界是多少。

第四，失效模式说明。明确器件首先因为什么失效，是模排序翻转、远场恶化、偏振崩坏、局部过热，还是注入均匀性丢失。没有失效模式说明的容差窗口，往往不利于后续工艺沟通。

在文风上，最推荐的报告方式是更具体地说：“在给定的孔半径、刻蚀深度与孔径错位统计区间下，目标模相对首个竞争模的净裕量仍保持为正，远场主瓣仍位于允许角窗内，峰值温升未超过设计上限，因此该方案在所定义的工艺窗口内通过率达到某一区间，而非说“该设计对工艺误差较鲁棒”。”如果没有统计样本支持，就不应使用“鲁棒”这个词。

易错点

只给名义参数点，不给通过率、不说首先在何种误差下失效，这样的“优化结果”通常只适合内部草图，不适合作为器件设计结论。鲁棒性设计的关键是把失效边界说明白，而非把均值做高。

22.10 前沿补充：quasi-BFIC 视角下的鲁棒性

近年的 quasi-BFIC (quasi-bound flat bands in the continuum) 研究提供了一个值得引入鲁棒性章节的新视角：是在更宽角谱内维持较高 Q 与可用耦合，而非只在单一动量点追求极高 Q 。对工艺容差的潜在收益是，当有限阵列和误差导致角谱展宽时，器件性能下降可能更平缓 [54, 33]。但这并不意味着 quasi-BFIC 自动“更优”：若目标是极窄主瓣和单点最大提取效率，传统 Γ 点 quasi-BIC 仍可能更合适。鲁棒性评估应把两类路线放在同一指标下比较：通过率、最坏样本模裕量、远场角窗保持率和热漂移后稳定性。更重要的是，quasi-BFIC 的鲁棒性判断必须与有限阵列证据联动，而不能停在无限周期层级 [52]。

报告规范

跨平台外推三元注释

- 外推来源：quasi-BFIC 机制文献 + 有限尺寸机制文献并用 [54, 52]
- 边界条件：必须给出有限阵列尺寸、边缘截断、热回写后模排序和远场角窗保持率
- 不可迁移项：不可把动量空间宽角高 Q 直接等价于器件级高通过率；不可忽略边界终止与注入非均匀带来的模式重排

常见错误：无限周期高 Q 结果的误用

错误理解：“无限周期模型显示宽角高 Q ，所以器件鲁棒性已经得到证明。” **正确做法：**必须在有限阵列与工艺误差样本上验证：主模排序是否保持、远场是否维持在角窗内、最坏样本净裕量是否仍为正。**结论边界：**无限周期结果只能证明“候选机制可行”，不能替代器件级容差统计。

报告规范**无限周期结果转有限阵列时的最小报告四要素**

1. **阵列尺寸：**至少报告 3 个横向尺寸样本，并明确主模/竞争模排序是否翻转。
2. **边界模型：**明确边缘截断方案、PML/端口设置与粗糙/无序代理模型。
3. **收敛曲线：**至少给出 1 组关键量（如 Q 、主瓣角、 ΔG_{*j} ）随尺寸或网格的收敛趋势。
4. **误差样本：**给出最坏样本或通过率统计，而不只给名义参数点。

22.11 把鲁棒性设计接回工作流与失败模式

现在可以把本章放回全书主线。第二十一章说明了为什么单模设计不能只盯名义值；本章进一步说明，即便名义值设计已经跨过单胞、有限阵列和器件工作点三层关口，仍然必须检查它对工艺误差是否稳健。第二十五章中的工作流，则应在“名义结构通过电热光验证”之后显式加入 tolerance analysis 阶段，而不是把它留到最后作为附加说明。第二十七章中的许多失败模式，其实正是在鲁棒性分析缺席时才暴露得最晚、代价也最高。

因此，器件导向的 PCSEL 设计流程更合理的顺序应是：先做名义值筛选，再做有限尺寸验证，再做器件级电热光回写，最后用 tolerance analysis 检查这个方案是否有足够宽的 robustness window。若一个方案只有较好的名义值，却没有工艺窗口，它更适合作为物理演示结构，而不适合作为器件候选。

回看本章

PCSEL 的鲁棒性问题是器件设计的核心组成部分，而非简单的“公差补充说明”。孔半径无序、晶格常数漂移、刻蚀深度误差、侧壁角、表面粗糙、金属错位、电流孔径错位和边缘截断，都会通过面内相位网络、辐射通道重分配以及注入与温升重定位三条路径改变模式排序、辐射损耗、偏振、远场、注入均匀性和热分布。小扰动、单模块、模式身份稳定的问题可先用微扰观点筛查；一旦涉及模式翻转、有限尺寸边界、注入路径重组或热反馈，就必须升级到全波或多物理场重算。真正的 tolerance analysis 追求的是在工艺扰动下仍有足够通过率的 robustness window，而非名义值最优。

练习题

1. [基础] 说明为什么孔半径无序与晶格常数随机起伏对大面积 PCSEL 的危害，往往首先表现为远场和模式排序问题，而不是单一频率漂移。
2. [基础] 针对刻蚀深度误差，分别说明单胞 RCWA、有限阵列 FDTD/FEM 和电热耦合模型各自能回答什么、不能回答什么。
3. [进阶] 给出一个你自己的 tolerance analysis 计划，至少包含参数分层、灵敏度筛查、联合抽样和通过率定义四步。
4. [综合]（思路提示：先定层级，再列关键式）解释为什么“金属错位”与“电流孔径错位”不能仅当作几何误差处理，而应看作光学、电学和热学的耦合误差。

延伸阅读

- 下一章将先把“名义设计是否成功”推进到起振以上的动态稳定性、调制与噪声问题，再在随后的性能章节中把这些动态结论翻译成线宽、相干性、亮度和光束质量等可观测指标。阅读时应同时核对本章变量在后续模型中的定义和适用范围。
- 建议与 第二十五章、第二十七章 对照阅读：前者告诉你 tolerance analysis 应放在工作流哪里，后者告诉你缺少鲁棒性分析时最容易出现哪些误判。阅读时应同时核对本章变量在后续模型中的定义和适用范围。

第二十三章 动态学、调制与噪声

本章关注 PCSEL 在高速调制下的动态行为及其在实际应用中的表现。我们将重点讨论两个核心问题：第一，小信号调制响应的物理机制和带宽限制因素；第二，相对强度噪声（RIN）等噪声动力学的特性和抑制思路。这些内容对于设计用于光通信、激光雷达和传感系统的 PCSEL 至关重要。

为了帮助读者直观理解 PCSEL 的小信号调制响应特性，图23.1展示了不同偏置电流条件下的示意响应曲线。图中三条曲线对应 $I/I_{th} = 1.5$ （蓝色）、2.5（红色）和 4.0（绿色）。这里的 GHz 数值与“最优偏置区间”应读作 [教学示意]，不是跨结构普适常数：弛豫振荡频率与带宽会随孔阵尺寸、寄生电容、热阻和封装条件显著变化 [10, 25]。更稳妥的结论是：偏置升高通常推动 f_R 上移并改变阻尼，但工程最优点必须在给定器件和热窗口下通过实测或标定模型确定。

读完本章，你能

- 从阈值条件出发建立 PCSEL 的最小动态学语言，理解载流子、光子数、增益压缩和模式竞争如何共同决定工作稳定性。
- 掌握弛豫振荡、小信号调制和线宽来源之间的逻辑链，而不是把它们割裂成不同专题。
- 明确哪些动态结论可以由确定性多物理场模型支撑，哪些结论必须进一步引入降阶噪声模型或实验闭环。

读前准备 建议已掌握 第七章、第十九章、第二十章和第二十二章 中关于阈值、有效模增益、电注入分布和热反馈的内容。

阅读主线 本章所处模型层级：L4

前面的章节已经把 PCSEL 的静态链条搭了起来：模式、阈值增益、阈值电流来源以及热-电-光工作窗口都已经出现。接下来必须补上的，是“过了阈值以后会怎样”。本章先从单模速率方程出发，建立最小动态学骨架；再把增益压缩、模竞争和热时间尺度纳入，说明为什么真实器件的稳定工作从来不是一句“高于阈值即可”这么简单；随后讨论小信号调制和弛豫振荡；最后把相位噪声、线宽增强因子与线宽来源串起来，明确哪些量可以从建模中直接得到，哪些量还需要实验和降阶噪声模型闭环。

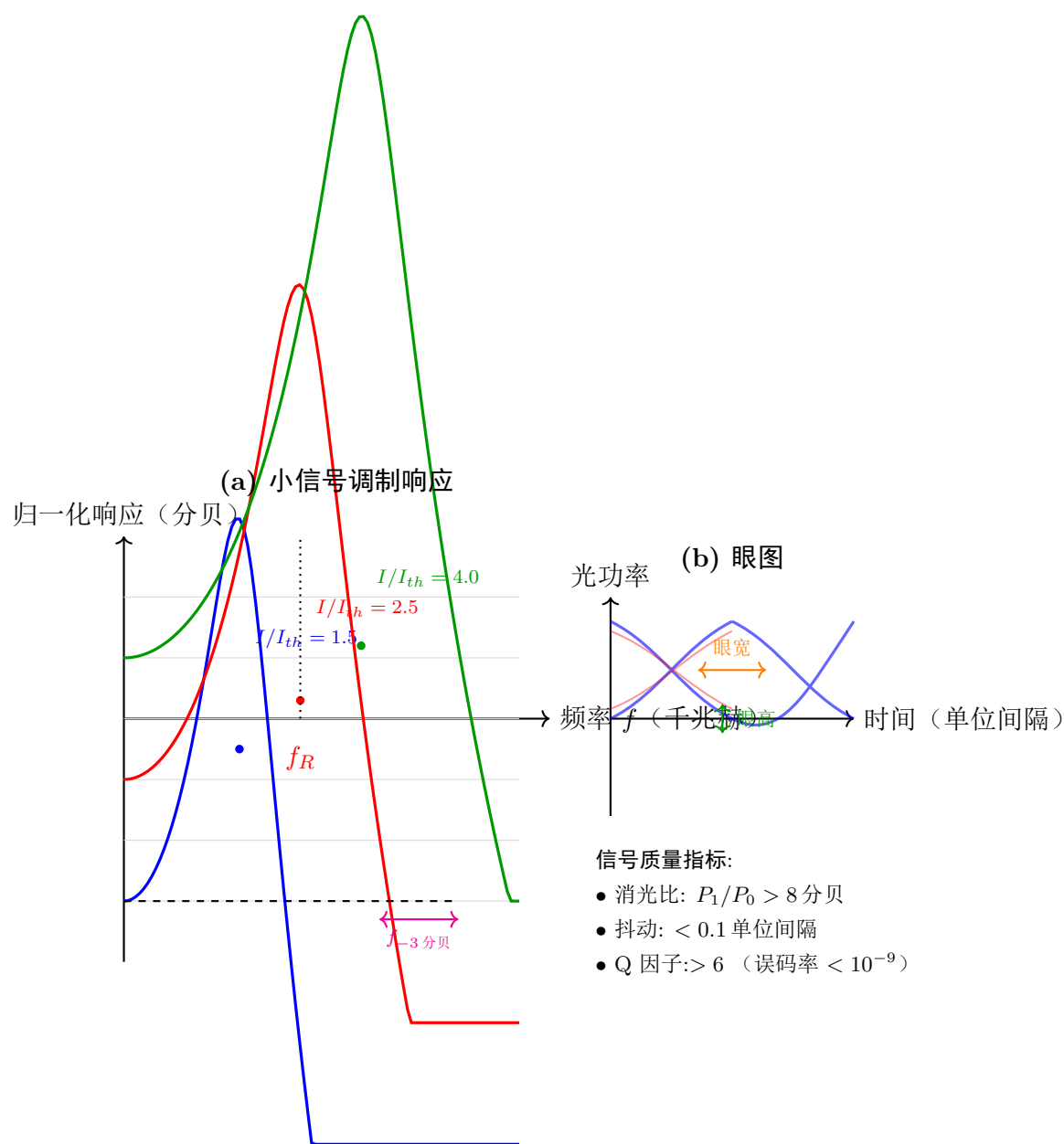


图 23.1: 光子晶体面发射激光器动态调制示意: 左图小信号频率响应随偏置电流升高而展宽, 右图高速数字调制下的眼图与主要质量指标。

23.1 为什么“跨过阈值”只是动态学的起点

第7章给出的阈值条件，本质上是线性起振判据：目标模的净增益刚好补偿总损耗。但激光器一旦进入起振以上区域，问题马上变成一个动态平衡问题。载流子继续被注入，光子数增长，受激辐射开始反过来抽空载流子，温升又在更慢时间尺度上改变增益谱和损耗分布。于是，“器件能亮”并不等于“器件稳定、单模、低噪声”。

对 PCSEL 来说，这个提醒尤其必要。因为它往往具有较大横向面积、多个接近的候选模、可工程化的偏振支路以及显著的热-电-光回写。只要其中任一支路在工作过程中失去净优势，器件表现就可能从“窄远场、单偏振、相对稳定”转为“远场变宽、偏振波动、模式跳变明显”。因此，本章讨论动态学，是为了说明为什么 PCSEL 的器件工作窗口必须用动态语言来定义。

直觉框

阈值只回答“哪一支模最先有资格起振”，动态学回答的则是“起振以后这支模能不能继续站住、能站多稳、对外界看起来有多干净”。

23.2 最小单模速率方程：把阈值语言延伸到起振以上

在最小单模、等效均匀有源区的近似下，可用载流子密度 N 与光子密度 S 写出一组真正可计算的最小速率方程

$$\frac{dN}{dt} = \frac{\eta_{\text{inj}} I}{qV_{\text{act}}} - R(N, T) - \Gamma v_g g(N, T) S, \quad (23.1)$$

$$\frac{dS}{dt} = \left[\Gamma v_g g(N, T) - \frac{1}{\tau_p} \right] S + \Gamma \beta_{\text{sp}} R_{\text{sp}}(N, T). \quad (23.2)$$

其中常把材料增益进一步近似为

$$g(N, T, S) \approx \frac{g_0(N - N_{\text{tr}})}{1 + \varepsilon_g S}, \quad (23.3)$$

这里 $R(N, T)$ 是总复合率， Γ 是目标模约束因子， v_g 是群速度， τ_p 是光子寿命， β_{sp} 是自发辐射耦合因子， ε_g 是增益压缩系数。对量子阱半导体激光器， ε_g 常见于 10^{-17} – 10^{-16} cm^3 量级；但在 PCSEL 里它应被视为平台与工作点相关的拟合参数，需绑定波段、孔阵尺度、寄生电容与热边界联合标定 [10, 25, 26]。因此本书仅将该区间作为 [工程经验] 敏感度扫描初值，而不作为“PCSEL 固定常数”。只要把稳态条件 $dN/dt = dS/dt = 0$ 代入，式 (23.2) 就会回到第7章的阈值形式；但一旦允许时间变化，它们便开始描述开机瞬态、工作点回稳以及小扰动响应。

式 (23.1) 和式 (23.2) 中的 v_g 必须读成速率方程中的等效光子-增益转换参数，而不能机械理解成“PWEM 带图在 Γ 点的 Bloch 群速度”。对 PCSEL 的 standing-wave / Γ 点模式来说，带图上的群速度趋零与速率方程里仍可出现一个有限的等效增益转换系数，并不矛盾，因为两者服务的是不同层级的降阶描述。

这个模型当然还很粗糙。它把空间分布压缩成了单个 N 和单个 S ，因此不能直接处理横向多模竞争和空间烧孔；它的适用前提——目标模已知、主导时间尺度相对分离、空间非均匀

性可被等效参数吸收——一旦被突破,就必须升级到多模或空间分辨模型。但它有重要教学价值:它让读者第一次看到阈值以上的闭环机制。注入提高 N , 增益提高; 增益提高后 S 上升; S 上升后又通过受激辐射消耗 N , 最终形成新的稳态。

23.3 增益压缩与多模竞争: 为什么工作点不会无限平滑

若只看式 (23.2), 似乎起振以上只要持续加电流, 目标模的光子数就会平稳增长。真实情况并非如此, 因为增益会饱和, 竞争模会共享同一载流子库, 局部热斑还会改写模式排序。式 (23.3) 是把谱烧孔、载流子加热和多体效应压缩成个工作点代理参数, 而非第一性原理定律。

若器件存在两个接近竞争的模式 $m = 1, 2$, 更实用的写法是

$$\frac{dS_m}{dt} = \left[\Gamma_m v_{g,m} g_m(N, T, \{S_n\}) - \frac{1}{\tau_{p,m}} \right] S_m + \beta_m R_{sp,m}. \quad (23.4)$$

一旦 S_1 增长, 它不仅会抽空共享的载流子库, 还会通过增益压缩削弱 S_2 的有效增益; 反过来, 若热漂移或注入非均匀使 g_2 上升更快, 竞争模也可能在较高电流下重新占据优势。这是大面积 PCSEL 中远场突然变宽、偏振切换和副瓣抬升的动态来源之一。

若某器件在冷腔分析中目标模相对竞争模只有很小的阈值增益裕量, 那么一旦工作电流上升, 局部温升导致增益峰偏离、注入分布中心外移, 竞争模就可能在起振后重新获得更高的有效增益。这时器件并非“阈值判断错了”, 而是进入了动态工作点改写模式排序的区域。

23.4 弛豫振荡与小信号调制: 从稳态到频率响应

把式 (23.1) 和式 (23.2) 在稳态点 (N_0, S_0) 附近线性化, 可得到一个 2×2 小信号系统。对其中的特征根进行分析, 便会出现半导体激光器最经典的动态特征之一: 弛豫振荡。若记小扰动为 $\delta N, \delta S$, 则常可得到一个形如

$$\delta \ddot{S} + 2\zeta \omega_R \delta \dot{S} + \omega_R^2 \delta S = K \delta I \quad (23.5)$$

的二阶响应式, 其中 ω_R 是弛豫振荡频率, ζ 是阻尼因子。定性上说, 注入把载流子往上推, 受激辐射又把它拉下来, 载流子和光子库之间便形成了一个耦合振荡对。

为了让读者能自己把这条式子推一遍, 最小线性化步骤可以写成下面这样。先令

$$N(t) = N_0 + \delta N(t), \quad S(t) = S_0 + \delta S(t),$$

并在工作点附近把材料增益近似展开为

$$g(N, T_0) \approx g_0 + g_N \delta N, \quad g_N = \left. \frac{\partial g}{\partial N} \right|_{(N_0, T_0)}. \quad (23.6)$$

再把稳态条件

$$\Gamma v_g g_0 = \frac{1}{\tau_p}$$

代入式 (23.1) 和式 (23.2)，并定义

$$\frac{1}{\tau_N} := \left. \frac{\partial R}{\partial N} \right|_{(N_0, T_0)}, \quad (23.7)$$

就得到最小线性系统

$$\delta \dot{N} = - \left(\frac{1}{\tau_N} + \Gamma v_g g_N S_0 \right) \delta N - \Gamma v_g g_0 \delta S + \frac{\eta_{\text{inj}}}{q V_{\text{act}}} \delta I, \quad (23.8)$$

$$\delta \dot{S} = \Gamma v_g g_N S_0 \delta N. \quad (23.9)$$

消去 δN 之后，就得到

$$\delta \ddot{S} + \left(\frac{1}{\tau_N} + \Gamma v_g g_N S_0 \right) \delta \dot{S} + \frac{\Gamma v_g g_N S_0}{\tau_p} \delta S = \Gamma v_g g_N S_0 \frac{\eta_{\text{inj}}}{q V_{\text{act}}} \delta I, \quad (23.10)$$

它正是式 (23.5) 的具体来源。由式 (23.10) 立刻可读出

$$\omega_R^2 \approx \frac{\Gamma v_g g_N S_0}{\tau_p}, \quad 2\zeta\omega_R \approx \frac{1}{\tau_N} + \Gamma v_g g_N S_0. \quad (23.11)$$

这条推导链的教学意义在于：弛豫振荡是载流子库和光子库围绕稳态点的线性耦合本征响应，而非凭经验命名的峰。

若把式 (23.3) 中的增益压缩保留到同一阶，等效微分增益会先被压低为

$$g_N^{\text{eff}} \approx \frac{g_N}{1 + \varepsilon_g S_0}, \quad (23.12)$$

而阻尼项则会额外获得与 $\varepsilon_g S_0 / \tau_p$ 同量级的正贡献。也就是说，随着功率上升， ε_g 一边降低有效微分增益，一边提高阻尼，最终使调制峰被压钝、带宽不再能靠继续加功率线性提升。这给出了高功率下不能无限快调制的物理解释。

这一步在 PCSEL 中的意义，不只是教读者认识一个标准术语，而是帮助判断器件到底更像“能稳定工作的大面积单模器件”，还是“在注入和热反馈下很容易触发模式起伏的边缘设计”。因为如果目标模与竞争模距离太近、增益压缩太强、热时间尺度又开始进入测量窗口，那么单一弛豫峰的理想小信号图像就会被明显扭曲。

对于小信号调制，最重要的是知道应当报告什么，而非死记带宽公式。合理的报告方式通常至少包括：偏置电流、温度、主导偏振、是否为单模区间、测得或估计的弛豫频率、阻尼情况以及是否出现模式竞争导致的额外滚降。否则，“器件调制带宽是多少”这种说法很容易失去上下文。若暂时忽略强阻尼、额外热极点和显著模式竞争，一个常见的教学近似是

$$f_{3\text{dB}} \approx \sqrt{3} f_R, \quad f_R = \frac{\omega_R}{2\pi}, \quad (23.13)$$

它的用途只是给出“单模小信号极限下的数量级”，而不是替代真实器件在复杂工作点下的带宽测量。

把教学带宽公式与真实 FM-PCSEL 对齐

Inoue 等人的双区 PCSEL 提供了一个带完整工作点的现实锚点：器件出光口径 500 μm 、激光波长 942 nm，在 3 A 偏置下输出超过 1 W；时变三维 CWT 预测反相电流驱动可把幅度调制压低，并把明显的频率调制响应保持到约 3 GHz，该上限对应其弛豫频率 [26]。实验在 1.1 W 平均功率下实现了 0.5 Gbaud 与 1.0 Gbaud 相干自由空间通信，对应链路预算分别为 88 dB 与 81 dB。

这些数字不能直接填进 $f_{3\text{dB}} \approx \sqrt{3}f_R$ ：3 GHz 是给定模型下的 FM/AM 响应边界，Gbaud 是系统传输速率，二者都不是未经协议说明的“强度调制 3-dB 带宽”。真实带宽必须连同偏置、功率、调制方式、探测器带宽和误码判据一起报告。

23.5 热时间尺度为什么会把快动态问题改写成慢漂移问题

如果只讨论载流子和光子，动态问题常集中在纳秒到皮秒尺度；但真实电注入 PCSEL 还有更慢的热时间尺度。温升不会像光子数那样瞬时响应，却会在微秒到毫秒甚至更长时间内逐步改写折射率、增益峰和吸收。于是很多实验上观测到的“线宽变宽”“偏振开始晃”“远场主瓣慢慢偏”并非快噪声，而是慢热漂移叠加在快动态之上。

若不先把快动态与慢漂移分开，读者就无法判断某个不稳定现象究竟来自弛豫振荡过冲、模式竞争，还是热回写造成的工作点漂移。

易错点

把所有时间变化都笼统称为“噪声”，会直接破坏模型判断。快动态、慢热漂移和外界扰动属于不同时间尺度、不同建模层级的问题，必须分开识别。

23.6 从相位噪声到线宽：Henry 因子为什么出现在这里

线宽之所以必须放在动态学之后讲，是因为它的来源并非“模式有一个 Q 就够了”，而是相位噪声如何在起振以上持续积累。对半导体激光器来说，一个核心机制是幅相耦合，即载流子扰动不仅改变增益，也改变折射率；于是强度噪声会被转写成相位噪声。常用的线宽增强因子写为

$$\alpha_H = -\frac{4\pi}{\lambda} \frac{\partial n / \partial N}{\partial g / \partial N}, \quad (23.14)$$

它衡量载流子波动引起折射率变化与增益变化之间的耦合强度。若再把它与输出功率、光子寿命和自发辐射耦合联系起来，便能得到 Henry 型线宽估计式。

在最小单模、近稳态近似下，常把这一层关系进一步写成

$$\Delta\nu_H \approx \frac{1 + \alpha_H^2}{4\pi} \frac{n_{\text{sp}}}{\tau_p N_{\text{ph},0}}, \quad (23.15)$$

其中 n_{sp} 是等效自发辐射因子， $N_{\text{ph},0}$ 是稳态腔内光子数（若前文使用光子密度记号，则 $N_{\text{ph},0} = S_0 V_{\text{ph}}$ ）。式 (23.15) 的用途，是把“载流子引起的折射率起伏会放大线宽”这件事压缩成一个

可估量的工作点公式；它的边界则同样必须一起写出：只有当器件近似单模、工作点近稳态、相位扩散可视为小扰动时，它才是可靠的第一近似。

本书在这里强调它的作用，是为了说明线宽为什么天然依赖**工作点、光子寿命、材料增益导数和热环境**。因此，后面的性能章节在讨论线宽时，只把确定性全波和器件模型视为中间量提供者，而不把它们误当成“一次仿真直接给出线宽”的完整线宽模型；若器件进入强热漂移、偏振切换、模式跳变或显著多模竞争区域， α_H 也不再是足以封装全部相位噪声的单参数。

23.7 RIN、模式分配噪声与快慢噪声分层

前面已经把线宽的相位侧讲清，但对实验上同样常见的**相对强度噪声**（relative intensity noise, RIN），本章还需要补一块最小框架。若输出功率记为 $P(t) = P_0 + \delta P(t)$ ，则常见的单边功率谱定义写成

$$\text{RIN}(f) = \frac{S_{\delta P}(f)}{P_0^2}, \quad (23.16)$$

其中 $S_{\delta P}(f)$ 是功率起伏的谱密度。这个定义的意义很直接：它把“强度抖得有多厉害”写成频率分辨的量，而不是只给一个总均方起伏。

对 PCSEL 而言，RIN 通常混有三层来源。第一层是载流子-光子库自身的快随机起伏，它和弛豫振荡峰往往联系最紧。第二层是模式竞争和模式分配噪声（mode partition noise）：不同偏振支路或不同横向模式在共享载流子库时，会把输出功率在多个通道间重新分配。第三层则是慢热漂移和外界供电/机械扰动，它们在低频端把 RIN 谱抬高，却不应被误判成“腔本身的快噪声变大”。因此，**RIN 与线宽一样，都必须按频率窗分层讨论。**

若某器件在高偏置下主模功率仍然上升，但低频 RIN 明显抬高，同时远场开始轻微摆动，那么比“腔品质变差”更合理的解释通常是：热慢漂移或弱模式竞争开始出现，而非单纯的快自发辐射噪声增强。

23.8 外腔反馈：为什么即使本书不展开，也必须给读者一个入口

很多 PCSEL 在实际测试和封装中都会遇到外腔反馈：窗口片、探测光路、准直镜甚至测试台上的反射面，都可能把一小部分光再次送回器件。即使反馈很弱，它也可能在动态学上引入新的慢快耦合。对教材而言，最小模型并不需要完整展开 Lang-Kobayashi 方程，但至少要让读者知道：**外腔反馈是一种独立于器件内部电热回写的动态扰动入口。**

在最小层面上，它会带来三类后果。第一，改变有效相位条件，使频率噪声谱和线宽对外部反射相位敏感。第二，改变强度噪声低频端，使 RIN 与偏振/模式稳定性一起恶化。第三，在更强反馈下诱发模式跳变或准周期起伏，使“单模近稳态”这一整套降阶假设本身失效。也正因此，本书把外腔反馈定位为**动态与实验接口章节的边界条件**，而不是把它塞回冷腔理论里当作一个小修正。

23.9 自发辐射耦合、Purcell 与 β_{sp} 因子：哪些微腔语言可用，哪些不能滥用

前面速率方程里已经出现过 $\beta_{\text{sp}} R_{\text{sp}}$ 项，但如果不把它和 Purcell 因子、自发辐射耦合以及 PCSEL 的开放大面积特征放在一起解释，读者很容易把整套微腔语言机械照搬过来。更稳妥的最小定义是

$$\beta_{\text{sp}} = \frac{R_{\text{sp} \rightarrow \text{mode}}}{R_{\text{sp}, \text{tot}}}, \quad (23.17)$$

即 β_{sp} 表示总自发辐射中有多少比例耦入了目标激光模。它衡量的是自发辐射分流，不是斜率效率，也不是出光提取效率。下文若无歧义，会口头简称“ β 因子”，但记号始终保持 β_{sp} 。

若进一步采用经典微腔近似，常见的 Purcell 因子写成

$$F_{\text{P}} \sim \frac{3}{4\pi^2} \left(\frac{\lambda}{n} \right)^3 \frac{Q}{V}, \quad (23.18)$$

其中 Q 是品质因子， V 是模体积。式 (23.18) 在教学上很有用，因为它建立了“局域光学态密度可被腔改写”的直觉；但它的适用前提也必须一起写出来：单模主导、弱耦合、发射体与腔模良好共振、模体积有清楚定义，且开放辐射通道没有把问题变成强多通道干涉。

对许多大面积、开放、Bloch-like 的 PCSEL 而言，这些前提并不天然成立。目标模往往延展在较大横向面积上，模体积不再像纳米腔那样是一个简单局域量；同时，法向辐射、侧向泄露、竞争模与有限尺寸边缘都会共同参与态密度重分配。因此：

- 把大面积 PCSEL 的高 Q 直接翻译成“强 Purcell 增强”，通常是不严谨的；
- 把 β_{sp} 因子直接当成“阈值高低的主控参数”，对多数器件导向 PCSEL 也往往是误判；
- 更合理的做法是把 β_{sp} 与 Purcell 视为噪声、起振附近自发辐射馈入、以及局域态密度工程的辅助语言，而不是替代增益、损耗、注入和热反馈的主语言。

什么时候可以先忽略 β_{sp} 若目标是分析电注入大面积 PCSEL 的阈值排序、模式竞争、调制趋势和热稳定窗口，而且已有证据表明起振后很快进入受激辐射主导区，那么把 β_{sp} 作为小参数、只保留它对起振附近噪声地板的影响，通常是合理的一阶近似。这也是为什么本书前面的阈值和器件章节没有把 β_{sp} 放到主线中央。

什么时候 Purcell / β_{sp} 不能被轻描淡写 若器件进入以下场景，就不应再把它们简单略去：目标模极高 Q 且空间局域性增强、工作点长期停留在阈值附近、需要解释自发辐射馈入主导的起振统计、需要比较不同模式对局域态密度的重分配，或者设计对象本身已更接近纳米腔/缺陷腔而不是典型大面积 PCSEL。此时 Purcell 与 β_{sp} 语言能够帮助解释“为什么起振附近的噪声、线宽起步行为或模式占优概率发生改变”，但仍然不应跳过注入、热和开放辐射多通道这些更大的器件背景。

“这个 PCSEL 的被动 Q 很高，所以 Purcell 因子一定很大，因此阈值一定很低”这类推理是典型的微腔语言滥用。对大面积开放 PCSEL，阈值首先仍由模增益、辐射与内部损耗、注入分布和热回写共同决定；Purcell 与 β_{sp} 最多是在某些工作区间帮助解释自发辐射馈入和噪声，而不是代替整条器件链。

23.10 哪些动态结论可以直接建模，哪些必须实验闭环

到这里必须明确区分三层可获得性。

第一层，是确定性模型可以较直接给出的量。例如稳态工作点、光子寿命、载流子分布、热时间常数趋势、模式间阈值裕量以及对偏置点的灵敏度。这些量可以由 RCWA / FDTD / FEM 与 CHARGE / Semiconductor / HEAT / Heat Transfer 的组合给出。

第二层，是可以由这些中间量进入降阶动态模型后较可靠估计的量。例如弛豫振荡频率随偏置上升的趋势、是否存在明显多模竞争风险、工作点附近的调制滚降来源以及线宽的量纲级变化。

第三层，是必须实验闭环才能真正定量的量。例如最终工作线宽、封装和外腔反射参与后的频率噪声谱、长时间漂移窗口内的偏振保持性以及高功率连续波下的完整动态稳定区间。这些量若脱离实验，仅靠确定性仿真难以严肃给出。

确定性多物理场求解器最擅长的是给出动态与噪声模型所需的中间量，而不是替代随机过程模型本身。若没有说明噪声源、统计窗口和实验边界条件，就不应把“动态趋势分析”写成“完整噪声预测”。

回看本章

PCSEL 的动态学从阈值条件自然延伸而来，但绝不能停留在“起振以上功率继续增加”这种静态印象上。最小速率方程把注入、复合、受激辐射和光子寿命组织成了起振以上的基础闭环；增益压缩、多模竞争和热回写则说明真实大面积器件的稳定工作窗口远比单模理想模型复杂。弛豫振荡和小信号调制告诉我们器件对快扰动怎样响应，而 Henry 因子与相位噪声则把线宽问题重新连回材料与工作点。自发辐射耦合、Purcell 因子与 β 因子则进一步提醒我们：微腔语言在 PCSEL 中并非全无用处，但必须放在“大面积开放腔、注入与热反馈主导”的背景下谨慎使用。到本章结束时，读者应当已经具备一条连续逻辑：阈值只是起点，动态稳定性决定工作窗口，噪声和线宽则是这条动态链在实验观测端的投影。

练习题

1. [基础] 解释为什么阈值条件成立并不自动意味着器件在起振以上会稳定单模工作。
2. [基础] 结合式 (23.4) 说明模式竞争如何同时改变远场、偏振和线宽趋势。
3. [进阶] 讨论为什么热时间尺度的引入会让“动态不稳定”和“慢漂移”在实验上表现得很像，但建模上必须区分。
4. [进阶] 说明式 (23.14) 为什么会把材料层面的折射率导数和增益导数直接带进线宽讨论。
5. [综合] (思路提示：先定层级，再列关键式) 结合式 (23.17) 和式 (23.18) 讨论为什么对大面积 PCSEL 不能把高 Q 直接等同于高 Purcell 因子和高 β 。

延伸阅读

- 下一章将把本章建立的动态与噪声语言，翻译成实验上真正汇报的线宽、相干性、 M^2 和亮度等指标。阅读时应同时核对本章变量在后续模型中的定义和适用范围。
- 若想补充速率方程、调制响应和线宽的经典背景，可参考 [10, 61, 19, 25]。阅读时应重点核对相关模型假设、参数范围与文献适用条件。
- 阅读性能章节时，建议始终追问一个问题：某个被报告的线宽或光束质量，到底对应的是哪一个动态工作窗口。阅读时应同时核对本章变量在后续模型中的定义和适用范围。

第二十四章 线宽、相干性、光束质量与亮度

读完本章，你应能

- 从严格定义出发理解线宽、相干性、 M^2 、亮度、偏振纯度与工作点之间的关系。
- 明确哪些性能量可以直接由确定性电磁/器件求解器给出，哪些量只能借助降阶噪声模型间接估计。
- 学会把性能指标写成“条件化、可复核、可对比”的结果，而不是脱离工作点的孤立数字。

读前准备 建议已掌握 第七章和二十至二十三 中关于阈值、单模工作窗口、鲁棒性窗口、动态稳定性和热-电-光反馈的内容。

阅读主线 本章所处模型层级：L4

前面两章已经分别回答了“名义设计对工艺扰动是否稳健”以及“起振以上的动态工作窗口如何理解”，而本章要回答的是“这些内部动态最终怎样投影成外界真正测到的指标”。这一步很容易再次发生层级混乱，因为实验测到的是线宽、远场、偏振、功率和亮度，而模型里最先算出的往往是复频率、 Q 、模场、温度场以及上一章中的动态中间量。我们将先建立性能指标的严格定义，再说明这些定义如何与 PCSEL 的工作点连接，随后专门讨论仿真和实验之间的接口，最后回答一个实践中最常被问到的问题：哪些量可以在 Lumerical 和 COMSOL 中直接算，哪些量只能通过它们提供的中间量来间接估计。

贯穿案例 R1：同一器件链采用同一报告协议 R1 的数字在本章按测量条件重新整理：脉冲功率、连续波功率、内禀线宽、远场角和亮度不能脱离口径与工作方式横向拼接。后文每次引用 500 μm 、1 mm 或 3 mm 器件，都会同时保留工作模式和指标定义；这也是 B0 模型输出能否与实验比较的最终检查。

24.1 为什么性能指标必须和工作点一起报告

对于普通课堂中的理想谐振腔，给出一个本征频率和一个 Q 往往已经能说明很多问题。对 PCSEL 来说，这远远不够。因为真实器件工作在外延层、电流路径和热反馈共同参与的环境里，同一个结构在不同电流、不同散热条件、不同封装反射和不同观测时间窗下，线宽、偏振纯度和远场都可能明显改变。

这意味着性能指标是工作状态的函数，不能作为脱离条件存在的“器件属性标签”。若不先说明工作点，就谈“该器件的线宽是多少”“该器件的光束质量是多少”，这种说法通常不严谨。教材式写法应当始终采用条件化表述，例如：“在 25 °C、连续波、某电流区间、某观测窗口下，线宽与偏振纯度落在什么范围内。”

这个原则也在防止错误建模：接受了“性能是工作点的函数”，就不会把冷腔结果直接冒充工作状态下的器件结论（红线 #2，见第2.10节）。

24.2 从一阶相干函数到线宽：定义先于经验公式

设输出复电场为 $E(t)$ ，一阶相干函数定义为

$$g^{(1)}(\tau) = \frac{\langle E^*(t)E(t+\tau) \rangle}{\langle |E(t)|^2 \rangle}. \quad (24.1)$$

若在某个近似区间内相位扩散可看作平稳随机过程，那么频谱与 $g^{(1)}(\tau)$ 互为傅里叶变换。在线形接近 Lorentz 形时，常把谱半高全宽记为线宽 $\Delta\nu$ 。这个定义看似标准，真正困难的是：PCSEL 常常同时受到快噪声和慢漂移影响，因此“线宽”可能依赖观测时间窗和拟合方法。

若进一步采用最经典的相位扩散模型，

$$g^{(1)}(\tau) = \exp(-\pi\Delta\nu_L|\tau|), \quad (24.2)$$

其傅里叶变换就是 Lorentz 谱线，而相干时间满足

$$\tau_c \approx \frac{1}{\pi\Delta\nu_L}. \quad (24.3)$$

式 (24.3) 只在 Lorentz 线形、单模、近稳态相位扩散主导的条件下成立；一旦慢漂移主导， τ_c 与“某个单独线宽数字”之间就不再能用这一条关系简单换算。

为了避免概念跳跃，最好把观测线宽写成分层诊断量

$$\Delta\nu_{\text{obs}} \approx \Delta\nu_{\text{int}} + \Delta\nu_{\text{thermal}} + \Delta\nu_{\text{env}}, \quad (24.4)$$

其中 $\Delta\nu_{\text{int}}$ 表示内禀相位噪声主导的部分， $\Delta\nu_{\text{thermal}}$ 表示温度起伏和慢热漂移引起的展宽， $\Delta\nu_{\text{env}}$ 则是外腔反射、机械振动和电源噪声等外禀项。这种写法属于**实验与建模对接时的诊断分解**，不构成严格的可加定理。它的意义在于提醒读者：只用一个瞬时测得的线宽数字，往往无法回答性能差异究竟来自腔本身、热反馈还是测试环境。

在半导体激光器近似理论中，人们常写出与输出功率、光子寿命和线宽增强因子有关的估计式。这里最常用的线宽增强因子就是 Henry 因子

$$\alpha_H = -2k_0 \frac{\partial n / \partial N}{\partial g / \partial N}, \quad (24.5)$$

其中 $k_0 = \omega/c$ 。[文献值] 对 GaAs 量子阱， α_H 常见于 2-5 区间 [19, 10]，该典型值主要对应单模行波半导体激光器语境。

PCSEL 中 Henry 因子的适用边界。式式 (24.5) 给出的 α_H 定义及其导致的 $(1 + \alpha_H^2)$ 线宽放大因子，是在以下假设下推导的：

1. 单模行波腔，振幅 - 相位耦合沿腔长均匀
2. 载流子浓度变化 δN 引起折射率变化 $\delta n = (\partial n / \partial N) \delta N$
3. 增益变化 $\delta g = (\partial g / \partial N) \delta N$ 与相位变化具有确定的比例关系

但对 PCSEL 驻波腔，这些假设可能不再成立：

- **空间烧孔效应：**驻波场的节点和反节点导致载流子消耗在空间上不均匀，有效 α_H 可能显著偏离材料值
- **多波耦合：**PCSEL 的四波或六波耦合结构中，相位响应是集体模式属性，不能简化为单一行波的 α_H
- **模式竞争：**近简并模之间的功率再分配会引入额外的低频相位噪声，这不属于 α_H 的描述范畴

因此，对 PCSEL 而言， $\alpha_H = 2-5$ 仅能作为数量级参考。当 (i) 主模与边模净裕量接近、存在模式跳变，(ii) 偏振分支在工作窗口内可切换，(iii) 温升回写显著改变模排序，或 (iv) 频率调制与相位噪声耦合成为系统指标时，应从单模 Henry 公式升级到多模/非正交或系统级噪声模型；这在高亮度与调制型 PCSEL 报道中尤为常见 [63, 25, 74]。

在行波近似下，线宽常相对理想 Schawlow-Townes 极限再乘上 $(1 + \alpha_H^2)$ 的放大量级。图24.1展示了 Henry 因子的定义及其导致的线宽展宽机理。右图中“ $\alpha_H = 3$ 时约放大 10 倍”应读作 [教学示意] 数量级，不是所有 PCSEL 的固定倍数。

24.3 频率噪声谱、短时/长时线宽与测量协议

若只报告一个“线宽”数字，而不说明它来自哪一种测量时间窗和拟合约定，结果往往不可比。更稳妥的现代写法，是先看频率噪声功率谱密度 $S_\nu(f)$ ，再讨论在给定观测协议下它如何投影成谱线宽。定性上说，高频白频率噪声更容易对应 Lorentz 型展宽，低频漂移和慢热起伏则更容易把谱线推向 Gaussian 或 Voigt 型轮廓。于是，“短时线宽”和“长时线宽”在实验上往往并非同一个数字。

把这条链写成最小公式，就是先引入相位与瞬时频率偏移的关系

$$\delta\nu(t) = \frac{1}{2\pi} \frac{d\phi}{dt}, \quad (24.6)$$

再用频率噪声谱 $S_\nu(f)$ 区分快慢贡献。若高频端可近似为白频率噪声，且采用常用单边谱约定，则对应的 Lorentz 线宽可写成

$$\Delta\nu_L \approx \pi S_\nu^{(\text{white})}. \quad (24.7)$$

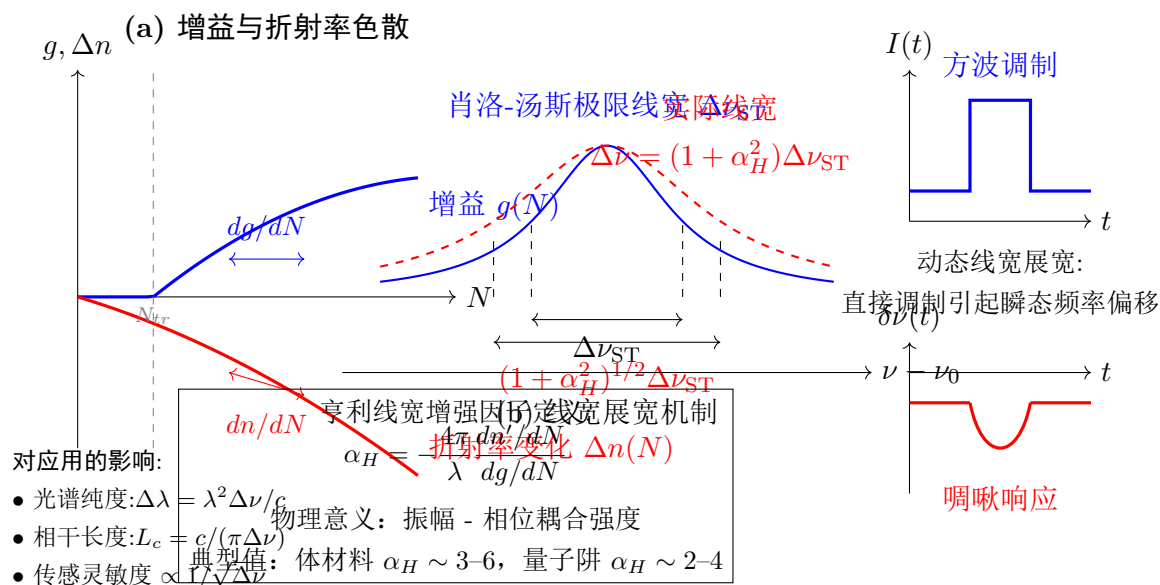


图 24.1: 亨利线宽增强因子 α_H 的物理来源及其对光子晶体面发射激光器光谱纯度的影响。(a) 左图示意载流子变化同时引起增益和折射率变化，从而产生振幅与相位耦合；(b) 右图示意 α_H 使实际线宽相对肖洛-汤斯极限进一步展宽，并在动态调制中引入频率啁啾。

该关系用于把频率噪声白噪声平台映射到短时 Lorentz 线宽，属于 [文献值] 经典口径 [19, 35, 10]。若低频端主要由慢漂移决定，频率分布更接近高斯，则常把 Gaussian 半高全宽写成

$$\Delta\nu_G = 2\sqrt{2\ln 2} \sigma_\nu, \quad (24.8)$$

其中 σ_ν 是观测窗口内的频率标准差。于是，当白噪声与慢漂移同时存在时，实验上常把谱线视为 Voigt 轮廓，并用近似公式

$$\Delta\nu_V \approx 0.5346 \Delta\nu_L + \sqrt{0.2166 \Delta\nu_L^2 + \Delta\nu_G^2} \quad (24.9)$$

估计总半高全宽 [10]。式 (24.7) 至式 (24.9) 的价值在于帮助读者建立一条清楚的诊断链：先看频率噪声来自快白噪声还是慢漂移，再决定线宽应按 Lorentz、Gaussian 还是 Voigt 去解释。

对教材读者而言，最小可执行协议至少应报告四件事：观测时间窗、频谱分辨率带宽、拟合线形 (Lorentz / Gaussian / Voigt)，以及是否做了偏振选择和模式选择。只有在这四个条件明确后，线宽比较才有物理意义。否则，同一器件在不同仪器设置下得到两个明显不同的“线宽”，并不一定意味着器件状态真的变了。

报告规范

跨平台外推三元注释

- **外推来源：**半导体线宽经典理论 (Henry、Petermann 与频率噪声法) 在 PCSEL 上的条件化迁移 [19, 35, 10]
- **边界条件：**需限定单模/多模状态、观测时间窗、偏振选择与外腔反馈条件，并给

拟合协议

- **不可迁移项**: 不可把边发射或 VCSEL 的单模近似线宽公式直接当作大面积 PCSEL 的通用终值; 尤其不可忽略热漂移与模式竞争

24.4 Petermann 因子与开放腔非正交性：什么时候必须提醒读者

对开放腔激光器，尤其是具有明显非厄米性质的模式，线宽讨论还会遇到一个额外修正语言：Petermann 因子。它最小地回答一个问题：**当激光模并非理想正交模时，噪声怎样被额外放大**。对 PCSEL 而言，这个语言不是每次都要展开成主公式，但至少要在教材里说明它的适用语境。

更稳妥的表述是：若目标模表现出明显开放性、非正交性增强，或器件工作点接近强非厄米模混合区域，那么把线宽只写成“Schawlow-Townes 乘一个 Henry 修正”可能不够。此时 Petermann 因子提供了一种提醒：开放腔中的模非正交会令相位噪声放大。它并非大面积 PCSEL 的默认主控量，但在讨论高 Q 、强开放、近简并或强模式混合区域时，不应完全缺席。

24.5 外腔反馈与环境耦合：为什么性能章也必须给出一句硬说明

第二十三章 已经把外腔反馈放进动态扰动入口，但本章还应从性能报告角度再补一句更硬的话：**任何线宽、相干性或稳定性指标，若未说明外腔反射与封装环境，就不应被自动读成“器件内禀极限”**。

这是因为 PCSEL 的面发射几何天然更容易和自由空间光路、窗口片和封装界面发生回馈耦合。实验上经常出现的现象是：冷腔和工作点都没有明显变化，但一旦更换外部光路或测试台，低频频率噪声、长时线宽和偏振稳定性就一起改变。对教材读者来说，最重要的是建立报告纪律：**内禀量、外禀量、短时量和长时量必须分开**。

24.6 相干性与线宽相关，但并不完全等价

教材中常把“线宽越窄，相干性越好”写成一句简单结论。这个说法在理想单模、稳定噪声主导的场景里有一定道理，但在 PCSEL 里需要补足两个限定。

第一，相干性有时间相干和空间相干两个侧面。线宽主要对应时间相干，而大面积 PCSEL 的空间相干则与横向模分布、相位一致性和模式竞争更直接相关。一个器件即使在短时间窗内线宽不宽，若横向上已经出现多模分区或局部失相干，其远场与亮度仍然会明显变差。

第二，线宽本身也可能同时混入慢漂移和快噪声。若观测窗口从微秒扩展到秒级，谱宽可能显著增加，但这不一定意味着腔本身的内禀噪声变大，而可能是温控和封装反射在更长时间尺度上显现出来。

因此，在 PCSEL 的研究报告中，应同时给出线宽、观测时间窗、偏振条件、工作电流和温度，避免孤立汇报“线宽”两个字。这样，线宽才能与相干性和模式稳定性建立明确对应。

24.7 M^2 、亮度和功率为什么必须一起看

高功率并不自动等于高亮度，这是大面积面发射器件最容易出现的误判。亮度常写为

$$B = \frac{P}{A\Omega}, \quad (24.10)$$

其中 P 是输出功率， A 是有效出光面积， Ω 是发散对应的立体角。若器件面积增大导致输出功率上升，但横向多模使远场发散明显变宽，则亮度可能不升反降。于是，对 PCSEL 来说，功率、 M^2 和亮度必须一起看，而不能只报其中一个最直观的数字。

M^2 则提供了把真实光束与理想高斯基模比较的桥梁。对于大面积 PCSEL，它是空间相干和模式纯度的一个非常直观的外部指标。若同一器件在电流升高后 M^2 明显变坏，同时偏振纯度下降、远场主瓣漂移增强，那么最合理的解释通常是模式竞争和热反馈已经开始破坏目标模的主导地位，而非“输出只是更强了”。

若采用二阶矩定义，并在某一横向方向上记近场标准差为 σ_x 、远场角谱标准差为 σ_{θ_x} ，则

$$M_x^2 = \frac{4\pi}{\lambda} \sigma_x \sigma_{\theta_x}, \quad M_y^2 = \frac{4\pi}{\lambda} \sigma_y \sigma_{\theta_y}. \quad (24.11)$$

对理想高斯基模有 $M_x^2 = M_y^2 = 1$ 。若光束近似圆对称、且希望做快速数量级判断，也常写成

$$M^2 \approx \frac{\pi w_0 \theta}{\lambda}, \quad (24.12)$$

其中 w_0 是近场束腰半径， θ 是远场半角。式 (24.11) 和式 (24.12) 的区别应在教材里说清：前者更严格，适合实验报告与复杂光束；后者更直观，适合教学估算。

例如，若某器件从 1 A 升到 2 A 时总功率上升 50%，但远场半角扩大一倍、 M^2 明显变差，那么亮度通常不会同步上升。此时应优先检查目标模是否已经失去对竞争模的净优势，而非继续追求更高电流。

例子

再给一个最小数量级例子。若把大面积 PCSEL 的主出光近场近似为直径 D 的均匀圆孔，则衍射极限半角可粗略写成

$$\theta_{1/e^2} \sim \frac{2\lambda}{\pi D}.$$

取 $\lambda = 980 \text{ nm}$ 、 $D = 200 \mu\text{m}$ ，得到

$$\theta_{1/e^2} \sim 3 \times 10^{-3} \text{ rad} \approx 0.18^\circ.$$

这个数字提供的是比较基线，不是适用于所有器件的统一标准：若实测发散角远大于该量级，就不应先把原因归咎于“大面积”，而应优先检查近场幅度不均、相位起伏、边缘泄漏、热透镜或多模共存。

Noda 等人的 2023 tutorial 给了一个很好的性能报告样本库：同样是“高亮度 PCSEL”， $500 \mu\text{m}$ 双晶格器件可在脉冲条件下达到 10 W 级、约 0.2° – 0.3° 发散；引入背面 DBR 后，脉冲峰值功率超过 20 W，亮度达到约 $1.5 \text{ GW cm}^{-2} \text{ sr}^{-1}$ ；1 mm 器件在 CW 下达到 10 W 级，并通过频率噪声白噪声平台给出低于 1.23 kHz 的内禀线宽估计；3 mm 器件进一步实现 50 W

CW、约 0.05° 发散和 $1 \text{ GW cm}^{-2} \text{ sr}^{-1}$ 亮度 [50, 75, 74]。这些数字必须和口径、脉冲/CW、热边界、线宽测量带宽、远场定义一起读（见表24.1）。

尤其要注意 3 mm CW 结果中的一个细节：论文报告的 M^2 约为 2.36，但器件仍被判定为单模工作；原因主要是均匀电流注入导致的超高斯近场强度分布，而不是传统意义上的多横模混合 [50]。因此，对 PCSEL 而言， $M^2 > 1$ 不能自动等价于“多模失败”，必须结合光谱、近场、远场、偏振和工作点一起解释。

M^2 在 PCSEL 中的适用性限制与替代表征

上面的 $M^2 \approx 2.36$ 案例背后有一个普遍问题：把二阶矩定义式 (24.11) 直接套在 PCSEL 上存在三个固有偏差。其一，近场强度包含晶格尺度的周期性调制，会使 σ_x 人为偏大，从而高估 M^2 ；其二，周期近场的傅里叶变换会在远场产生高阶衍射瓣，这些旁瓣对二阶矩的贡献可能主导计算结果；其三，对这类周期调制光束，偏大的 M^2 反映的可能是“光束携带周期结构信息”，而非实际可聚焦性退化。

因此在报告 PCSEL 光束质量时，建议辅以下列替代表征之一，并说明口径：

1. **围封功率 (encircled power)**：计算包含 86.5% 总功率的圆斑直径，与理想高斯光束比较；
2. **Strehl 比**：主瓣峰值强度与同功率理想高斯光束峰值强度之比， $S > 0.8$ 通常认为接近衍射极限；
3. **主瓣发散角**：仅对中心主瓣（排除旁瓣）拟合并提取发散角；
4. **亮度**：使用 $B = P/(\lambda^2 M_x^2 M_y^2)$ 时需注明 M^2 的计算方法和是否滤除旁瓣。

最稳妥的做法是同时报告全光束与主瓣两个口径的 M^2 ，以区分周期性调制与真实的光束质量退化。

仿真中获取 M^2 的典型流程：在近场监测面记录复振幅 $E(x, y)$ ；傅里叶变换得到远场 $\tilde{E}(k_x, k_y)$ ；分别计算强度与角谱强度的二阶矩；代入式 (24.11)。对非旋转对称结构（如椭圆孔、矩形晶格）， M_x^2 与 M_y^2 可能不相等，应分别报告而非给单一标量。

表24.1 训练的是一种读文献的顺序：先看工作条件，再看指标定义，最后才比较数字。若忽略这一步，50 W 连续波、20 W 脉冲峰值、1.23 kHz 内禀线宽估计和 $M^2 \approx 2.36$ 很容易被错误地并排解读；实际上，它们分别回答的是功率可扩展性、辐射提取、时间相干和空间光束质量的不同问题。

24.8 偏振纯度、远场稳定性和模式稳定性本质上连在一起

PCSEL 的偏振与远场并非两个孤立测量量。它们都来自近场对称性和辐射通道的组织方式，而这又会被注入和温升动态改写。于是，在工作点扫描中常见到这样的联动：偏振纯度下降的同时，远场主瓣变宽或偏移；又或者远场出现副瓣增强的同时，线宽也开始展宽。这些现象不应被当作互不相干的“多个问题”，它们通常是同一个模式竞争过程在不同观测端口上的

表 24.1: 从 Noda 2023 tutorial 中抽出的性能报告矩阵。数值必须和工作条件、口径、远场定义和测量协议一起读。

器件阶段	代表性性能	必须绑定的条件	正确读法
500 μm 双晶格	10 W 级脉冲输出, 约 $0.2^\circ\text{--}0.3^\circ$ 发散	脉冲工作、有限孔阵直径、远场角定义	证明双晶格可扩大模式裕量, 但不能直接外推为任意 CW 口径下稳定
背面 DBR 优化	脉冲峰值功率超过 20 W, 亮度约 $1.5\text{ GW cm}^{-2}\text{ sr}^{-1}$	DBR 相位、上下出光比例、热负荷较弱的脉冲条件	说明外部垂直结构也会进入非 Hermitian 耦合预算, 不能把 PC 层单独看完
1 mm CW 器件	10 W 级 CW, 频率噪声白噪声平台对应低于 1.23 kHz 的内禀线宽估计	线宽分析仪带宽、白噪声平台提取方式、热沉条件	线宽是噪声协议下的推断量, 不是冷腔仿真可直接输出的单个数字
3 mm CW 器件	50 W CW, 约 0.05° 发散, 亮度约 $1\text{ GW cm}^{-2}\text{ sr}^{-1}$, $M^2 \approx 2.36$	连续波热工作点、超高斯近场、光谱与远场共同判据	$M^2 > 1$ 不自动等价于多模; 必须同时检查光谱、近场、远场、偏振和工作窗口

投影。

因此, 本章讨论性能时必须和前一章连起来看: 当目标模相对于竞争模的裕量缩小时, 最先暴露异常的, 不一定是阈值或总功率, 而往往是偏振纯度、远场或短时间稳定性。性能指标之所以是高价值指标, 恰恰因为它们能比“器件还能不能亮”更早地暴露设计边界。

24.9 哪些量可以在 Lumerical 和 COMSOL 中直接算, 哪些只能间接估计

这一节直接回答建模实践中的常见问题。Lumerical 和 COMSOL 本质上都以确定性电磁场、输运方程和热传导方程为核心。它们最擅长的, 是给出模式、损耗、场分布、载流子分布、温度分布以及这些量对参数的灵敏度。而线宽和相干性则牵涉随机噪声与相位扩散, 通常不能直接通过一次确定性全波仿真“求出来”。

表24.2 给出一张实用区分表。

如果要说得再直接一点:

- 想算候选模、远场、偏振、被动损耗和有限阵列出射, 优先使用光学求解器;
- 想算电流扩展、阈值电流的器件级来源、串联电阻和非均匀增益, 必须进入 CHARGE 或 Semiconductor 一类漂移-扩散求解器;
- 想算温升、热源和热漂移, 必须进入 HEAT 或 Heat Transfer 一类热求解器;
- 想直接“从一次仿真得到线宽”, 通常做不到, 只能用上述求解器提供的中间量建立降阶噪声模型。

严格地说, 全波求解器给出的是线宽模型所需的参数, 而不是线宽本身的全部统计信息。

表 24.2: 性能指标与 Lumerical / COMSOL 的典型建模接口。重点在于区分“可直接计算量”和“需降阶估计量”。

指标	物理来源	Lumerical 中常见做法	COMSOL 中常见做法
共振频率、模场、 Q	冷腔或含损耗光学	FDTD / MODE / RCWA 提取复频率、衰减率、场分布	Wave Optics 本征频率或频域求解 + PML
远场、偏振、上下出光比例	开放辐射通道	周期 RCWA 或有限阵列 FDTD + near-to-far	Wave Optics 频域 + 端口 / 远场后处理
约束因子、损耗重叠	模场与层栈重叠	MODE / FDTD 后处理重叠积分	Wave Optics 场积分后处理
电流扩展、J-V、载流子分布	漂移-扩散电学	CHARGE	Semiconductor
温度场、热漂移灵敏度	热传导 + 热源	HEAT	Heat Transfer
线宽、相干时间	噪声与相位扩散	由 Q 、功率、热灵敏度等中间量进入降阶噪声模型	同左，通过中间量进入降阶模型
亮度、 M^2	远场与功率联合定义	由远场与功率后处理得到	由远场与功率后处理得到

24.10 性能报告的正确写法：条件、区间和可复核性

到了收束阶段，最值得坚持的写法是把结果写成“条件 + 区间 + 提取方法”的结构。与其写“器件线宽为多少、亮度为多少”，不如写成：在何种电流、温度、观测窗口和偏振选择下，线宽、 M^2 、偏振纯度和远场主瓣角分别落在哪些范围，提取算法是什么，仿真或实验是否重复验证过。

这种写法看起来更繁琐，但它有两个长远好处。第一，它天然阻止跨层级偷换概念，因为每个指标都带着自己的求解或测量上下文。第二，它使模拟与实验更容易对接，因为双方都知道哪些中间条件必须一致才能比较。

易错点

把某一电流点、某一时间窗、某一偏振状态下得到的“最优线宽”当成器件代表性能，是典型的过度报告。对于 PCSEL，这种写法几乎必然掩盖模式竞争和热漂移带来的窗口效应。

24.11 2024–2026 前沿补充：高相干与系统化应用

近两年的面发射路线更强调“系统可用性”而非单一指标极值。一个明显趋势是：高亮度 PCSEL 开始直接面向长距离相干自由空间通信与高速调制场景，性能报告不再只给静态线宽，而是联动给出调制方式、链路稳定性与环境约束 [26, 74]。另一条趋势是谐振腔嵌入式 PCSEL 路线，通过结构耦合改善模式选择与噪声约束，使“线宽-功率-远场”三者协同优化

更具可操作性 [9]。在理论侧, linewidth-power 乘积模型为大面积单模 PCSEL 的线宽缩放提供了更直接的判据, 并与近年的窄线宽高功率器件结果形成了可对照闭环 [70, 6]。这意味着本章的线宽讨论不应停在“单点最小线宽”, 而应转向“给定功率区间内的线宽斜率、模式稳定性和测量协议是否一致”。对本书主线的含义是: 性能章应从“报单点最好值”升级为“报工作窗口 + 报系统条件 + 报可复核协议”。与全书统一前沿矩阵可对照 表1.6。

回看本章

PCSEL 的性能指标必须在工作点上下文中理解。线宽来自谱形定义并依赖观测窗口, 相干性既有时间侧也有空间侧, M^2 和亮度要求把功率与发散一起看, 偏振纯度和远场稳定性则是模式稳定性的外部投影。频率噪声谱、短时/长时线宽、Petermann 因子和外腔反馈进一步提醒我们: 开放腔激光器的“性能数字”总是带着测量协议和环境边界条件。在建模层面, Lumerical 与 COMSOL 都擅长直接计算模式、辐射、载流子和温度等确定性中间量, 但线宽和相干性通常需要借助这些中间量进入降阶噪声模型间接估计。可靠的性能报告, 是报告一个可复核的工作窗口, 而非“报一个最直观的数字”。

练习题

1. [基础] 解释为什么在 PCSEL 中, 线宽不能脱离观测时间窗单独比较。
2. [基础] 结合 式 (24.10) 说明为什么高功率并不必然意味着高亮度。
3. [进阶] 试着把“偏振纯度下降、远场变宽、线宽增加”三件事串成一条可能的模式竞争链。
4. [综合] (思路提示: 先定层级, 再列关键式) 选择一个你关心的性能指标, 说明它在 Lumerical 或 COMSOL 中是直接计算量, 还是需要通过中间量间接估计。

延伸阅读

- 下一章将把前面所有理论主线压缩成一套可执行的 PCSEL 仿真 workflow, 并明确不同软件链如何协同。阅读时应同时核对本章变量在后续模型中的定义和适用范围。
- 若想更深入理解半导体激光器中的线宽与相干, 可参考 [61, 10, 19] 中关于相干函数、幅相耦合、线宽和光束质量的内容。阅读时应重点核对相关模型假设、参数范围与文献适用条件。

第五部分

workflow、例题与失败模式

第二十五章 一套实用的 PCSEL 仿真 workflow

读完本章，你能

- 把前面二十余章建立的概念、方程和近似组织成一条可执行的研究 workflow。
- 明确不同阶段该计算什么、向下一阶段交付什么，以及哪些结论在该阶段绝不能提前宣称。
- 在求解器无关的原则下，理解 Lumerical 与 COMSOL 分别更适合承担哪些计算，并掌握最小可执行建模过程。

读前准备 建议已掌握 第十六章、第十七章、第十九章、第二十章和二十二至二十四 中关于方法分工、外延层、电学、热学、鲁棒性分析、动态稳定性和性能报告的内容。

本章结构

本章所处模型层级：L1→L4

到这里，全书已经把 PCSEL 的主要物理链条铺开了。剩下的问题不再是“某个公式怎么推”，而是“研究时应该先做什么、后做什么、哪些结果能信、哪些结果还只是候选”。本章把这些环节整理成可执行的工作流。我们先建立分层结论和数据账本，再逐阶段说明纵向模检查、周期单胞筛选、有限器件验证、电热建模、热-电-光回写以及鲁棒性分析的输入输出，随后专门讨论两条常见实现路径：以 Lumerical 为主的光学优先链，以及以 COMSOL 为主的多物理场优先链。最后给出停机判据和结果交付模板。图25.1 把指标、结构、能带、阈值、远场、热与容差串成七步闭环；其中 $\Delta T \lesssim 50\text{ K}$ 和几何偏差 $\pm 5\text{ nm}$ 仅是 [本书算例] 筛选门槛，发布级判据仍须绑定波段、封装散热与工艺统计重新定义 [46, 17]。

25.1 本章如何与下一章例题和失败模式章配合使用

如果只把本章当成“流程图章节”，读者容易知道名词，却不知道每一步真正长什么样；如果只看下一章例题，又可能误以为那只是某一个特定结构的演示，而不是一般工作流的落地版本。较稳妥的读法是把 第二十五章至第二十七章 连成一套使用：本章给出正向流程，第二十六章 把它落实成一次完整作答，而 第二十七章 则告诉你每一步最常见的失误会长成什么样。

因此，建议不要孤立阅读本章。较合适的顺序是：先读本章建立流程骨架，再在 第二十六章 中按表完成各阶段检查，最后用 第二十七章 反向审计整条链。

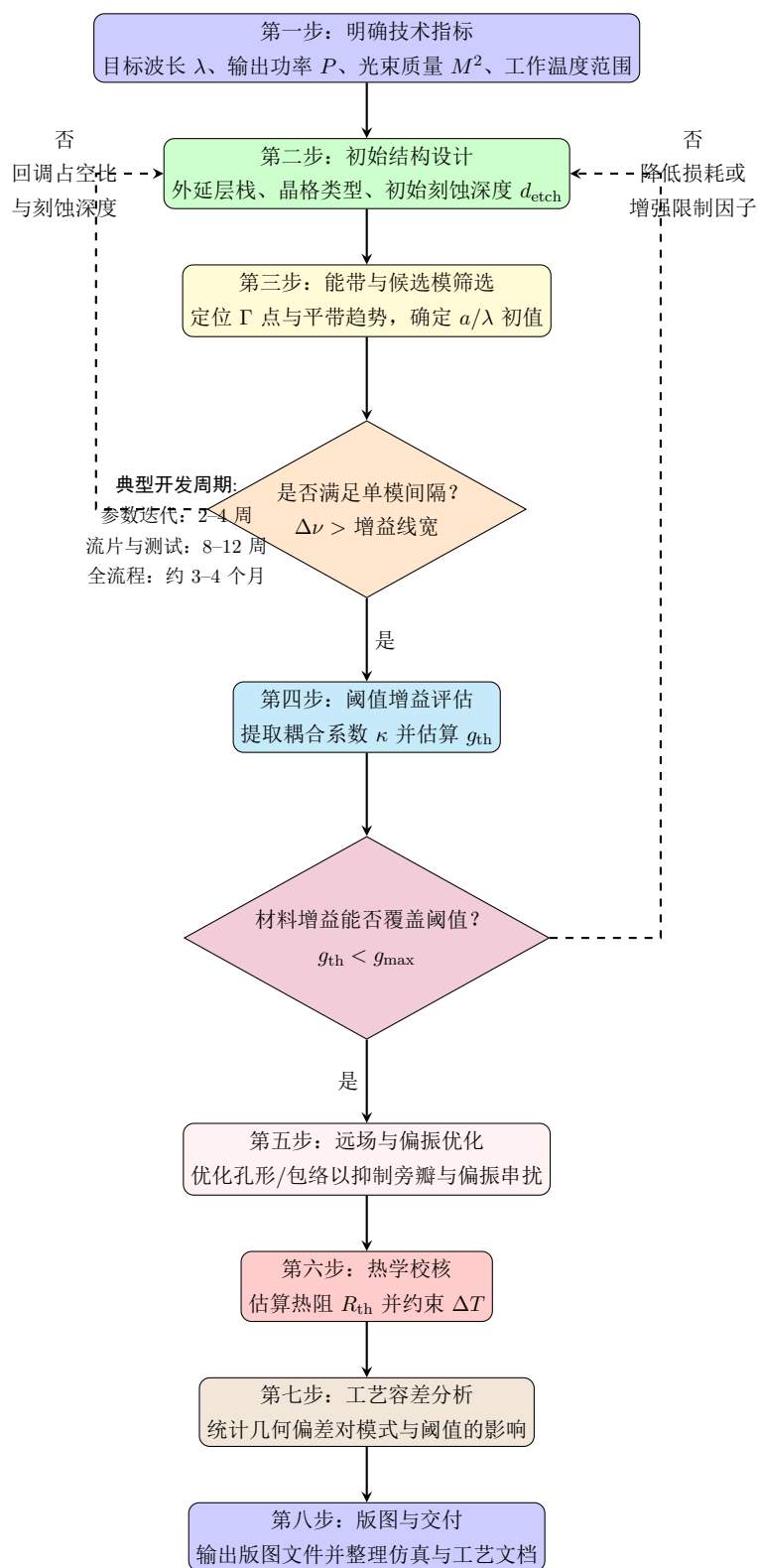


图 25.1: 光子晶体面发射激光器从指标定义、结构初设、候选模筛选、阈值评估、远场与热优化，到容差分析和版图交付的完整工程流程。图中“温升约束”和“几何容差”仅作为 [本书算例] 示例门槛，需按目标波段、散热边界和工艺统计重标定 [46, 17]。

表 25.1: workflow、例题与失败模式的对应关系。

阶段	本章任务	在 第二十六章 中的落地位置	若这一步做错，在 第二十七章 中最接近的失败类型
A	纵向模与层栈检查	例题 B 的“第一步：纵向模检查”	把错误层栈背景带入后续所有计算；误用有效折射率近似
B	周期单胞筛选	例题 B 的“第二步：单胞候选模筛选”	把单胞开放结果直接当器件结论；误读被动排序
C	有限器件全波验证	例题 B 的“第三步：有限阵列检查”	有限尺寸建模不足；边界条件与几何定义不一致
D	电注入与热建模	例题 B 的“第四步：简化电学/热学回写”	把均匀增益当真实注入；忽略热回写
E/F	工作窗口与鲁棒性	例题 B 的“中间结果汇总与结论分层”	名义值最优但无工艺窗口；对工作窗口层级越权宣称
G	实验回写链	例题 B 的“哪些问题必须交给实验回写链”	用已拟合量冒充预测能力；缺少回验量

表 25.2: workflow 各阶段的典型计算资源量级。这里给的是项目排程意义下的数量级估计，而不是普适硬门槛。

阶段	主要任务	典型耗时量级	资源判断
A	纵向模与层栈检查	秒到分钟	二维模态问题，内存压力通常较小，适合大量参数初筛
B	PWEM / RCWA 单胞筛选	秒到分钟；做参数扫时可到小时	周期单胞便宜，但参数扫会迅速放大总成本
C	有限阵列 FDTD / FEM	小时到天	三维全波往往是第一道主要算力门槛，常需多核、大内存或 GPU
D	电注入与热建模	分钟到小时	比全波便宜，但网格与非线性收敛会明显受几何复杂度影响
E/F	热回写与鲁棒性分析	小时到天	单次求解未必最慢，但多轮迭代和多样本统计常使总成本最高
G	实验回写链与校准	组织成本最高	计算时间未必最长，但参数整理、版本控制和回验最消耗研究日程

注记

FDTD 内存需求的快速估算。对三维 FDTD 仿真，内存占用可由网格数量粗略估算：

$$\text{RAM (GB)} \approx \frac{N_x N_y N_z \times N_{\text{fields}} \times 8 \text{ bytes}}{10^9},$$

其中 N_x, N_y, N_z 是各方向的网格数， $N_{\text{fields}} \approx 6-12$ （电场、磁场、辅助变量、PML 等）。对典型 PCSEL 仿真：

- 单胞周期结构： $100 \times 100 \times 50$ 网格 $\rightarrow$ 约 0.5–1 GB
- 有限阵列（ 10×10 晶格）： $500 \times 500 \times 100$ 网格 $\rightarrow$ 约 15–30 GB
- 大尺寸器件（ 100×100 晶格）： $5000 \times 5000 \times 100$ 网格 $\rightarrow$ 约 1.5–3 TB

因此，对有限阵列 PCSEL 的全波仿真，通常需要至少 32–64 GB 内存的工作站，或使用子网格、对称性、周期边界等技巧降低网格总量。

25.2 工作流的第一原则：先定义结论层级，再跑仿真

仿真失败常常是因为在第一步就没有定义清楚自己打算得到哪一层结论，而非因为不会点界面操作。对 PCSEL 来说，最安全的做法是把结论分为四层：

1. L1 结构候选层：只谈候选模、对称性、被动辐射趋势和纵向重叠；
2. L2 光学阈值层：只谈开放结构中的阈值增益、偏振趋势、上下出光比例和有限尺寸模式排序；
3. L3 电学映射层：开始谈注入分布、J-V、串联电阻、阈值电流区间的来源；
4. L4 热-电-光工作层：开始谈工作窗口、稳定性、亮度、偏振稳定性与热漂移后的器件表现。

如果还停留在 L1 或 L2，就绝不能提前宣称阈值电流、线宽窗口或高温稳定工作。这个分层是为了防止求解器输出和物理结论错位，而非文风问题。

严格表述（难度）

PCSEL 工作流中最重要的纪律是：**低层结论可以进入高层模型，高层结论不能倒过来由低层结果“脑补”出来。**被动 Q 可以进入阈值增益估计，但被动 Q 本身不能直接变成阈值电流或工作线宽。

图25.2 把这一纪律画成阶段门禁。每跨过一级，是必须增加一种新的物理证据，而非多跑一个软件名；缺少对应证据时，结论应停留在较低层级。

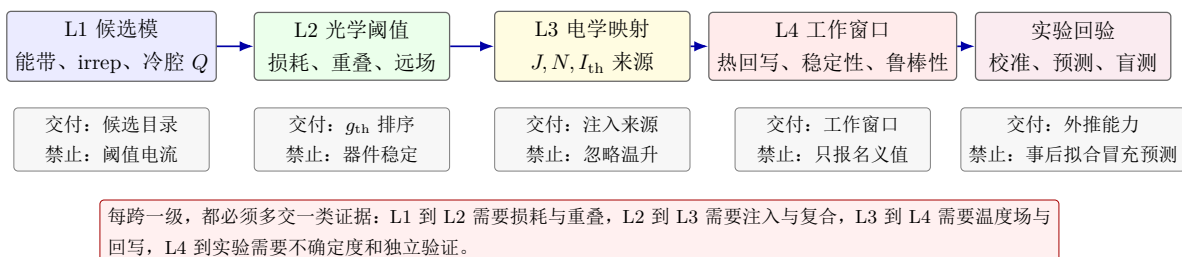


图 25.2: PCSEL 仿真 workflow 中的证据门禁。L1 允许给出候选模目录, L2 允许给出光学阈值排序, L3 才开始讨论注入与阈值电流来源, L4 才能讨论热回写后的工作窗口; 实验回验需要独立不确定度和未参与拟合的观测量。

25.3 工作流的第二原则: 先建数据账本, 再建几何

在任何求解器启动之前, 先建立参数账本 $\mathcal{D}$ 。它至少应记录四类信息: 几何参数、材料参数、外延层参数和电热参数。每个参数都至少带四个元信息: 单位、来源、适用温度或频段、当前不确定度。没有元信息的参数, 不应直接进入主结果模型。

这个步骤看似偏向项目管理, 但直接影响后续结果的可复核性。因为 PCSEL 的仿真链跨越多个求解器, 若不在一开始统一单位、坐标和参数来源, 后面最容易发生的是参数错配, 而非数值误差。例如, 光学模型用的是冷态折射率, 电热模型却默认了工作温度; 或者 RCWA 的单胞孔径与 CHARGE 的电极开窗尺寸根本不是同一个版本。一个看似微小的版本错配, 足以让整条 workflow 失去可复核性。

把“数据账本”落实成可执行的 manifest 为了让这条原则不止停留在原则表述, 建议把参数账本直接写成一个可版本控制的 manifest。最小字段至少应包括: 参数唯一 ID、显示符号、文件字段名、单位、数值、来源、适用波长/温度、被哪些阶段使用, 以及当前不确定度。一个最小 YAML 模板可以长成下面这样:

```

parameter:
  id: MAT-GAAS-N-001
  name: n_GaAs
  symbol: n
  file_key: n_gaas
  unit: 1
  value: 3.42
  source: literature
  valid_wavelength_nm: [930, 990]
  valid_temperature_C: [20, 80]
  uncertainty: 0.02
  used_in:
    - stage_A_waveguide
  
```

```

- stage_B_unitcell
- stage_C_finite_array
version: v1.2
last_updated: 2026-03-01

```

在项目实施中，这个 manifest 还应再加两列：`propagates_to`（它会影响哪些输出量）和 `status`（direct-measured / literature / fitted / inferred）。这样处理后，读者第一次独立搭建建模链时，就可以避免把文献冷态参数与实验反演的工作点参数混入同一张无版本表格。

25.4 最小命令行复现骨架：即使使用商业软件，也应把文件流做成终端原生

本书不把 GUI 界面操作教程当主线，但这不等于 workflow 只能停在“软件名词表”。较稳妥的做法，是先把目录、参数文件、阶段脚本和输出制品固定成一条 terminal-native 骨架，再决定每一步内部到底由 Lumerical、COMSOL 还是其他求解器来完成。一个最小项目骨架可以写成：

```

pcsel_project/
  params.yaml
  01_layers/
  02_unitcell/
  03_finite_array/
  04_electrical_thermal/
  05_postprocess/
  run_stage_A.py
  run_stage_B.py
  run_stage_C.py
  run_stage_D.py
  run_stage_E.py

```

其中每一步至少应有固定的输入、输出、sanity checks 与失败回扫项：

- Stage A: 输入层栈与材料表，输出 n_{eff} 、 Γ_{QW} 、 Γ_{loss} ；sanity check 是模式主峰位置和量纲自洽。
- Stage B: 输入单胞几何与层栈背景，输出候选模频段、被动损耗排序、上下辐射比例；sanity check 是 PWEM / RCWA 频段与对称性标签一致。
- Stage C: 输入有限阵列几何，输出近场、远场、主模排序和尺寸收敛；sanity check 是尺寸/PML/网格继续加密后主结论不翻转。
- Stage D: 输入外延、电极与接触参数，输出 $J(\mathbf{r})$ 、 $N(\mathbf{r})$ 、 $T(\mathbf{r})$ ；sanity check 是边界条件、单位和功率守恒一致。

- Stage E: 输入前四阶段结果, 输出回写后的净模裕量、工作窗口和鲁棒性报告; sanity check 是结论层级不越权。

这套骨架的目的是强制读者把文件流、参数流和结论层级固定下来, 而非承诺“整本书都提供开源全复现脚本”。只要这一步做对, 后续无论内部调用商业软件还是自写脚本, 研究链条都会清楚得多。

25.5 阶段 A: 先做纵向模和层栈检查, 而不是急着扫平面晶格

PCSEL 的第一步不应直接扫单胞孔径, 而应先确认外延层给出的纵向模式背景是否合理。也就是说, 先解层栈中的纵向模、有效折射率、量子阱重叠和损耗重叠。若这一关没有通过, 后续平面内晶格优化很可能建立在错误的纵向模型上。

在 Lumerical 路径里, 这一阶段通常最适合用 MODE/FDE 之类的波导模求解器: 建立二维横截面层栈, 定义材料色散和损耗, 求出目标偏振附近的导模, 输出 n_{eff} 、场分布、 Γ_{QW} 与与损耗层重叠积分。最小建模过程可以写成:

1. 画出外延层截面并赋予折射率、掺杂相关吸收代理;
2. 在工作波长附近做本征模求解;
3. 提取目标模的 n_{eff} 、纵向场峰值位置和重叠因子;
4. 对层厚与折射率做小扰动扫描, 判断结果是否过于脆弱。

在 COMSOL 路径里, 这一阶段通常用二维横截面的 Wave Optics 本征/模态分析来完成。最小过程是: 建立二维层栈几何、指定材料、用适合导模的边界条件和本征求解得到传播常数, 再做后处理积分得到重叠与损耗趋势。这里的核心是输出必须一致: 都要产出一个可供下一阶段使用的纵向模式背景, 而不仅仅是一张场图, 而非软件界面差异。

25.6 阶段 B: 周期单胞筛选, 目标是排除错误结构而不是宣布最终器件

有了可靠的纵向背景, 才进入单胞筛选阶段。这里的任务是回答: 在无限周期、开放结构、小信号框架下, 哪些 Gamma 点候选模、哪些偏振状态和哪些辐射通道值得继续投入有限器件验证。

这一阶段最好再拆成两个子步骤。先按第五章、第十三章的方法做一次被动无限周期候选模筛选: 用 PWEM 建立候选频段、Gamma 点简并关系和主要平面波分量; 再把已经缩小后的候选模窗口交给 RCWA 或周期频域模型, 去检查开放辐射、上下出光比例和偏振趋势。这样做的好处是, RCWA 不必在过大的参数空间里盲扫, 而是围绕已经识别出的候选模做精修。

Lumerical 路径中, 若目标是快速扫描单胞参数和辐射分配, 优先使用 RCWA; 若需要更灵活地看时域响应或近场细节, 也可用周期边界下的 FDTD, 但成本通常更高。一个最小 RCWA 过程可以概括为:

1. 建立单胞几何、上/下介质和层栈；
2. 设定法向附近的入射波或端口条件，并扫描波长、孔径、刻蚀深度和微扰参数；
3. 读取反射、透射、衍射效率、局域场增强和上下辐射分配；
4. 用收敛的傅里叶阶数确认候选模排序与偏振趋势。

COMSOL 路径中，这一步通常用三维单胞频域模型完成：在横向使用相位周期边界，在上下使用端口或开放边界，并通过频率扫或参数扫读取散射量。若要说得再具体一点，就是“用 Wave Optics 的频域单胞模型来做 RCWA 所擅长的那类周期散射判断”，只不过实现框架是以有限元频域周期散射问题出现，而非以 RCWA 名义出现。

这一阶段的交付物应当是候选模目录，而不是“最终器件最低阈值”的宣称。目录中至少应包含：候选频段、PWEM 给出的 Gamma 点模族和主要傅里叶分量、目标/竞争模的被动损耗排序、偏振趋势、上下出光比例和对工艺扰动的敏感性。

25.7 阶段 C：有限器件全波验证，检查无限周期假设到底失效到什么程度

通过单胞筛选的结构，下一步就要离开无限周期近似，进入有限孔阵模型。这里的任务不再是看单胞散射特征，而是看真实有限阵列中模式是否仍保持正确排序、边缘是否引入新的泄漏通道、远场是否还能维持目标主瓣和偏振状态。

在 Lumerical 路径里，这一步最常见的是 FDTD。最小过程通常是：

1. 把候选单胞复制成有限阵列，并加入真实上下层栈；
2. 在开放方向使用 PML，在必要时利用对称边界降低计算量；
3. 用脉冲源、偶极源或与目标模匹配的激励源激发结构；
4. 通过时域衰减、频谱、场分布和近远场变换提取复频率、 Q 、远场和主模排序；
5. 对阵列尺寸、PML 厚度、网格和仿真时间窗做收敛检查。

在 COMSOL 路径里，这一步常用三维 Wave Optics 的频域或本征频率研究来做。若重点是共振位置和模式形状，可用本征频率研究并在开放边界加 PML；若重点是给定频率下的响应和远场，则更适合频域研究。需要强调的是，COMSOL 的有限元框架在处理复杂材料依赖和后续多物理场联动时很灵活，但三维有限孔阵的计算代价往往不小，因此必须严格控制模型尺寸和收敛验证。

这一阶段的输出至少应包括：有限阵列主模与竞争模的排序、目标模近场与远场、边缘敏感性、以及“无限周期结论在有限阵列中是否仍成立”的判断。

25.8 阶段 D：电注入与热建模，开始把阈值增益映射成阈值电流区间

如果研究目标是器件而不是纯光学结构，那么到这里就必须进入电学与热学。任务不再是找模，而是问：电流怎样进入器件、载流子密度怎样分布、哪里掉压、哪里发热、这些分布怎样改写模增益与模损耗。

在 Lumerical 路径里，常见的分工是用 CHARGE 解漂移-扩散和泊松方程，用 HEAT 解热传导。最小过程可以概括为：

1. 依据真实外延层和电极定义半导体几何与掺杂；
2. 在 CHARGE 中指定接触边界、复合模型和工作电压/电流，求得 $J(\mathbf{r})$ 与 $N(\mathbf{r})$ ；
3. 依据 Joule 热和非辐射复合构造热源，送入 HEAT 得到 $T(\mathbf{r})$ ；
4. 把 N 和 T 场转写成折射率修正、增益修正和吸收修正，回写给光学阶段。

在 COMSOL 路径里，对应的是 Semiconductor 与 Heat Transfer 接口。几何通常可与光学模型共享，只是在网格、边界和研究类型上要做针对性修改。最小过程是：

1. 在 Semiconductor 接口中设置接触、掺杂、复合与偏置，解出电势和载流子分布；
2. 在 Heat Transfer 接口中加入 Joule 热与复合热源，求稳态或瞬态温度场；
3. 将温度和载流子结果通过自定义材料关系回写到折射率和增益代理中；
4. 重新运行 Wave Optics，检查目标模与竞争模的净裕量是否改变。

这里必须再次提醒：除非你真的完成了这个映射链，否则不能把“阈值增益估计值”写成“阈值电流”。

25.9 阶段 E：热-电-光迭代，不追求一轮算完，而追求工作窗口收敛

初次建模时，容易把多物理场耦合理解为“把所有接口同时打开，一次求解结束”。对 PCSEL 来说，这种理解往往既低效也不稳健。较实用的策略通常是分阶段迭代：光学给出模式和重叠，电学给出注入分布，热学给出温度场，再把这些结果回写到光学模型，重复直到目标指标收敛。

workflow 应检查的是工作窗口是否稳定，而非某一次迭代里的最佳单点。一个结构若只有在第一次冷腔迭代中表现良好，而一旦写回温升就发生模排序翻转，它就不能算成功设计。相反，一个方案即使冷腔 Q 不是最高，若热回写后仍保持目标模的净裕量为正，它在器件层面上更有价值。

25.10 阶段 F：鲁棒性验证，检查工作窗口是否同时具有工艺窗口

若阶段 A 到阶段 E 用于判断“名义器件在当前模型下可行”，那么阶段 F 要回答的是“这是不是一个能加工出来、能批量重复出来的器件方案”。也就是说，在名义工作窗口已经

建立之后，还应把 第二十二章 中的 tolerance analysis 显式纳入流程。最小动作包括：定义关键误差参数、先做局部灵敏度筛查、对高敏感参数做联合抽样，并用通过率而不是单一样本图来报告结果。

这个阶段之所以不能省略，是因为很多设计在名义值上成功，却在孔半径、刻蚀深度或孔径错位上异常脆弱。若不把阶段 F 写进 workflow，所谓“可行设计”往往只是在数值环境里可行，而不是在工艺环境里可行。

25.11 阶段 G：实验回写链，不让模型停在“可运行”而要走到“可校准”

可用于研究和器件开发的 workflow，还必须再多走一步：把模型接回实验。最小回写链应写成“测量 → 参数提取 → 分层校准 → 预测 → 回验”。这里的关键是把它当成模型可用性的约束，而非把实验当成结果展示。

具体做法通常是：先用截面、几何、电学和热学的直接测量值约束几何与边界条件；再用 I-V、L-I、热阻或谱漂移数据分别校准电学和热学子模型；最后才把这些中间结果回写给光学与增益模型，去预测未参与拟合的阈值漂移、远场变化或偏振窗口。只有当“未参与校准的观测量”也能被解释时，模型才具备受检验的外推能力。

这一阶段的更细致写法，会在 第二十九章 中系统展开。本章只强调一个纪律：不要把所有实验量都同时拿来拟合同一套自由参数。那样做容易得到“处处都像、处处都不可信”的过参数化模型。

25.12 阶段 H：代理模型与 AI 加速（新增）

2024–2026 的一个明显趋势是：PCSEL 设计正在把“快速近似模型”进一步扩展为“近似模型 + 代理模型 + 全波验收”的混合流程 [36, 53, 71]。这一步可降低参数扫描成本，但可靠性取决于是否保留物理验收门槛。

本书建议的最小门禁是：

1. 数据来源可追溯：训练集必须标注生成器（CWT/RCWA/FDTD）、参数范围与版本；
2. 外推边界可声明：仅在训练分布内做优化建议，超出分布必须回退到物理求解；
3. 关键结论可复核：所有准备进入发布级结论的点，必须用 RCWA/FDTD/FEM 复算；
4. 器件级回写链不跳步：代理模型不能替代电热回写与实验校准阶段。

代理模型应定位为筛选加速层，而不是真值层。它在流程中的合理位置是阶段 B/C 的前置筛选，而不是阶段 G 的替代。

25.13 Lumerical 与 COMSOL 的两条常见实现路径

前面分阶段讲的是物理链，下面把它压缩成两条常用工程路径，供读者实际启动项目时参考。需要强调，这两条路径都是**推荐流程**，不是唯一流程。

路径一：以 Lumerical 为主的光学优先链

这条路径适合“先把单胞和有限阵列光学迅速扫清，再把最有希望的结构接到器件级电热模型”这一工作风格。典型顺序是：

1. MODE/FDE：检查纵向模、 n_{eff} 、 Γ_{QW} 与损耗重叠；
2. RCWA：扫单胞、提取候选模、上下出光和偏振趋势；
3. FDTD：有限阵列验证、提取远场和边缘敏感性；
4. CHARGE：求注入分布和串联电阻来源；
5. HEAT：求温升和热热点；
6. 脚本化回写：把 $N(\mathbf{r})$ 与 $T(\mathbf{r})$ 回写给光学模型，更新模增益和损耗排序。

这条路径的特点是光学前端筛选较快，但多物理场耦合依赖明确的接口管理和脚本化回写。

路径二：以 COMSOL 为主的多物理场优先链

这条路径适合“在统一几何和 PDE 框架下做电-热-光之间的参数回写和灵敏度分析”这一工作风格。典型顺序是：

1. 二维或简化三维 Wave Optics：建立纵向模背景；
2. 三维单胞 Wave Optics：做周期单胞散射与偏振筛选；
3. 三维有限阵列 Wave Optics：做有限器件近场、远场与本征模验证；
4. Semiconductor：求注入和载流子分布；
5. Heat Transfer：求温度场与热热点；
6. 参数回写与分阶段迭代：更新材料参数、重跑光学、检查工作窗口。

这条路径的优点是多物理场接口统一、参数回写自然，缺点是大尺寸三维光学问题的计算成本往往更高，对网格和求解设置更敏感。

若研究目标是快速筛选 50 组单胞设计，再从中挑 2 组做器件级深入验证，通常更适合先走 Lumerical 光学优先链；若研究目标是深入分析一个给定外延与电极方案中，温升如何改变模式竞争，则更适合走 COMSOL 多物理场优先链。这里的“更适合”是工程判断，不是排他性规则。

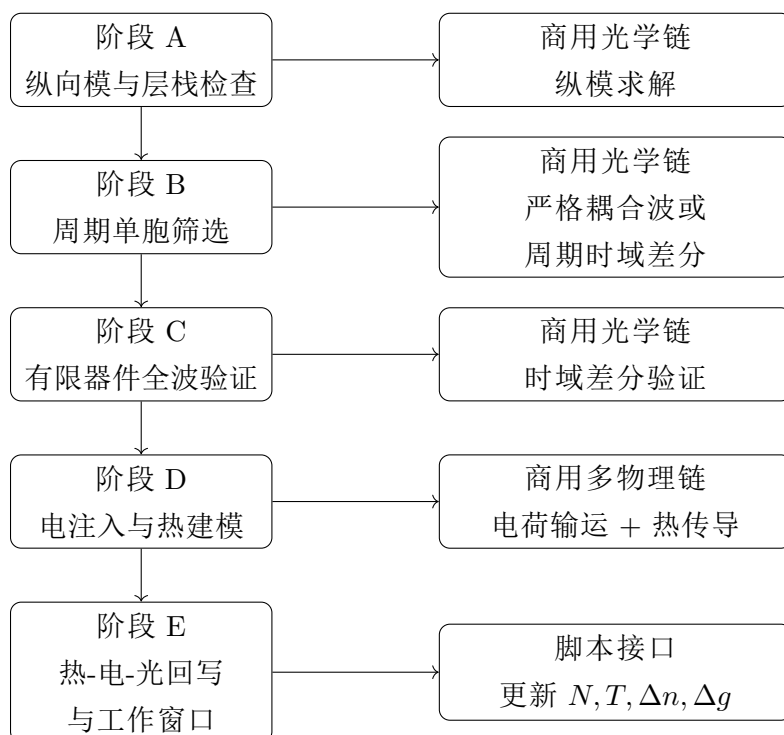


图 25.3: 光子晶体面发射激光器仿真 workflow 示意。图中给出商用光学链到多物理链的常见映射；不同软件路径遵循同一物理阶段划分。

25.14 何时停止迭代：停机判据比“再跑一轮看看”更重要

完整 workflow 容易陷入无止境试错，因此必须在开工前就设定停机判据。最有用的停机条件通常有三类：

1. 数值收敛停机：网格、时间窗、PML、傅里叶阶数继续加密时，关键输出变化已低于预设阈值；
2. 物理一致性停机：功率守恒、参数量纲、层级结论和趋势解释均已自洽；
3. 目标窗口停机：在目标工作区间内，目标模对竞争模的净裕量、远场和温升都满足要求。

如果这三条没有满足，即使结果图平滑，也不应进入最终结论。

易错点

“已经跑出了一个看起来不错的图，所以先写结论再说”是仿真 workflow 中最常见的失控点。对 PCSEL 来说，有价值的是一条从参数到账本、从候选模到工作窗口都可追溯的证据链，而非一张结果图。

本章到此给出的是正向流程骨架与停机判据。当流程走到需要把仿真数字与实验测量放在同一张桌上时，完整的仿真-实验对标协议（层级对齐五项前置检查、三级对标体系、关键物理量技术规范、不符时的诊断树与报告伦理）已收入 第二十九章 的证据链部分（第29.12节），建议在进入 第二十六章 的阶段 G 之前通读一遍。

回看本章

一套实用的 PCSEL 仿真 workflow 必须以结论分层和参数账本为起点，然后依次经历纵向模检查、周期单胞筛选、有限器件验证、电学与热学建模，以及热-电-光回写。Lumerical 与 COMSOL 都能承载这条链，但常见工程分工不同：前者通常更擅长高效光学筛选，后者通常更擅长统一多物理场耦合。无论选哪条路线，应追求的都是一个通过收敛和物理一致性检查的工作窗口，而非“某一轮最优图片”。若把本章和后面两章一起看，本章给出的是正向流程骨架，第二十六章 给出的是一次完整走法，第二十七章 给出的则是反向诊断顺序。

练习题

1. [基础] 为一个你感兴趣的 PCSEL 方案写出 L1 到 L4 的结论层级，并说明每一级允许说什么、不允许说什么。
2. [基础] 分别给出阶段 A 到阶段 E 的输入与输出变量，特别指出哪些量需要在不同求解器之间传递。
3. [进阶] 比较以 Lumerical 为主和以 COMSOL 为主的两条 workflow，说明它们各自最可能在哪一步最耗时、最容易出错。
4. [综合]（思路提示：先定层级，再列关键式）为你的项目设计三条停机判据，并解释为什么这些判据能防止你过早下结论。

延伸阅读

- 下一章将把本章 workflow 落实到完整示例，给出参数、建模顺序、求解设置模板和可信度检查。阅读时建议对照 表25.1，逐步核对每一阶段到底在例题中对应哪一步。
- 结合 第二十七章 阅读，可以把本章 workflow 直接转化为项目级排错流程；最有效的用法，是先按本章正向走，再按失败模式章反向复盘。

第二十六章 教材式例题：从玩具模型到完整建模框架

本章先解一道具体问题：

给定一套简化但合理的层栈与平面晶格参数，如何判断它是否存在值得继续设计的面发射候选模；若存在，这一判断依赖哪些中间输出，又还缺少哪些步骤才能升级为器件级结论？

读完本章，你能

- 通过一个双层例题，学会把 PCSEL 问题从“物理图像”推进到“可执行建模步骤”，而不只是记住 workflow 名词。
- 在纸上完成一个最小四波 CWT 玩具模型的模式分裂、损耗排序与阈值判断，从而理解为什么有些模虽然亮，却不一定先激射。
- 在数值层面走完整条链：纵向模、单胞候选模筛选、有限阵列检查、简化电学/热学回写，并学会判断每一步到底能说什么、还不能说什么。

读前准备 建议已掌握 第十章、第十三章、十七至二十、第二十二章和第二十五章 中关于最小 CWT、PWEM、完整 workflow、外延层、电注入、热回写与鲁棒性分析的内容。

本章结构

本章所处模型层级：L1→L4

图 26.1 先给出本章双层例题的输入/输出边界，后文每一步都按这张桥接图检查“已经证明了什么、还缺什么”。

前面的例题更像一份任务说明：它告诉读者应该先做什么、后做什么，但还没有把“为什么这样做”和“手上到底应该算出什么”讲到足够像教科书例题。本章因此重写成双层结构。第一层是纸上可走完的玩具模型，我们明确写出一个四波耦合矩阵，算出模式分裂、损耗排序与一个最小阈值判断。第二层是数值全流程例题，我们固定一组简化但合理的参数，逐步完成纵向模检查、单胞候选模筛选、有限阵列验证以及简化电热回写，并在最后汇总每一步的中间结果、可下的可靠结论与仍然悬而未决的问题。读者完成本章后，应能把“workflow”落实为一组可检查的输入、输出和判断边界，而不是停留在抽象流程图。

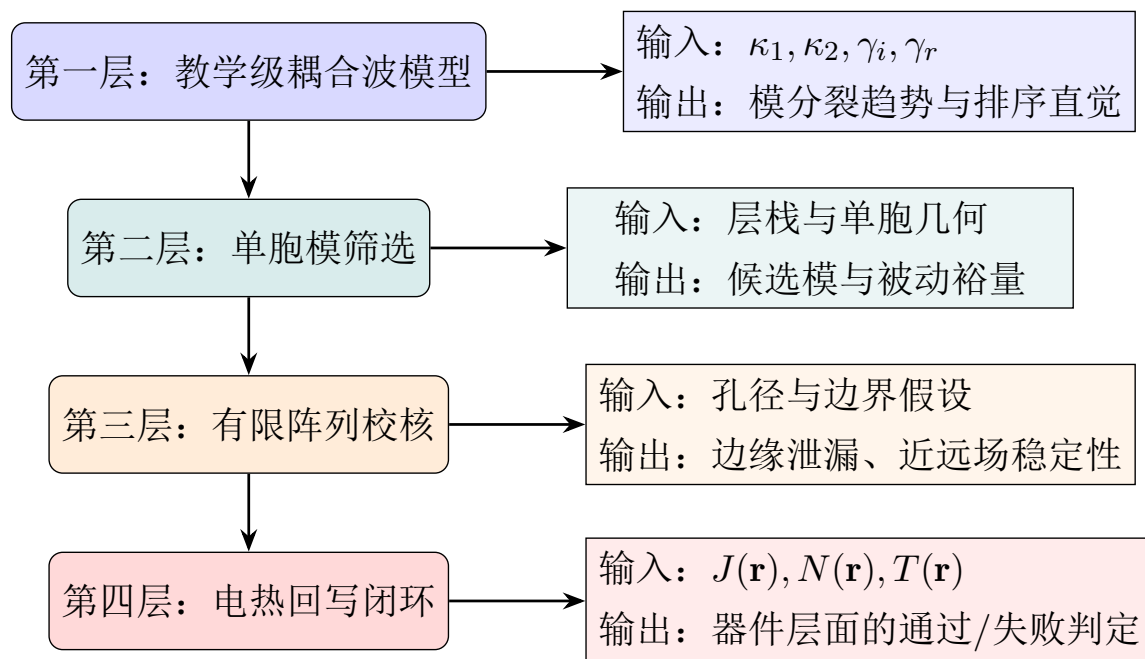


图 26.1: 算例桥接图：每一层都给出明确输入、输出与结论边界，读者可沿证据链追踪“本层到底证明了什么”。

这个题目研究一个教学型方格晶格 air-hole slab PCSEL。它看似保守，却同时要求回答四类问题：Gamma 点附近是否有合理候选模；有限阵列边界会不会推翻无限周期排序；电流注入和温升会不会改写模竞争；最后能汇报的究竟是“光学候选性”，还是“器件级可行性”。

直觉框

本章的目标是带你学会一种解题习惯：每往前走一步，都知道自己手里多了什么证据，而非带你“跑完一遍软件”；每做一个判断，都知道它还缺哪一块物理才能升级。

26.1 第一层例题：纸上可完成的四波 CWT 玩具模型

26.1.1 例题 A 的设定

先暂时不碰真实层栈和复杂几何，只抓住 Gamma 点附近最基本的四个面内基本波：

$$\mathbf{a} = \begin{bmatrix} R_x & S_x & R_y & S_y \end{bmatrix}^T,$$

其中 R_x, S_x 表示沿 $\pm x$ 方向传播的基本波， R_y, S_y 表示沿 $\pm y$ 方向传播的基本波。我们考虑一个理想方格晶格，假设：

1. 沿同一坐标轴的前后向散射强度相同，记为 κ_1 ；
2. x 向与 y 向波之间的交叉耦合相同，记为 κ_2 ；
3. 四个基本波共享同一个背景频率 ω_0 ；

4. 所有模式都具有相同的内部损耗 γ_i ;
5. 只有完全对称的亮模额外耦合到垂直辐射通道, 这一辐射损耗记为 γ_r 。

假设审查: 辐射矩阵的均匀性假设

式 (26.2) 中的辐射矩阵采用了一个简化假设: 四个基本波以相同强度耦合到垂直辐射通道, 即辐射矩阵元全部相等 ($\zeta_{ij} = \gamma_r/4$)。这对应于理想方格晶格且孔径远大于波长的极限情况。

更一般地, 辐射矩阵可写为

$$\mathbf{\Gamma}_{\text{rad}} = \begin{bmatrix} \zeta_{RR} & \zeta_{RS_x} & \zeta_{RS_y} & \cdots \\ \zeta_{S_x R} & \zeta_{S_x S_x} & \cdots & \cdots \\ \vdots & & \ddots & \end{bmatrix}, \quad (26.1)$$

其中矩阵元 ζ_{ij} 依赖于晶格形状、占空比和辐射方向图。对非理想结构 (如椭圆孔、矩形孔或存在倾斜入射), 辐射矩阵元可能不再均匀, 导致暗模也获得有限的辐射损耗, 甚至出现模式间的辐射耦合。本例题采用均匀辐射矩阵是为了突出对称性保护的物理图像, 实际器件建模需通过严格电磁仿真提取辐射矩阵。

在这些假设下, 最小非厄米有效矩阵可以写成

$$\tilde{H} = (\omega_0 - i\gamma_i) \mathbf{I} + \begin{bmatrix} 0 & \kappa_1 & \kappa_2 & \kappa_2 \\ \kappa_1 & 0 & \kappa_2 & \kappa_2 \\ \kappa_2 & \kappa_2 & 0 & \kappa_1 \\ \kappa_2 & \kappa_2 & \kappa_1 & 0 \end{bmatrix} - i\frac{\gamma_r}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \quad (26.2)$$

这里的第一项给出共同背景频率和内部损耗, 第二项表示面内布拉格散射, 第三项则表示只有“所有基本波同相叠加”的那一个组合会显著向垂直方向辐射。这个模型是故意简化的, 但它已经足以回答两个教材里最关键的问题: 模式为什么会分裂, 以及“亮”为什么不自动等于“先激射”。

26.1.2 第一步: 把矩阵换到对称性基底

在原始基底直接解式 (26.2) 并不困难, 但教学上更好的做法是先引入四个更贴近物理图像的线性组合:

$$\mathbf{u}_B = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_X = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \quad \mathbf{u}_Y = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \quad \mathbf{u}_D = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}. \quad (26.3)$$

其中下标 ‘B’ 可理解为最亮的全同相组合, ‘X’ 表示 x 与 y 两组波之间的相对反相, ‘Y, D’ 则属于在完美方格对称下仍简并的一对暗模组合。

把式 (26.2) 作用到这四个向量上, 可逐个得到本征值。先看亮模:

$$\begin{aligned}\tilde{H}\mathbf{u}_B &= (\omega_0 - i\gamma_i)\mathbf{u}_B + (\kappa_1 + 2\kappa_2)\mathbf{u}_B - i\gamma_r\mathbf{u}_B \\ &= (\omega_0 + \kappa_1 + 2\kappa_2 - i(\gamma_i + \gamma_r))\mathbf{u}_B.\end{aligned}\quad (26.4)$$

因此

$$\tilde{\omega}_B = \omega_0 + \kappa_1 + 2\kappa_2 - i(\gamma_i + \gamma_r). \quad (26.5)$$

对第二个组合 $\mathbf{u}_X$, 因为各分量求和为零, 辐射矩阵对它不起作用, 于是

$$\tilde{\omega}_X = \omega_0 + \kappa_1 - 2\kappa_2 - i\gamma_i. \quad (26.6)$$

再对 $\mathbf{u}_Y$ 与 $\mathbf{u}_D$ 计算, 可得到同样的本征值

$$\tilde{\omega}_Y = \tilde{\omega}_D = \omega_0 - \kappa_1 - i\gamma_i. \quad (26.7)$$

到这里, 第一层例题的主要模式分裂已经在纸上完成了。四个原始基本波在周期散射下重组为四个超模, 其中一个最亮, 一个与 x/y 组反相有关, 另外两个在完美方格晶格下保持简并。

26.1.3 第二步: 读出模式分裂和损耗排序

现在不要急着代数字, 而是先解释式 (26.5)至式 (26.7) 的物理意义。

第一, κ_1 控制的是同一轴前后向波之间的耦合, 因此它会同时抬高亮模和 X 模、压低 Y/D 这对模的频率。第二, κ_2 控制的是 x 轴与 y 轴两组波之间的耦合, 因此它主要负责把亮模和 X 模再拉开一个 $4\kappa_2$ 的频率间隔。第三, 辐射损耗 γ_r 只落在亮模上, 这意味着“最强法向辐射”与“最小损耗”不是同一件事。

若把小信号净增长率写成

$$\lambda_j = g_{\text{mod},j} - \gamma_j, \quad (26.8)$$

其中 $\gamma_j = -\text{Im}(\tilde{\omega}_j)$, 则亮模的光学线性阈值为

$$g_{\text{th},B} = \gamma_i + \gamma_r. \quad (26.9)$$

暗模 ($X/Y/D$) 的总损耗仅为 γ_i , 但它们垂直辐射耦合为零 ($\gamma_{r,\text{dark}} = 0$), 因此无法作为表面发射器件激射。它们代表边发射通道, 只能通过腔体边缘耦合输出。

这一步就是本例题最值得读者记住的第一结论: **面发射最亮的模, 不一定是最先跨过线性阈值的模**。若垂直辐射耦合太强, 亮模反而会因为额外辐射损耗而需要更高的阈值增益。对表面发射 PCSEL 而言, 正确的品质因数不是最小阈值, 而是**最大斜率效率**:

$$\eta_{\text{slope}} \propto \frac{\gamma_r}{\gamma_i + \gamma_r}. \quad (26.10)$$

暗模虽然表观阈值更低, 但由于 $\gamma_r = 0$, 其表面发射斜率效率为零, 因此不会在表面发射中竞争胜出。

26.1.4 第三步：在简并子空间中加入一个最小对称性破缺

完美方格晶格下， $\mathbf{u}_Y$ 与 $\mathbf{u}_D$ 保持简并。但真实器件里，轻微矩形畸变、双晶格位移或工艺各向异性都可能把这对简并打破。教材里最干净的写法，是直接在 $\{\mathbf{u}_Y, \mathbf{u}_D\}$ 这个二维子空间内写一个微扰矩阵：

$$\tilde{H}_{\text{deg}} = (\omega_0 - \kappa_1 - i\gamma_i) \mathbf{I} + \begin{bmatrix} \delta & \eta \\ \eta & -\delta \end{bmatrix}, \quad (26.11)$$

其中 δ 描述两个暗模对某种几何各向异性的不同响应， η 描述它们之间被该微扰重新混合的强度。于是新的本征值立刻写成

$$\tilde{\omega}_{\pm} = \omega_0 - \kappa_1 - i\gamma_i \pm \sqrt{\delta^2 + \eta^2}. \quad (26.12)$$

这个结果虽然简单，但给出了明确的物理信息：对称性破缺的第一作用，往往是在原来已经存在的简并子空间里把模式重新旋转并拉开，而非立刻“创建一个新模”。若这两个新模对辐射或增益的重叠不再相同，那么它们的损耗排序和偏振稳定性也会随之改变。这对应于第八章、第十二章中“偏振选择”和“模式分裂”在教材例题里的最小落地版本。

例子

取一组纯教学型无量纲数值： $\kappa_1 = 2$ ， $\kappa_2 = 1$ ， $\gamma_i = 0.5$ ， $\gamma_r = 1.5$ 。则

$$\tilde{\omega}_B = \omega_0 + 4 - i2.0, \quad \tilde{\omega}_X = \omega_0 - i0.5, \quad \tilde{\omega}_Y = \tilde{\omega}_D = \omega_0 - 2 - i0.5.$$

这说明亮模频率最高、法向辐射最强，但其线性阈值也最高。若材料增益峰更接近 $\omega_0 - 2$ 一侧，或者目标是最先起振而非最亮输出，那么最先获利的模很可能不是亮模。这个结论完全不依赖具体软件，是纸上就能读出的结构性事实。

26.1.5 第四步：这一层玩具模型到底回答了什么

到这里，例题 A 已经完成。它回答了四件实用的事。

第一，它告诉我们完美方格晶格里四个基本波如何重组为亮模和暗模。

第二，它告诉我们模式分裂受哪两类耦合主导：同轴前后向耦合 κ_1 和轴间交叉耦合 κ_2 。

第三，它告诉我们“亮”不等于“低阈值”，因为亮模通常要支付额外辐射损耗。

第四，它告诉我们对称性破缺首先在简并子空间里起作用，因此偏振选择和模式分裂可以被看成同一个二维微扰问题。

但这一层仍然有明确边界。它没有真实层栈，没有真实三维辐射，没有有限阵列边缘，也没有电流注入和温升。因此它是一个有效的纸上判断器，却还不是器件结论。

易错点

若读者在玩具模型这一步就宣布“某个亮模一定是器件工作模”，那就相当于把一个低维线性模型越级使用到了有限尺寸、有源、热反馈器件。这也是本书一直反对的跳步。

给玩具模型补一点数量级感 为了不让 $\kappa_1, \kappa_2, \gamma_i, \gamma_r$ 永远停留在抽象符号层，可以给它们配一个最小数量级直觉。若参考波长取 $\lambda_0 \sim 1 \mu\text{m}$ ，则参考角频率约为 $\omega_0 = 2\pi c/\lambda_0 \sim 1.9 \times 10^{15} \text{ rad/s}$ 。若某组参数对应相对模分裂

$$\frac{\Delta\omega}{\omega_0} \sim 10^{-4},$$

则大致对应 0.1 nm 量级的波长分裂；若某模满足

$$\frac{\gamma}{\omega_0} \sim 10^{-5},$$

则对应的被动品质因子约为 $Q \sim \omega_0/(2\gamma) \sim 5 \times 10^4$ 。这种换算的作用，是让读者知道玩具模型里的“一个小参数”落回实验量级后大概代表多大的频移和多大的被动损耗。

26.2 第二层例题：数值全流程全解

26.2.1 例题 B 的设置与已知量

现在进入软件建模例题。我们仍然使用方格晶格 air-hole slab PCSEL，但不再停留在抽象四波基底，而是固定一组简化但合理的名义参数：

$$a = 280 \text{ nm}, \quad r = 0.30a, \quad t_{\text{pc}} = 160 \text{ nm}, \quad t_{\text{wg}} = 260 \text{ nm}, \quad (26.13)$$

$$t_u = 1.2 \mu\text{m}, \quad t_d = 1.5 \mu\text{m}, \quad N_x = N_y = 31, \quad z_{\text{QW}} \approx 0. \quad (26.14)$$

它们是一个教学用名义设计点，而非校准后的器件参数。我们的任务不是“报出最终性能”，而是回答：

在这组名义参数下，是否存在值得继续推进的面发射候选模；如果要把这个候选模升级为器件级结论，还必须补哪些证据？

26.2.2 解题前先写定性预期

教材式例题不应从“开软件”开始，而应从“先写预期”开始。对本例，我们先明确四条定性预期。

第一，纵向层栈应支持一个与量子阱区有较好重叠的主要导模，否则平面晶格结果再好也没有意义。

第二，单胞层面应出现 Gamma 点附近的面发射候选模，而且目标模与竞争模在辐射排序上至少要有可解释差异。

第三，有限阵列边界可能改变单胞下看到的排序，因此需要重新审查主模、边缘泄漏和远场。

第四，哪怕纯光学阶段一切良好，电流扩展和热回写仍可能改变目标模的净模裕量。

这四条预期的作用是给后面的每一步都设立“要验证的命题”。这样处理后，仿真就是逐题作答，而非盲扫。

表 26.1: 例题 B 的名义参数。所有数值仅用于教材推演，不应被视为某一已校准器件的真实数据库。为避免误读，表中新增“来源/用途”列。

参数	名义值	来源/用途	在例题中的作用
晶格常数 a	280 nm	[本书算例] 教学名义点	设定平面周期和归一化频率标尺
孔半径 r	$0.30a$	[本书算例] 教学名义点（与随书 PWEM 脚本一致）	控制单胞傅里叶分量、耦合和辐射趋势
PC 层厚度 t_{pc}	160 nm	[本书算例] 教学名义点	影响面内散射和纵向剖面
波导总厚度 t_{wg}	260 nm	[本书算例] 教学名义点	决定纵向模与量子阱重叠
上/下包层厚度 t_u, t_d	$1.2\mu\text{m}, 1.5\mu\text{m}$	[本书算例] 教学名义点	决定上下辐射不对称与导模背景
阵列尺寸 $N_x \times N_y$	31×31	[本书算例] 起始尺寸（需收敛复核）	给出有限阵列验证的起点尺寸
量子阱位置 z_{QW}	波导中心附近	[设计意图] 提升重叠	使目标纵向模与有源区尽量重合
电极/掺杂设置	上 p、下 n、接触层重掺杂	[工程经验] 最小器件背景	为后续电流扩展与热源建模提供最小器件背景

26.2.3 第一步：纵向模检查

题目 在给定层栈下，判断是否存在一个可作为后续 PCSEL 平面建模背景的主要纵向导模。

应建的最小模型 在 Lumerical 路径中，用 MODE/FDE 建立二维横截面层栈；在 COMSOL 路径中，用二维横截面的 Wave Optics 模态模型。无论哪条路线，都要把问题简化成同一个物理任务：求纵向导模、有效折射率和与有源区/损耗层的重叠。

应得到的输出

1. 有效折射率 n_{eff} ;
2. 纵向场分布 $E(z)$ 或等效模场剖面;
3. 量子阱重叠 Γ_{QW} ;
4. 重掺杂层与金属邻近损耗重叠 Γ_{loss} ;
5. 对 t_{wg} 、 t_{pc} 和 z_{QW} 的局部灵敏度。

这一步现在不再只是“应当求一个纵向模”的任务清单。脚本 `code/vertical_tmm_mode/run_vertical_tmm_mode.py` 的 CSV 保存了 $n(z)$ 与归一化场剖面，JSON 保存了层栈、全部束缚模根和重叠积分。模型仍是无损、标量、一维代理，因此不能给出金属损耗或全矢量偏振混合；它的职责是为后续二维平面模型提供一个可复核的 n_{eff} 与 Γ_{QW} 起点。

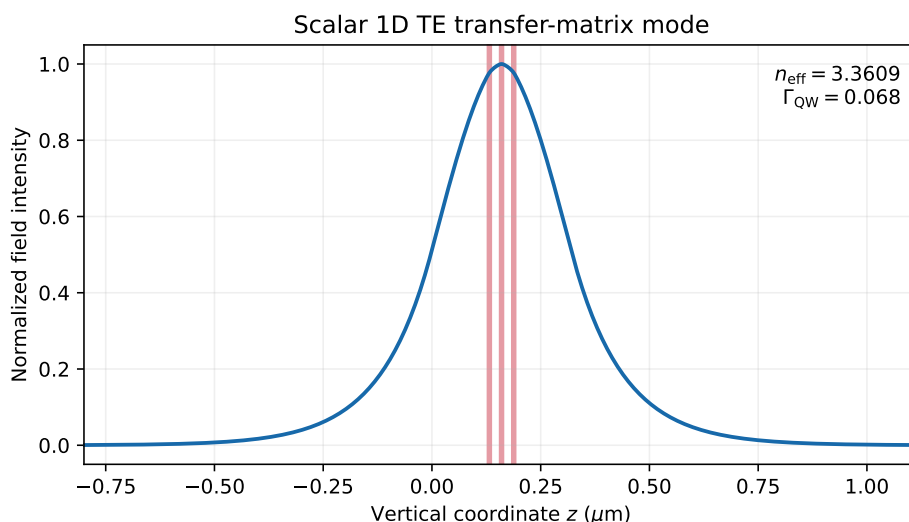


图 26.2: 例题 B 的可复现一维纵向模。随书脚本对 940 nm、总厚度 320 nm 的对称弱导层栈求解标量 TE 传输矩阵本征条件, 得到 $n_{\text{eff}} = 3.3609$ 、三个 8 nm 量子阱的能量加权重叠 $\Gamma_{\text{QW}} = 0.068$ 。红色窄带标出量子阱位置。

教材式判断 若 Γ_{QW} 本来就很小, 说明纵向模式背景设置不合理, 此时不应继续扫孔半径。若 n_{eff} 对层厚极端敏感, 说明后续平面设计会非常依赖工艺厚度控制。若 Γ_{loss} 很大, 说明电学低电阻与光学低损耗之间的冲突已经很强, 后面必须在 第十七章至第二十章 的器件主线上重做权衡。

这一阶段能说什么 可以说“存在一个合理的纵向候选导模”或“不存在”。不能说“器件阈值已经低”。

26.2.4 第二步：单胞候选模筛选

题目 在名义单胞参数下, 判断 Gamma 点附近是否有值得继续投入有限阵列验证的面发射候选模。

应建的最小模型 这一阶段建议分成两个子步骤。

1. 先用 第十三章 的 PWEM 框架做被动无限周期候选模筛选, 必要时可直接调用本项目脚本 `code/pwem_square_hole/run_pwem_square_hole.py` 作为正方晶格圆孔的第一轮筛选工具; 但教学上不应只停在“脚本完成运行”, 而应把输出与 图13.4和表13.2 对照, 确认本书示例中的 Gamma 点候选频带大致落在 $\Omega \approx 0.34$ 到 0.46, 其中典型高对称点值包括 0.3402, 0.4074, 0.4074, 0.4611;
2. 再用 RCWA 或 COMSOL 的三维周期频域单胞模型, 检查开放辐射、上下出光比例和偏振趋势。

应得到的输出

1. Gamma 点附近候选频带所在频段；
2. 候选模的主要对称性和主导傅里叶分量；
3. 上/下辐射比例；
4. 目标模与竞争模的被动损耗排序；
5. 对孔半径和刻蚀深度的局部灵敏度。

教材式判断 若 PWEM 只显示相关频段里没有合适候选模，则无需进入开放单胞。若 PWEM 给出候选模，而 RCWA 显示目标模在开放结构中辐射排序并不占优，则应回到几何与层栈参数重新调整。若单胞结果显示目标模确实有合理的辐射和偏振趋势，那么它才获得进入有限阵列审判的资格。

最小定量筛选门槛 为了把“教材式判断”落实成第一次可执行的筛选动作，至少还应加三条定量门槛。第一，候选模频率应落在目标材料增益峰附近的可接受失谐窗口内，而不是只凭“在同一个大频段里”；第二，目标模相对于最危险竞争模的被动损耗排序，应在小幅孔径、刻蚀深度或层厚扰动下保持不翻转；第三，上下辐射比例和偏振标签应能用前文的对称性与辐射通道语言解释，否则即便数值结果好看，也要先怀疑模式身份是否认对了。

这一阶段能说什么 可以说“该结构在无限周期开放单胞下存在合理候选模”。不能说“有限器件一定单模”。

26.2.5 第三步：有限阵列检查

题目 把单胞里通过初筛的候选模放进有限器件里，检查边缘是否改写模式排序和远场。

应建的最小模型 Lumerical 路径中最自然的是三维 FDTD；COMSOL 路径中最自然的是三维 Wave Optics 频域或本征频率研究，并辅以 PML。 31×31 阵列在本章是 [本书算例] 起点，不是通用最优尺寸；必须至少对邻近尺寸做一次收敛与边缘敏感性检查。

报告规范

无限周期结果转有限阵列时的最小报告四要素

1. **阵列尺寸**：至少报告 3 个横向尺寸样本，并明确主模/竞争模排序是否翻转。
2. **边界模型**：明确边缘截断方案、PML/端口设置与粗糙/无序代理模型。
3. **收敛曲线**：至少给出 1 组关键量（如 Q 、主瓣角、 ΔG_{*j} ）随尺寸或网格的收敛趋势。
4. **误差样本**：给出最坏样本或通过率统计，而不只给名义参数点。

应得到的输出

1. 有限阵列主模与竞争模的排序;
2. 近场分布与边缘泄漏位置;
3. 远场主瓣、旁瓣和偏振趋势;
4. 阵列尺寸、PML 和网格变化后的稳定性。

教材式判断 如果目标模在单胞中很好,但在有限阵列中出现强边缘泄漏或排序反转,那么单胞结论已经失效。此时最合理的动作是先承认“无限周期不等于有限器件”这一关还没过,而非急着引入增益。若有限阵列中目标模仍占优,而且远场和偏振保持合理,则纯光学证据链进一步加强。

这一阶段能说什么 可以说“在给定名义阵列尺寸下,纯光学有限器件支持某个优势候选模”。不能说“器件阈值电流已知”。

26.2.6 第四步:简化电学与热学回写

题目 判断在真实电极、注入与温升参与后,纯光学候选模是否仍保持净优势。

应建的最小模型 Lumerical 路径中,用 CHARGE 解简化漂移-扩散与泊松问题、用 HEAT 解温度场;COMSOL 路径中,用 Semiconductor 与 Heat Transfer 完成同样任务。这里不追求一上来就做全耦合,而是先做一个简化回写代理:

1. 用简化截面或对称几何得到 $J(\mathbf{r})$ 、 $N(\mathbf{r})$;
2. 由 Joule 热和非辐射复合得到 $T(\mathbf{r})$;
3. 把 N 与 T 映射成 $\Delta g(\mathbf{r})$ 、 $\Delta n(\mathbf{r})$ 和 $\Delta \alpha(\mathbf{r})$;
4. 再用这些空间分布回写光学模型,比较目标模和竞争模的净模裕量。

应得到的输出

1. 电流密度分布 $J(\mathbf{r})$;
2. 载流子分布 $N(\mathbf{r})$;
3. 温度分布 $T(\mathbf{r})$ 与热点位置;
4. 回写后的目标模净裕量变化趋势;
5. 目标模与竞争模是否发生排序翻转。

教材式判断 若 $J(\mathbf{r})$ 与目标模主场区域明显错位, 说明“光学模场合理”不等于“器件注入合理”。若 $T(\mathbf{r})$ 在边缘或金属邻近区形成热点, 而回写后竞争模获益更多, 则说明纯光学阶段的优势还不能升级为器件结论。只有当简化回写后目标模仍保持足够净裕量, 才可以开始谈器件级可行性的初步证据。

这一阶段能说什么 可以说“在该简化电热代理模型下, 目标模在工作窗口内仍保持/不保持优势”。仍然不能说“绝对阈值电流、线宽和功率已经被完全预测”。

26.2.7 给例题 B 补一页可核对的“实算结果页”

前面各步已经说明了“应该产出什么”。为了让本章不只像执行模板, 而更像可跟做的教材例题, 这里把本题已经固定下来的教学值、脚本实算值和验收窗口集中成一张结果页。需要强调, 下面的数字分三类: 一类来自本项目已附带的 PWEM 脚本实算; 一类来自本章名义层栈下的教学代理值; 还有一类是有限阵列与电热回写阶段的验收窗口, 而不是高保真器件终值。

报告规范

本章算例 provenance (可复核元数据)

- **PWEM:** code/pwem_square_hole/run_pwem_square_hole.py; 输入来自脚本内 PARAMS 字典 ($r/a = 0.30$, $\varepsilon_b = 3.4^2$, cutoff=2, bands=6, points_per_segment=16); 输出为 summary JSON 与能带 CSV。
- **纵向 TMM:** code/vertical_tmm_mode/run_vertical_tmm_mode.py; 输入为 940 nm、 $n_{\text{clad}} = 3.30$ 、320 nm 对称层栈与三个 8 nm 量子阱; 输出为场剖面 CSV、summary JSON 与 PDF 图。
- **PWEM 文件哈希 (SHA256)** : 脚本 31515F62...30FF7; summary 1045867A...7A266; csv 4836CB7E...A38D2
- **口径说明:** 上述实算值属于 [本书算例], 用于教学复核与数量级锚定; 若脚本或参数改动, 需同步更新结果页窗口和可追溯元数据。

若把表26.2与前文四个阶段对照起来, 可以看到本章现在至少已经给出了两类对新手最关键的信息。第一类是**真的可核对的数字**, 例如 PWEM 脚本在高对称点输出的归一化频率; 第二类是**一旦偏离就应该立刻停下排查的窗口**, 例如 Γ_{QW} 太小、 Γ_{loss} 过大、有限阵列结果对尺寸/PML 过敏等。对第一次独立上手的读者来说, 这比只知道“要算哪些量”更有用, 因为它把“算对了大概长什么样”也写出来了。

表26.2故意没有给出绝对阈值电流、绝对线宽或高功率热滚降数字。原因是因为这些量若没有完整增益标定、接触参数、热边界与实验回验, 就不应被伪装成“本题已经算完的终值”, 而非本章回避结果。本章给出的, 是一张**足够教材化的中间结果页**, 而不是一份冒充器件定量预测的表。

表 26.2: 例题 B 的“实算结果页”。这里把脚本实算、教学代理值和验收窗口并排列出，目的是给第一次独立做题的读者提供数量级与判错锚点，而不是伪造完整器件终值。

中间量	本题采用/得到的数值	来源/身份	读者应如何使用
n_{eff}	3.3609	一维标量 TE-TMM 脚本实算值	作为后续二维平面筛选的有效折射率起点，并用全矢量层栈模型复核
Γ_{QW}	0.068	TMM 能量加权重叠实算值	说明三个量子阱位于纵向主场区；更换层厚或阱位后必须重算
Γ_{loss}	$\approx 0.005\text{--}0.02$	[本书算例] 名义损耗重叠窗口	若与 Γ_{QW} 同量级甚至更大，后续器件阈值通常会明显吃亏
$\Omega_{\text{T}}^{(2,3,4,5)}$	0.3402, 0.4074, 0.4074, 0.4611	PWEM 脚本实算值	用于锁定单胞筛选频窗，并核对高对称点简并是否正确
单胞候选频窗	$\Omega \approx 0.34\text{--}0.46$	由 PWEM 实算反推的 RCWA 频窗	不再盲扫全频段，而是围绕候选模族做开放单胞精修
有限阵列起始尺寸	31×31	本题显式给定	至少再与相邻尺寸做一次主模排序与远场收敛比较
有限阵列验收	目标模近场主峰居中、边缘泄漏不主导；远场主瓣保持中心集中特征	教学验收窗口	若一换尺寸/PML/网格就翻转，说明当前还不能进入电热回写
电热回写验收	$J(\mathbf{r})$ 与目标模高场区不严重错位； $T(\mathbf{r})$ 热点不把竞争模推成净裕量更大	教学验收窗口	这里建立的是“初步器件可行性”，不是绝对 I_{th} 与线宽

26.2.8 第五步：把中间结果收束成一张教材式汇总表

到了这里，最重要的教学动作是学会把中间结果整理成一张读者可以复用的判断表，而非继续堆更多输出。表26.3 给出这道题的答案格式。如果把本章当成对 第二十五章 的一次完整执行，那么这张表就是 workflow 阶段 A–G 在一个具体算例中的投影；若再与 表25.1和表26.4 对照阅读，就能立刻看出每一条中间量在失败诊断里对应哪一类风险。

表 26.3: 例题 B 的中间结果汇总表。表中是规定每一步必须产出的输出类型、可做的判断和结论边界，而非伪造终值。

步骤	必须产出的中间量	现在可以做的判断	现在还不能做的判断
纵向模	n_{eff} 、 Γ_{QW} 、 Γ_{loss} 、层厚 灵敏度	外延层是否提供了合理光学模式背景	阈值电流、器件功率
单胞筛选	Gamma 点候选模、被 动辐射排序、上下出光 比例、偏振趋势	是否存在值得继续推进的候选模	有限器件一定单模
有限阵列	主模排序、近场、远场、 边缘敏感性、尺寸收敛	无限周期结论是否在有限器件中仍成立	电注入后的真实工作模
电热回写	$J(\mathbf{r})$ 、 $N(\mathbf{r})$ 、 $T(\mathbf{r})$ 、 Δg 、 Δn 、净裕量变化	纯光学候选模在简化器件代理下是否 仍占优	绝对阈值电流、线宽、老化寿命

26.3 这一章到底给出了哪些“可靠结论”

教材式例题的最后一步，是把结论分层说清，而非再出一张图。

在本章范围内可以可靠建立的判断

- 通过玩具模型，可以可靠理解四个基本波如何分裂成亮模与暗模，以及为什么亮模不一定最先跨过线性阈值。
- 通过纵向模与单胞筛选，可以可靠判断名义层栈是否提供了合理的导模背景，以及 Gamma 点附近是否存在值得继续投入的候选模。
- 通过有限阵列检查，可以可靠判断单胞下的排序是否在给定尺寸范围内仍成立，远场与边缘泄漏是否已经出现明显问题。
- 通过简化电热回写，可以可靠判断“纯光学优势是否在最小器件代理下仍然保留”。

还不能被本章完全关闭的判断

- 绝对阈值电流仍需要经过材料增益、接触电阻、真实电流扩展和热阻校准。
- 连续波高功率工作时的热滚降、模式跳变和封装相关效应仍需要更高保真度模型。
- 线宽和相干性仍需要噪声模型，而不是仅靠确定性全波或多物理场解。

- 大批量工艺可制造性仍需要 第二十二章 的 tolerance analysis 和 robustness window 报告。

哪些问题最终必须交给实验回写链

- 材料增益模型是否与真实外延片一致；
- 金属接触、电流扩展层和热边界条件的真实参数；
- 表面粗糙、孔径偏差和阵列边缘截断的统计分布；
- 实际工作偏振、远场和阈值电流是否与模型趋势相符。

这并不意味着仿真不重要。恰恰相反，仿真的价值在于把实验回写链的问题压缩成少数清楚、可检验的命题，而不是让实验在一团没有层次的参数空间里摸索。

本章最应该保留下来的纪律是：纯光学结果只能建立“候选性”，简化电热回写只能建立“初步器件可行性”，器件级定量结论仍需要更高保真度模型和实验回写链。这是对证据层级的严格区分，而非保守。

26.4 把例题映射回 workflow 与失败模式

完成计算后，再把答案放回全书框架：第二十五章 给出正向项目骨架，第二十七章 给出反向诊断骨架，本章位于两者之间。表26.4 适合在做完例题后使用，用来核对每一步增加了哪类证据，以及走错时最可能落入哪种误判。

表 26.4: 本章双层例题在 workflow 与失败模式框架中的对应关系。

本章步骤	对应 workflow 阶段	本章要产出的核心证据	若走错，最容易落入的失败模式
玩具模型：四波 CWT 纸上分析	A/B 之前的概念压缩	模式分裂、辐射排序与最小阈值判断的物理图像	把亮模直接等同于最低阈值模；把最小模型结论过度外推到真实器件
数值第一步：纵向模检查	A	n_{eff} 、 Γ_{QW} 、 Γ_{loss} 与层栈敏感性	把不合理层栈带入后续所有光学步骤；误用 effective index
数值第二步：单胞候选模筛选	B/C	Gamma 点候选模、被动辐射排序、偏振趋势、上下辐射不对称	把被动单胞排序直接读成器件排序；把被动共振当有源阈值
数值第三步：有限阵列检查	D	有限尺寸下的主模排序、近场、远场与边缘敏感性	把无限周期结论直接外推到有限器件；阵列尺度未收敛
数值第四步：简化电热回写	E/F	$J(\mathbf{r})$ 、 $N(\mathbf{r})$ 、 $T(\mathbf{r})$ 与目标模净裕量变化	把均匀增益当真实注入；忽略热点导致的模式翻转
最终汇总与结论分层	G	哪些判断已可靠、哪些仍需实验回写链	越级报告阈值电流、线宽或长期稳定性

回看本章

这一章把例题从“任务清单”升级成了“双层例题”。第一层玩具模型在纸上完成了四波 CWT 的模式分裂、损耗排序和简并破缺，因此读者能够理解 Gamma 点候选模为什么会亮、为什么会分裂、为什么亮模不一定最先激射。第二层数值全流程例题则把同一个问题依次推进到纵向模、单胞筛选、有限阵列验证和简化电热回写，并在每一步明确输出类型、判断边界和未决问题。掌握这一章的标志，是学会：每做一步计算，都知道它增加了哪一层证据，而非记住某一组数字；每下一个结论，都知道它还缺哪一层证据。若把 第二十五章 视为正向项目骨架，那么本章就是一次完整执行；若再把 第二十七章 视为反向审计骨架，那么本章的每一步都应能够被逐项回查。

练习题

1. [基础] 在玩具模型中，若把辐射损耗矩阵改成同时作用于 u_B 和 u_X ，请重新写出本征值，并解释这对应了怎样的辐射物理图像。
2. [基础] 设 κ_2 从正变负，讨论亮模与 X 模的频率排序如何改变，并说明这意味着哪一类几何傅里叶分量可能发生了主导变化。
3. [进阶] 按本章的格式，为你自己的器件写一张“中间结果汇总表”，并明确指出每一步不能越级宣称什么。
4. [综合]（思路提示：先定层级，再列关键式）说明为什么本章即便完成了简化电热回写，也仍不能在没有实验校准的情况下报告绝对阈值电流和线宽。

延伸阅读

- 先回到 第二十五章，把 表25.1和表26.4 对照阅读，可把本章例题逐步映射回完整项目工作流，理解玩具模型、全波仿真和器件模型各自在流程中的职责。
- 再结合 第二十二章、第二十七章 阅读，可把本章例题继续扩展成“名义设计 + tolerance analysis + 失败模式诊断”的完整项目模板，并用失败模式章对本章每一阶段做反向自检。

第二十七章 PCSEL 建模常见失败模式

读完本章，你能

- 学会按结论层级、物理模型和数值实现三层顺序诊断 PCSEL 建模失败。
- 理解常见错误为什么会在 PCSEL 中尤其致命，而不是把它们当作普通求解器小毛病。
- 掌握在 Lumerical 与 COMSOL 工作流中最常见的误判路径和优先修复动作。

读前准备 建议已掌握 第十六章、第二十二章、第二十五章和第二十六章 中关于工作流、鲁棒性分析、方法分工和示例执行链的内容。

本章结构

本章所处模型层级：L1→L4

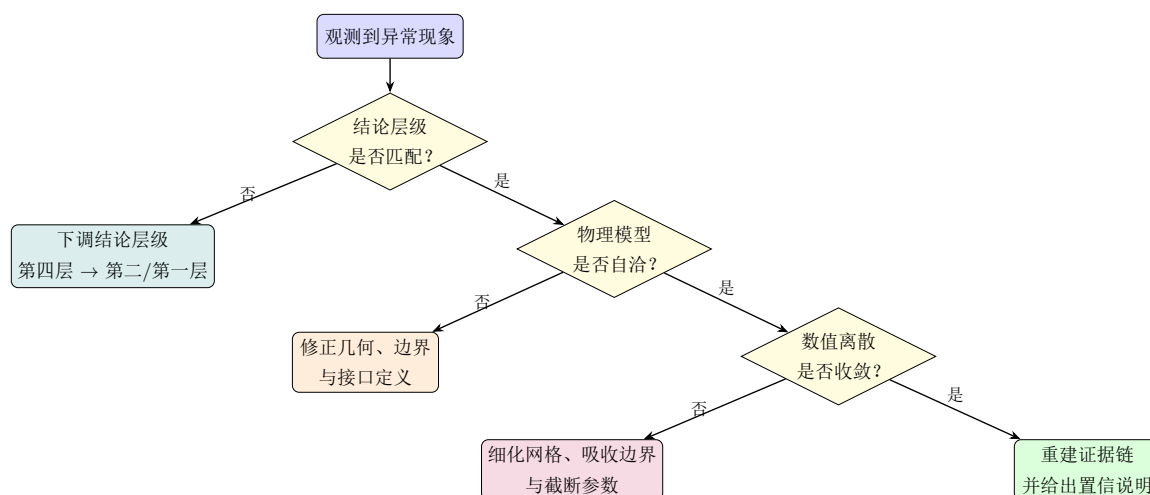


图 27.1: 本章建模排错流程：按“结论层级匹配 → 物理自洽 → 数值收敛”顺序逐级定位问题。

建议先按 图27.1 的顺序定位问题层级，再展开各节细节；这样可以把排错从“盲调参数”转成“证据链审计”。

排错是建模能力本身的一部分，而非附录工作。本章不按“软件功能表”写，而按“证据链在哪里断了”来写。我们先说明为什么 PCSEL 的错误往往先发生在结论层级而不是求解

器界面操作上，再按冷腔误判、边界条件误判、有效折射率误用、均匀增益误用、有限尺寸不足、鲁棒性分析缺席、热漂移忽略和单一求解器过度解读几个主类展开。随后再把这些错误映射到 Lumerical 与 COMSOL 中最常见的症状，给出优先排错顺序。

27.1 把本章当作工作流和例题的反向读法

第二十五章 的主要任务是告诉读者“正确顺序该怎么走”，第二十六章 则按这条顺序给出示例。本章应当被当作这两章的反向读法：当项目出现异常时，是先问自己，问题究竟出在工作流的哪一段，以及若把自己的结果放回 表25.1、表26.3和表26.4，最先对不上的那一栏是什么，而非重新从头盲目扫参数。这样做的好处是，排错会从“软件经验主义”转成“证据链审计”。

从实践角度看，一个成熟的项目通常会同时保存三张表： workflow 章节的阶段映射表、本章例题的中间结果汇总表，以及本章的失败症状表。前两张表告诉你项目应该长成什么样，第三张表告诉你它偏离时会以什么症状表现出来。三者同时保留，才能使排错过程具备可追溯的输入、输出和修正依据。

27.2 PCSEL 中最需要警惕的失败，不是算不出来，而是错误结果被误认为可靠结论

很多仿真项目的失败是显性的，例如求解器不收敛、网格内存爆掉、时域仿真提前终止。这类问题虽然麻烦，但至少容易察觉。PCSEL 建模中更危险的失败是隐性的：求解器给出了平滑曲线和平滑场图，但结论的层级已经错了，或者近似的边界已被悄悄越过。这样的结果最容易进入汇报和论文，因为它表面上具备合理性。

这也是为什么本章首先强调诊断顺序。遇到问题时，先不要立刻怀疑软件，也不要立刻去调求解器参数，而应先问：我现在试图下的结论，是否超出了当前模型本身的适用层级。如果这个问题没过，后续数值调参只会改善错误结论的表面一致性。

直觉框

PCSEL 建模里最昂贵的错误，通常是拿着一条层级错误的证据链往下越走越远，而非少跑了几轮仿真。先澄清“当前模型能证明什么”，通常比先纠正网格更值钱。

27.3 第一类失败：把冷腔结果直接读成器件结论

这是全书最反复强调、也是实际中最常见的失败模式。症状通常包括：用被动 Q 直接宣布低阈值；用单胞反射谱最窄特征直接宣布器件最低阈值模式；用有限阵列的近场图直接宣布阈值电流或线宽优越。

根因很明确：冷腔结果只包含开放电磁问题，而器件结论至少还需要材料增益、注入分布和温度场。若这些变量还没有进入模型，所谓“阈值电流更低”“高温更稳定”“线宽更窄”都只是推测，不是结论。

修复顺序也应当明确：先把结论降级到 L1 或 L2，即“候选模更优”“光学阈值要求更低”“有限阵列远场更理想”；随后再决定是否值得进入电学和热学阶段。在研究报告中，结论降级并非退让，而是对证据层级的明确标注。

若模型中尚不存在 $N(\mathbf{r})$ 、 $T(\mathbf{r})$ 或与工作点有关的增益/吸收回写，就不应输出器件级阈值电流、工作线宽窗口或高温稳定性结论。这是对模型适用范围的诚实描述，而非保守。

27.4 第二类失败：边界条件和几何定义在不同求解器里并不一致

这类失败尤其容易出现在跨软件或跨阶段工作流中。比如，单胞 RCWA 使用的是严格周期边界，有限阵列 FDTD 却因为复制方式或阵列边界定义不当，使真实几何与单胞假设不一致；又或者 COMSOL 的单胞频域模型采用了某一相位周期条件，而后续有限器件模型却没有保持同样的坐标与材料定义。

症状通常表现为：不同求解器得到完全不同的频率排序；网格一加密结果反而跳来跳去；远场主瓣方向和偏振在两个模型中无法对上。遇到这种症状，最常见的误判是以为“某个求解器不可靠”。更常见的真实原因，其实是两个模型解的根本不是同一个物理问题。

修复这类问题时，最有效的做法是先退回到一个最小对照算例：只保留一层、一个孔、一个清楚的周期方向和一组共享参数，然后逐步确认边界、坐标、单位和材料定义完全一致，再把复杂层栈加回来，而非在完整模型上继续试错。

27.5 第三类失败：把有效折射率近似当成严格三维模型

有效折射率近似是前期筛选的好工具，但它极易被误用成“已经足够描述器件”。症状通常是：二维模型预测的频率、偏振和远场都很好，但一进入三维验证就发生明显偏移；或者二维模型显示目标模有极佳重叠，三维模型却发现模式被上接触层或金属邻近区强烈拉偏。

根因在于有效折射率只压缩了纵向模的部分影响，并不能保留全部上下辐射不对称、层间损耗差异和强矢量效应。在 PCSEL 中，只要纵向剖面显著参与模式选择，这个近似就必须被当作代理模型，而不是最终结构模型。

修复方法也很直接：把二维模型的角色重新定义为“快速筛选器”，只让它决定参数方向，不让它独立承担最终结论；一旦候选结构进入关键排序比较，就必须回到三维矢量模型检查结果是否仍成立。

27.6 第四类失败：把均匀增益默认当作真实电注入分布

这是从光学走向器件时最危险的习惯性错误之一。很多初学者会在 RCWA、FDTD 或 Wave Optics 模型里直接给有源层加一个均匀增益，并据此默认自己已经“考虑了器件激射”。实际上，这最多只是小信号线性框架下的均匀增益代理问题。真实电注入器件中的增益分布取决于电流扩展、电流拥挤、漏电、串联电阻和温升，而这些都不在均匀增益假设里。

症状通常是：纯光学模型预测某个目标模最有利，但一旦把电流路径和热源加入，热点和高注入区却恰好偏向竞争模。此时若继续坚持均匀增益模型，就会把问题误判成“实验偏差太

大”。修复时应尽快引入 CHARGE 或 Semiconductor 一类电学模型，把均匀增益的假设显式替换成 $N(\mathbf{r})$ 诱导的空间分布。

27.7 第五类失败：有限尺寸建模不足，导致把单胞排序误读成器件排序

PCSEL 的很多关键特性都和有限孔阵有关，例如边缘泄漏、横向包络、主模尺寸和远场主瓣形成。若模型停留在单胞或极小阵列，就容易高估无限周期结论对真实器件的适用性。症状包括：单胞预测模式很理想，但实际有限阵列的远场副瓣很强；或者当阵列尺寸从 21×21 改到 31×31 时，目标模排序居然翻转。

修复这类问题时，最重要的是先做尺寸收敛曲线，而非立即追求更大阵列。也就是说，至少把阵列尺寸作为一个显式自变量来扫，观察目标模与竞争模的排序、远场主瓣和边缘损耗是否收敛。若连这个趋势都不稳定，就没有资格把结果上升为器件级结论。

对执行层面而言，更稳妥的最小判据是：至少连续比较三个阵列尺寸，并同时跟踪三类量

$$\omega_{\star}(N_a), \quad Q_{\star}(N_a), \quad \mathcal{F}_{\text{FF},\star}(N_a), \quad (27.1)$$

其中 N_a 表示阵列周期数。若目标模相对首个竞争模的排序、主瓣方向或主瓣半高全宽在这三个尺寸之间仍翻转，则当前阵列尺寸就还不够大。式 (27.1) 的目的是把“有限尺寸足够”改写成必须展示的收敛趋势，而非把“多少个周期才够”固定成某个固定数字。

严格表述（难度）

有限阵列尺寸收敛判据。

若用最小辐射衰减长度近似，阵列半径可先按 [工程经验] 启发式写成 $R_{\text{array}} \gtrsim 3L_{\text{rad}}$ ，其中 $L_{\text{rad}} \approx 1/(2\alpha_{\text{rad}})$ 。这不是严格定理，必须配合尺寸扫描验证 [52]。它的作用是提醒读者：高 Q 、低辐射损耗模往往需要远大于直觉的横向尺寸才能显露“接近无限周期”的行为。对亮模或中等 Q 模，三维全波仍可能可行；对极高 Q 的 quasi-BIC 候选模，所需阵列尺寸会很快超出直接三维 FDTD 的现实能力，此时应转向 RCWA 单胞加横向包络或边缘修正的混合策略。对 2D 圆阵列，更严格的处理需使用 Bessel 函数展开 [32]。辐射场的径向分布满足

$$E(r) \propto J_0(k_r r) e^{-\alpha_{\text{rad}} r}, \quad (27.2)$$

其中 k_r 是横向波矢， α_{rad} 是辐射衰减常数。此时收敛判据应修正为

$$R_{\text{array}} \gtrsim \frac{2.5}{\alpha_{\text{rad}}}, \quad (27.3)$$

其中系数 2.5 对应第一类零阶 Bessel 函数第一个零点的位置 ($j_{0,1} \approx 2.405$)。这说明对辐射衰减常数较小的模式 (如 $\alpha_{\text{rad}} \approx 10 \text{ cm}^{-1}$)，所需阵列半径可达 $R_{\text{array}} \gtrsim 250 \mu\text{m}$ ，远超直觉估计。

从执行可行性角度，也可以记一个 [工程经验] 粗窗口：几十微米量级的孔阵通常仍由边缘泄漏主导；进入 $100\text{ }\mu\text{m}$ to $300\text{ }\mu\text{m}$ 后，有限阵列全波开始有讨论价值。该窗口必须通过“尺寸- Q -远场”三联收敛曲线确认，不能直接当作固定门槛 [52, 74]；若再上到数百微米甚至更大且目标模仍是高 Q 候选，就应把“FDTD 是否仍然可行”当成 workflow 决策，而不是默认继续堆网格。

27.8 第六类失败：忽略热漂移，默认冷腔频率就是工作频率

热效应之所以在 PCSEL 中危险，不只是因为它会让阈值变高，而是因为它会主动改写模式竞争。折射率上升、增益峰位漂移、自由载流子吸收变化和横向温度梯度，都会把原本在冷腔中占优的模式重新排序。症状通常是：冷腔仿真与低功率实验能对上，但一到连续波或高占空比工作就出现远场偏移、偏振不稳或模式跳变。

这一类失败常见的不当应对，是继续在冷腔里调几何参数，希望“把结构调到足够好”来抵消热漂移。更有效的应对策略是承认问题已经进入 L4 层级：必须把热源、温度场和光学回写显式纳入 workflow。若还停留在冷腔里调参，通常只会越调越远。

27.9 第七类失败：过度相信单一求解器或单一图像

不同求解器各有优缺点。RCWA 强于周期单胞开放散射，FDTD 强于宽带有限阵列响应，COMSOL 的频域有限元强于复杂材料和多物理场联动，CHARGE / Semiconductor 强于漂移-扩散电学，HEAT / Heat Transfer 强于温度场求解。任何单一求解器都不可能完美覆盖所有层级问题。若一个结论只在单一求解器中成立，且无法由第二种独立框架解释其趋势，则应暂时作为待验证结果处理。

这里要防止两个极端。一个极端是“某个软件最可靠，所以只信它”；另一个极端是“两个软件数值不一样，所以一定有一个错”。更成熟的做法是比较不变量和趋势，例如模式排序、主辐射方向、对参数扰动的单调性，而不是强迫两个求解器在每一个绝对数字上完全一致。

若 Lumerical 的 RCWA 与 COMSOL 的单胞频域模型都显示某个微扰会增强上方辐射并拉开两支偏振模的排序，那么即便两者给出的绝对 Q 不完全相同，这个趋势仍然是有价值的。反过来，若两者对趋势都不一致，就应优先回查几何、边界和材料定义，而不是急着宣布哪个软件“更可信”。

27.10 可复现故障案例：低截断怎样制造伪频移与伪分裂

图27.2 来自脚本 `code/pwem_cutoff_failure/run_pwem_cutoff_failure.py` 的实际计算。三组算例使用相同的正方晶格圆孔和 H_z 标量 PWEM，只改变倒格矢基底：对称 cutoff 4 保留 81 个平面波；对称 cutoff 1 只保留 9 个；第三组保留一个 5×3 的低阶矩形窗口，它不在 C_{4v} 旋转下闭合。

数字比图形更能说明问题：参考计算的第三、第四带在 Γ 点均为 0.39016；对称低截断把它们同时推到 0.49244，没有破坏简并，却产生了约 26% 的伪频移；非对称低基底则给出

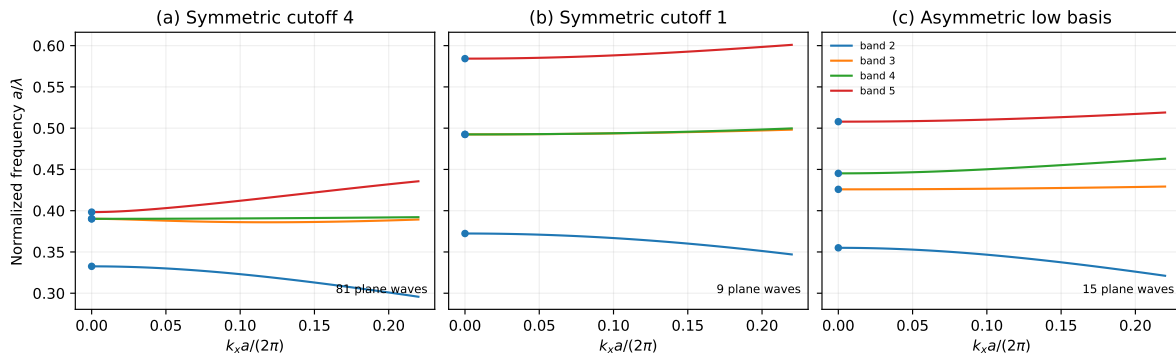


图 27.2: PWEM 截断故障的可复现实例。左：对称 cutoff 4 参考；中：低但仍保持方格对称的 cutoff 1，频率整体明显偏移，但 Γ 点双重态仍保持简并；右：不在 C_{4v} 下闭合的低阶非对称基底制造出假的 Γ 点分裂。

0.42586 与 0.44524，制造 0.01938 的伪分裂。由此可见，低截断本身首先造成不收敛，只有当截断窗口还破坏群对称性时，才会进一步伪造本不该存在的简并解除。检查 PWEM 不能只看“平面波数量是否够多”，还要检查保留基底是否在晶格点群下闭合。

27.11 这些失败在 Lumerical 与 COMSOL 中通常长什么样

为了让排错更贴近实战，表 27.1 把常见症状映射到两条常见软件链中最典型的表现形式。注意，这个表的目的是帮助你更快定位“这更像是层级错误、物理模型错误，还是数值实现错误”，而非软件评测。

表 27.1: 常见失败症状在 Lumerical 与 COMSOL 工作流中的典型表现。

症状	Lumerical 中常见表现	COMSOL 中常见表现	优先动作
目标模频率 / Q 对网格异常敏感	FDTD 网格或仿真时间窗不足，PML 过近，或 RCWA 阶数不足	有限元网格不足、PML 设置不当，或本征频率研究捕捉到伪模	先做单因子收敛
周期单胞与有限阵列结果矛盾	RCWA 与 FDTD 的阵列复制、边界或材料版本不一致	单胞周期模型与有限模型的几何或边界不一致	先做最小对照算例
引入增益后“阈值”过于乐观	把均匀增益直接当真实注入，未接 CHARGE / HEAT	在 Wave Optics 中直接加均匀增益，未做 Semiconductor / Heat Transfer 回写	降级为光学阈值结论
高电流下远场和偏振突然恶化	未把温度场回写到折射率和增益代理	热场与光场未联动，仍在用冷态材料参数	进入热-电-光回写
电学求解看似收敛但结果不合理	接触边界、掺杂或电流归一化设置有误	Semiconductor 接口的材料参数、边界或单位不一致	做量纲和边界审计

可按以下顺序记忆：

- 在 Lumerical 光学链里，最常见的危险信号是“结果对网格、PML 或仿真时间窗过分敏感”；
- 在 COMSOL 光学链里，最常见的危险信号是“本征频率或频域解很平滑，但其实抓到的是错误边界条件下的模式”；
- 在两条电热链里，最常见的危险信号都是“结果数值稳定，但物理上明显不自洽”，例如热点位置与电流路径常识不符。

27.12 推荐排错顺序：先层级，再物理，再数值

当结果不可信时，最有效的排错顺序通常是三步。

第一步，检查结论层级。问自己：我现在试图声称的东西，是否超出了模型当前可支持的层级。

第二步，检查物理模型。问自己：边界条件、几何、材料、增益、注入和热源是否与研究问题匹配，是否存在被遗忘的近似。

第三步，最后才检查数值实现。包括网格、时间步、傅里叶阶数、PML、求解容差和 study 设置。

这个顺序看似反直觉，因为很多人更愿意先调数值参数。但对 PCSEL 来说，若结论层级本来就错了，继续调数值只会让错误结果更平滑，而不会让它更接近真实物理。

易错点

看到“两个求解器数值不同”就立刻开始加密网格，往往不是最高收益动作。若它们解的本来就不是同一个几何、同一组边界或同一层级问题，再细的网格也不会让它们自动一致。

回看本章

PCSEL 建模中最需要警惕的失败，并非显性的求解器报错，而是层级错误、近似越界和物理问题错配导致的“看起来合理的错误结论”。最常见的失败模式包括：把冷腔结果读成器件结论、边界条件和几何定义跨求解器不一致、误用有效折射率近似、把均匀增益当真实电注入、有限尺寸建模不足、忽略热漂移，以及过度依赖单一求解器。排错的正确顺序应是先检查结论层级，再检查物理模型，最后才检查数值实现。保持这一顺序，可把许多仿真异常转化为可定位的问题。第二十五章 给出正向流程，第二十六章 给出示例执行，本章则提供反向审计表。

练习题

1. [基础] 从你最近的一次仿真中选一个“结果表面合理，但你并不完全相信”的案例，按本章的三步顺序写出排错计划。

2. [基础] 说明为什么把均匀增益模型直接用于电注入 PCSEL 的器件结论会造成系统性误判。
3. [进阶] 设计一个最小对照算例，用来验证你在两个不同求解器中的边界条件和材料定义是否一致。
4. [综合] (思路提示：先定层级，再列关键式) 结合 表27.1，写出一个同时适用于 Lumerical 与 COMSOL 的项目级排错清单。

延伸阅读

- 结合 第二十五章、第三十章 与 表25.1 阅读，把本章中的失败模式直接转成你自己的项目模板和调试清单。
- 若准备开始器件建模，建议把本章与 第二十六章和表26.3和表26.4 对照阅读：前者告诉你哪里最容易错，后者告诉你正确顺序该怎么走，以及每一步应该先产出什么证据。

第六部分

附录与索引

第二十八章 数学附录

读完本章，你能

- 把全书反复使用的时间因子、傅里叶展开、归一化、复频率和微扰公式整理成一套统一数学语法。
- 理解这些公式各自依赖的物理前提，避免把本来只对封闭或弱开放系统成立的表达式机械外推到所有 PCSEL 问题。
- 建立开放 Maxwell 问题中 QNM、伴随模、双正交和广义归一化的最小数学框架。
- 学会把这些数学工具用于检查仿真结果、统一跨章节符号，以及快速估计参数变化对模态的影响。

读前准备 建议已掌握 第三章、第四章、第七章和第十二章 中关于 Maxwell 方程、Bloch 展开、阈值条件和微扰观点的内容。

阅读主线 本章所处模型层级：跨层数学支撑（L1→L4）

图 图28.1 先把本章数学对象之间的关系固定下来，再进入各节公式推导，可明显降低“符号会看、结构不清”的阅读负担。

首次阅读建议路径 若是第一次阅读本附录，建议按以下顺序进行：先读时间因子与傅里叶约定，再读周期/Bloch 展开与卷积结构，再读 QNM 与归一化，最后读互易与微扰公式。其余段落可在需要时回查。

本附录是把正文反复调用、却不宜每次都从头展开的数学骨架集中讲清楚，而非把若干常见恒等式堆在一起的公式仓库。我们先统一时间因子和频域约定，因为一切损耗、增益和复频率的符号都依赖它；随后建立周期介质中的傅里叶与 Bloch 展开，解释卷积结构为什么在第五章 的能带概念、第十三章 的 PWEM、以及后续 RCWA 和 CWT 中一再出现；再讨论内积、归一化和开放系统中的复频率与 Q ；最后给出互易性与一阶微扰公式，并明确说明它们在哪些条件下可靠、在哪些条件下只能作为工程近似。

28.1 为什么数学附录不是装饰，而是主结论的一部分

PCSEL 的建模链横跨电磁场、半导体输运和热传导。只要这些模块之间的时间因子、损耗符号、傅里叶约定或归一化不一致，最终就会出现一种很危险的情况：看起来像是“物理结

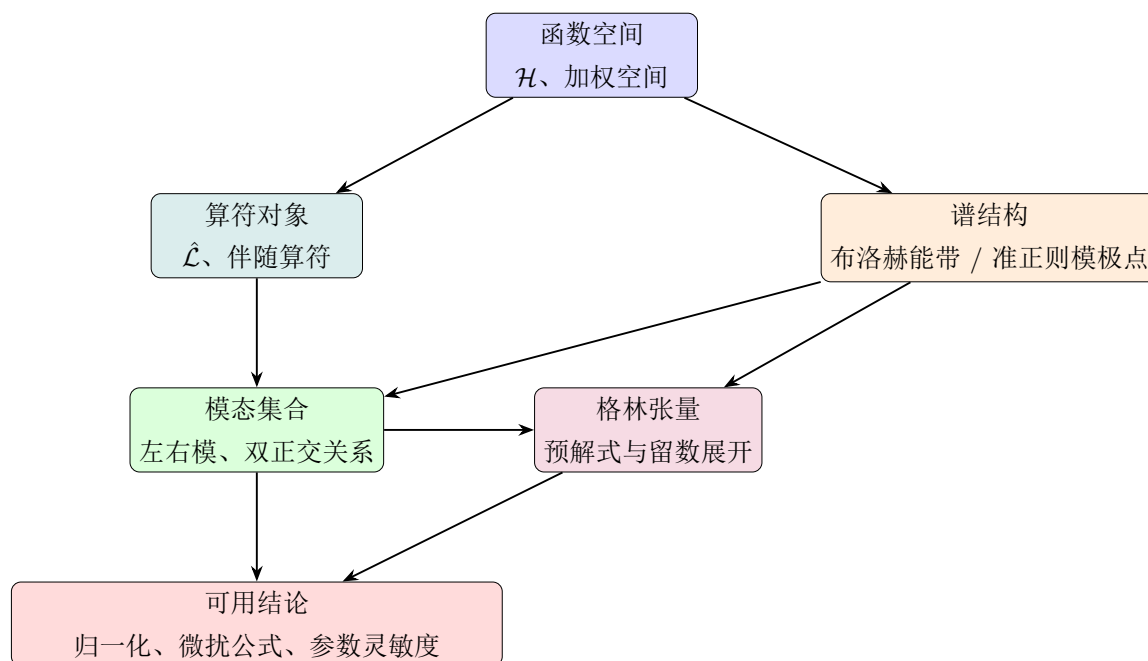


图 28.1: 本附录的数学对象关系图：从函数空间与算符出发，连接到谱结构、模态展开与格林张量，最终落到可直接使用的公式。

论不同”，实际却只是数学约定变了。最典型的例子就是时间因子的选择。一旦一个章节使用 $e^{-i\omega t}$ ，另一个局部推导却默默改成 $e^{+i\omega t}$ ，那么介质虚部、损耗符号、复频率虚部和 Q 的定义就会全部翻转。若读者没有意识到这种符号翻转，很容易把纯符号问题误读成腔损耗差异或增益符号差异。

因此，本附录的目标是把正文各章默认共享的数学地基显式写出来，而非补充“额外数学”。只要地基统一，后面的开放系统、耦合波理论、RCWA、FDTD/FEM 提取和器件级回写才可能真正接成一条链。

直觉框

如果把整本书看成一套多物理场模型，那么数学附录的作用就像接口协议。接口协议一旦不统一，上层结果可能只是数值上偶然一致，而不能视为可复用结论。

28.2 时间因子与频域约定：先把增益和损耗的符号说清楚

全书统一采用时间因子

$$\mathbf{E}(\mathbf{r}, t) = \Re\{\mathbf{E}(\mathbf{r})e^{-i\omega t}\}, \quad \mathbf{H}(\mathbf{r}, t) = \Re\{\mathbf{H}(\mathbf{r})e^{-i\omega t}\}. \quad (28.1)$$

在这一约定下，频域 Maxwell 方程写成

$$\nabla \times \mathbf{E} = +i\omega \mathbf{B}, \quad (28.2)$$

$$\nabla \times \mathbf{H} = \mathbf{J} - i\omega \mathbf{D}. \quad (28.3)$$

若采用线性各向同性本构关系

$$\mathbf{D} = \varepsilon_0 \varepsilon \mathbf{E}, \quad \mathbf{B} = \mu_0 \mu \mathbf{H}, \quad (28.4)$$

则在无外加电流时可得到熟悉的波动方程。重要的是它对损耗和增益符号的约束，而非方程形式本身。

在约定式 (28.1) 下，若介质为被动介质，则通常要求

$$\Im \varepsilon(\omega) > 0, \quad (28.5)$$

因为这对应正的平均耗散功率。这个结论可以从时均功率密度看出来。对无外加源的谐波场，时均耗散功率密度满足

$$P_{\text{abs}} = \frac{\omega \varepsilon_0}{2} \Im \varepsilon |\mathbf{E}|^2, \quad (28.6)$$

其中假设磁损耗可以忽略或已单独处理。于是，在本书采用的符号体系里，正的虚部表示吸收，若要以等效介电常数虚部的方式表示增益，则会对应负的有效虚部。这条符号约定看似基础，却是所有“阈值增益符号到底该写成正还是负”的根源。

式 (28.5) 和式 (28.6) 都依赖于本书统一采用的时间因子 $e^{-i\omega t}$ 。若改用 $e^{+i\omega t}$ ，则介质虚部和复频率虚部的物理判据会整体翻转。任何引用外部文献中的损耗或增益公式时，都必须先检查其时间因子约定。

为了避免正文不同章节在符号上“看起来一样、其实约定不同”，这里把本附录与主文完全一致的傅里叶变换约定写清楚。时间傅里叶变换采用

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{+i\omega t} dt, \quad f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{-i\omega t} d\omega, \quad (28.7)$$

它与时间因子式 (28.1) 保持同一符号体系。平面上的空间傅里叶变换采用

$$\tilde{f}(k_x, k_y) = \iint f(x, y) e^{-i(k_x x + k_y y)} dx dy, \quad (28.8)$$

$$f(x, y) = \frac{1}{(2\pi)^2} \iint \tilde{f}(k_x, k_y) e^{+i(k_x x + k_y y)} dk_x dk_y. \quad (28.9)$$

把这一步写明的价值在于：后面近远场变换、角谱传播、频率噪声谱以及 RCWA/FDTD/FEM 的谱量定义，都能回到同一组符号，而不会在章节切换时悄悄换掉正负号。

28.3 周期介质中的傅里叶展开：为什么卷积会反复出现

正文的 PWEM、RCWA 和 CWT 之所以看起来形式不同，却总会出现相似的矩阵耦合项，根源在于周期介质中的傅里叶卷积。设平面内周期单胞为 A_c ，倒格矢为 $\mathbf{G}$ 。任意满足该周期的标量场 $f(\mathbf{r}_{\parallel})$ 都可展开为

$$f(\mathbf{r}_{\parallel}) = \sum_{\mathbf{G}} f_{\mathbf{G}} e^{i\mathbf{G} \cdot \mathbf{r}_{\parallel}}, \quad f_{\mathbf{G}} = \frac{1}{A_c} \int_{A_c} f(\mathbf{r}_{\parallel}) e^{-i\mathbf{G} \cdot \mathbf{r}_{\parallel}} d^2 r. \quad (28.10)$$

若场满足 Bloch 形式, 则还应再乘上相位因子 $e^{ik \cdot r_{\parallel}}$ 。于是任意 Bloch 场都可写为

$$\mathbf{E}_k(\mathbf{r}_{\parallel}, z) = e^{ik \cdot r_{\parallel}} \sum_G \mathbf{E}_{k+G}(z) e^{iG \cdot r_{\parallel}}. \quad (28.11)$$

关键的是乘积项的展开。若介电函数也按同样方式展开,

$$\varepsilon(\mathbf{r}_{\parallel}) = \sum_G \varepsilon_G e^{iG \cdot r_{\parallel}}, \quad (28.12)$$

那么 $\varepsilon \mathbf{E}$ 的傅里叶系数不再是简单乘法, 而是卷积:

$$(\varepsilon E)_G = \sum_{G'} \varepsilon_{G-G'} E_{G'}. \quad (28.13)$$

因此周期结构中的本征问题总会变成“一个分量与许多倒格矢分量耦合”的矩阵问题。PWEM 把这个卷积直接写成平面波矩阵; RCWA 把它写进层内本征模矩阵; CWT 则主动只保留少数最重要的耦合项。

若单胞几何只保留最强的四个一级傅里叶分量, 那么 CWT 中保留的四波耦合, 本质上就是把式 (28.13) 从无限维卷积截断成最小闭合子空间。它的物理解释更强, 但严格性当然不如 RCWA 或大基底 PWEM。

28.4 开放 Maxwell 问题的函数空间直觉: 为什么普通 Hilbert 空间不够

从一个简单例子理解开放模式的复频率

考虑一个法布里 - 珀罗腔, 两端镜面反射率为 $R < 1$ 。腔内往返一次的振幅衰减因子为 $r^2 = R$ 。若初始时刻 $t = 0$ 的场振幅为 A_0 , 则经过一次往返时间 $\tau_{rt} = 2L/c$ 后变为 $A_1 = A_0 \sqrt{R}$ 。这等价于一个复频率阻尼振荡:

$$A(t) = A_0 e^{-i\omega_0 t} e^{-\gamma t}, \quad e^{-\gamma \tau_{rt}} = \sqrt{R}.$$

于是 $\gamma = -\ln(R)/(2\tau_{rt})$ 。这就是开放腔复频率 $\tilde{\omega} = \omega_0 - i\gamma$ 的最简单范例。QNM 理论要做的, 就是把这种直观推广到任意形状的开放谐振腔。

在封闭、无损、厄米问题中, 我们习惯把本征模视为某个 Hilbert 空间里的正交基。这个直觉很有用, 但它并不能原封不动地搬到 PCSEL 的开放模问题上。根源在于开放系统的共振模满足的是纯出射边界条件, 而不是封闭边界条件。对一个复频率模

$$\tilde{\omega}_n = \omega_n - i\gamma_n, \quad \gamma_n > 0, \quad (28.14)$$

其远场渐近形式可写成

$$\mathbf{E}_n(\mathbf{r}) \sim \frac{e^{i\tilde{k}_n r}}{r} \mathbf{f}_n(\hat{\mathbf{r}}), \quad \tilde{k}_n = \frac{\tilde{\omega}_n}{c}. \quad (28.15)$$

由于

$$e^{i\tilde{k}_n r} = e^{i\omega_n r/c} e^{+\gamma_n r/c}, \quad (28.16)$$

模场若按纯出射条件延拓到无穷远，空间上通常不是衰减而是增长。这并非说模式“物理发散”，而是说它不再属于普通平方可积的束缚态函数空间。开放模更自然的数学身份，是 Green 张量或散射矩阵在复频率平面上的极点态 [39, 3]。

这条函数空间直觉非常关键，因为它直接解释了后面三个常见困惑：

- 为什么开放共振模的归一化不是唯一的；
- 为什么非厄米微扰不能直接照搬 $\int \delta\epsilon |\mathbf{E}|^2$ 形式；
- 为什么在 PML 中求得的“模”虽然数值上有限，却仍对应同一开放系统共振。

封闭腔本征模更像“盒子里的驻波”；开放共振模更像“散射振幅在复频率平面上的极点”。只要把这一步想通，QNM、伴随模和广义归一化就不再是后贴的数学技巧，而是开放问题自然逼出来的语言。

28.5 QNM / resonant states 的最小定义

本书中所说的准正规模（quasinormal mode, QNM）或共振态（resonant state），最小可以定义为满足无源 Maxwell 方程且只允许出射波的非平凡解：

$$\hat{\mathcal{L}}(\tilde{\omega}_n) \mathbf{E}_n^R = 0, \quad \text{outgoing boundary condition only.} \quad (28.17)$$

这里上标 R 表示右模（right mode），是为了和稍后出现的左模区分。对开放系统，这个定义与“ $\tilde{\omega}_n$ 是 Green 张量或散射矩阵极点”是等价的 [60, 3]。两种说法分别对应两种直觉：

- 方程视角强调它是无源 Maxwell 问题的本征解；
- 散射视角强调它是实际可测散射响应的极点结构。

这也是为什么 PCSEL 中的开放模分析天然应使用复频率语言。只要目标问题涉及法向辐射、上下出光不对称、quasi-BIC、远场整形或辐射损耗排序，把模式写成 QNM / resonant states 就比把它硬塞回封闭腔本征模更贴近真实物理。

例子

最小的一维开放腔例子，是长度为 L 的均匀 Fabry–Perot 腔，两端反射系数分别为 r_1, r_2 ，并要求边界外只有出射波。此时共振条件写成

$$r_1 r_2 e^{2i\tilde{k}L} = 1. \quad (28.18)$$

只要 $|r_1 r_2| < 1$ ，其解 $\tilde{k}$ 必为复数，等价地 $\tilde{\omega} = c\tilde{k}$ 也必为复数。这个最小例子非常值得记住，因为它说明：复频率是开放边界条件本身逼出来的本征值，而非后来为了描述损耗而“补上去”的修正。

28.5.1 QNM 理论的核心三个数学问题（进阶选读）

准正规模理论在实际应用中面临三个相互关联的数学挑战：(i) 模场的空间增长导致传统 L^2 内积发散；(ii) 非厄米算符的本征模不构成正交系而需采用双正交基；(iii) QNM 集合的完备性问题。本节逐一解决这些问题。

1. 发散积分的正则化处理

考虑一维 Fabry-Perot 腔的 QNM 解。在腔外区域 ($|x| > L/2$)，模场具有如下渐近形式：

$$E_n(x) \sim E_0 e^{i\tilde{k}_n|x|}, \quad \tilde{k}_n = \frac{\tilde{\omega}_n}{c} = k_n - i\kappa_n, \quad (28.19)$$

其中 $\kappa_n > 0$ 对应模式的辐射衰减率。将复波矢代入模场表达式可得，当 $x \rightarrow \infty$ 时， $|E_n(x)|^2 \sim e^{2\kappa_n x}$ 指数增长，导致传统的 L^2 范数积分发散：

$$\int_{-\infty}^{\infty} |E_n(x)|^2 dx = \infty. \quad (28.20)$$

为解决这一问题，现代 QNM 理论采用以下几种等价的正则化方案：

(a) 体积极限正则化 (Volume regularization): 先将系统限制在有限体积 $V_L = [-L, L]$ 内计算，最后取 $L \rightarrow \infty$ 极限并提取有限部分 (finite part)。具体而言，将发散积分写为：

$$\int_{-L}^L |E_n(x)|^2 dx = A e^{2\kappa_n L} + B(L), \quad (28.21)$$

其中 A 是发散的系数， $B(L)$ 是随 L 缓变的部分。定义正则化后的内积为去掉指数发散项后的剩余部分：

$$\langle E_n, E_m \rangle_{\text{reg}} := \lim_{L \rightarrow \infty} \left[\int_{-L}^L E_n^*(x) E_m(x) dx - \delta_{nm} \frac{A_n}{2\kappa_n} e^{2\kappa_n L} \right]. \quad (28.22)$$

(b) 解析延拓法 (Analytic continuation): 将 QNM 视为复波数平面上的亚纯函数，通过将实轴积分路径旋转到复平面的适当位置来实现收敛。设 $k = \tilde{k}_n + s e^{i\theta}$ ，选择旋转角 θ 使得被积函数沿新路径指数衰减。这种方法与留数定理自然衔接。

(c) PML 正则化: 引入完美匹配层坐标变换 $x \rightarrow \tilde{x}(x) = x + i f(x)$ ，其中 $f(x)$ 在计算域外部单调递增。在此变换下，原发散的被积函数变为：

$$\int_{-\infty}^{\infty} |E_n(x)|^2 dx \longrightarrow \int_{-\infty}^{\infty} |E_n(\tilde{x})|^2 \frac{d\tilde{x}}{dx} dx, \quad (28.23)$$

由于 $\Im[\tilde{k}_n \tilde{x}] < 0$ 在外区成立，积分自动收敛。这是商业软件中最常用的实现方式。

这三种方法在数学上是等价的，都给出了相同的广义归一化结果 [34, 60]。

2. 双正交基的严格构造

对于非厄米 Maxwell 算符 $\hat{\Theta}$ ，其右本征模 $\{|R_n\rangle\}$ 满足：

$$\hat{\Theta}|R_n\rangle = \tilde{\omega}_n^2 |R_n\rangle, \quad (28.24)$$

但这些右模之间并不正交: $\langle R_m | R_n \rangle \neq 0$ ($m \neq n$)。为了构建有效的模态展开, 必须引入伴随算符 $\hat{\Theta}^\dagger$ 的左本征模 $\{|L_n\rangle\}$:

$$\hat{\Theta}^\dagger |L_n\rangle = (\tilde{\omega}_n^2)^* |L_n\rangle. \quad (28.25)$$

关键结论是: 左右模构成双正交系 (biorthogonal system):

$$\langle L_m | R_n \rangle = N_n \delta_{mn}, \quad (28.26)$$

其中 N_n 是依赖于归一化约定的常数。这里的内积应采用前述正则化方案之一来计算。

双正交系的几何图像 在非厄米系统中, 左模和右模不再是彼此的共轭转置, 而是两个独立的向量集。右模 $\{|R_n\rangle\}$ 张成状态空间, 而左模 $\{|L_n\rangle\}$ 作为其对偶基底 (dual basis)。任意物理态 $|\Psi\rangle$ 按右模展开时, 展开系数由对应的左模投影给出: $c_n = \langle L_n | \Psi \rangle / N_n$ 。这种关系类似于协变矢量和逆变矢量——通过配对缩并得到标量。

3. 完备性关系与模态展开

若 QNM 集合 $\{|R_n\rangle\}$ 是完备的, 则任意物理态 $|\Psi(t)\rangle$ 可以展开为:

$$|\Psi(t)\rangle = \sum_n c_n(t) |R_n\rangle, \quad c_n(t) = \frac{\langle L_n | \Psi(t) \rangle}{\langle L_n | R_n \rangle}. \quad (28.27)$$

完备性的数学表述是闭包关系 (closure relation):

$$\sum_n \frac{|R_n\rangle \langle L_n|}{\langle L_n | R_n \rangle} = \hat{I}, \quad (28.28)$$

其中 $\hat{I}$ 是单位算符。

然而, 对于开放系统, QNM 的完备性并非总是成立。Leung 等人 [39] 证明了以下重要结论:

- **有限尺寸散射体:** 若介电常数扰动 $\Delta\epsilon(\mathbf{r})$ 具有紧支集 (compact support), 则 QNM 集合在散射体外部的区域内是完备的。
- **色散介质:** 当 $\epsilon(\omega)$ 显含频率时, 需要在扩展的状态空间 (包含介质极化自由度) 中讨论完备性。此时 QNM 可能与连续谱混合, 形成所谓的“quasinormal continuum”。
- **高 Q 值近似:** 对于高品质因子腔 ($Q \gg 1$), 前几个主导 QNM 通常足以精确描述腔内的场分布, 即使数学上不完备, 工程上也足够准确。

从 Green 函数的角度看, 完备性等价于说 Green 算符可以用 QNM 展开:

$$\overleftrightarrow{G}(\mathbf{r}, \mathbf{r}'; \omega) = \sum_n \frac{\mathbf{E}_n(\mathbf{r}) \otimes \mathbf{E}_n^L(\mathbf{r}')}{2\tilde{\omega}_n(\omega - \tilde{\omega}_n)}, \quad (28.29)$$

其中 $\mathbf{E}_n^L$ 是左模的电场分量。这个展开式在远离分支切割 (branch cut) 的区域一致收敛, 而在接近连续谱阈值时需要额外小心处理 [60, 34]。

实际应用建议 对于 PCSEL 仿真，推荐采用以下工作流程：

1. 使用 COMSOL/Lumerical 的 QNM 求解器获取前 N 个主导模式的复频率 $\tilde{\omega}_n$ 和场分布 $\mathbf{E}_n(\mathbf{r})$
2. 在同一平台上计算对应的左模 $\mathbf{E}_n^L$ (或通过互易性关系间接获得)
3. 验证模态截断误差：先计算双正交投影系数

$$c_n = \frac{\langle L_n | \Psi_{\text{target}} \rangle}{\langle L_n | R_n \rangle},$$

再以重构残差定义误差

$$\eta = \frac{\left\| \Psi_{\text{target}} - \sum_{n=1}^N c_n R_n \right\|_W}{\left\| \Psi_{\text{target}} \right\|_W}$$

4. 若 $\eta < 1\%$ ，则认为当前 QNM 基组足够用于后续的微扰分析和耦合计算

28.6 光子学中的 Green 函数方法简介

前面已经从本征模的角度讨论了开放系统的 QNM 框架。但在现代纳米光子学中，还有一个更加统一的语言——Green 函数方法。它不仅包含了本征模的所有信息，还能自然地处理光源激发、散射问题和自发辐射耦合等物理过程。

28.6.1 从标量波动方程的 Green 函数出发

考虑最简单的标量 Helmholtz 方程：

$$[\nabla^2 + k_0^2 \varepsilon(\mathbf{r})] \psi(\mathbf{r}) = -s(\mathbf{r}), \quad (28.30)$$

其中 $s(\mathbf{r})$ 是源项。我们希望找到一个“单位点源的响应” $G(\mathbf{r}, \mathbf{r}')$ ，使得对任意源分布都能通过叠加原理写出解：

$$\psi(\mathbf{r}) = \int G(\mathbf{r}, \mathbf{r}') s(\mathbf{r}') d^3 r'. \quad (28.31)$$

这就要求 G 满足：

$$[\nabla^2 + k_0^2 \varepsilon(\mathbf{r})] G(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'). \quad (28.32)$$

这就是 Green 函数的定义：它是算符 $\hat{L} = \nabla^2 + k_0^2 \varepsilon(\mathbf{r})$ 的逆算符的核 (kernel)。

因果性与推迟 Green 函数 对于时谐问题，还需要指定边界条件来确定 Green 函数的唯一性。物理上合理的要求是**因果性**：响应不能超前于激励。在频域中，这等价于要求当 $r \rightarrow \infty$ 时只有出射波：

$$G(\mathbf{r}, \mathbf{r}') \sim \frac{e^{ik_0 n r}}{4\pi r}, \quad r \rightarrow \infty. \quad (28.33)$$

这个条件被称为 Sommerfeld 辐射条件。满足它的 Green 函数称为**推迟 Green 函数** (retarded Green's function)。

例子

[自由空间的标量 Green 函数] 在均匀介质中 ($\varepsilon = \text{常数}$), 可以直接验证:

$$G_0(\mathbf{r}, \mathbf{r}') = \frac{e^{ik_0 n |\mathbf{r} - \mathbf{r}'|}}{4\pi |\mathbf{r} - \mathbf{r}'|} \quad (28.34)$$

满足式 (28.32) 和辐射条件。这个表达式的重要性在于: 任何复杂结构的 Green 函数都可以写成 $G = G_0 + \Delta G$ 的形式, 其中 ΔG 包含了散射体的全部信息。

28.6.2 Dyson 方程及其物理含义

设我们把介电常数写成背景加扰动: $\varepsilon(\mathbf{r}) = \varepsilon_{bg}(\mathbf{r}) + \Delta\varepsilon(\mathbf{r})$ 。相应地, Green 函数也可以分解为背景部分 G_0 和散射修正 ΔG 。将它们代入定义式并整理, 可得到著名的 Dyson 方程:

$$G(\mathbf{r}, \mathbf{r}') = G_0(\mathbf{r}, \mathbf{r}') + \int G_0(\mathbf{r}, \mathbf{r}'') k_0^2 \Delta\varepsilon(\mathbf{r}'') G(\mathbf{r}'', \mathbf{r}') d^3 r''. \quad (28.35)$$

这个方程的物理图像极其清晰: 总响应等于直接响应加上所有可能的单次/多次散射路径的总和。如果把这个方程迭代展开, 就得到 Born 级数:

$$G = G_0 + G_0(k_0^2 \Delta\varepsilon)G_0 + G_0(k_0^2 \Delta\varepsilon)G_0(k_0^2 \Delta\varepsilon)G_0 + \cdots \quad (28.36)$$

$$= G_0 + G_0 T G_0, \quad (28.37)$$

其中 T 算符总结了所有阶次的散射过程。

Dyson 方程的核心思想是“递归”: 要知道总的散射响应, 可以先问“第一次散射发生在哪里”, 然后再问“散射之后的传播又如何”。这种自相似的结构正是 Green 函数方法强大之处的源泉。

28.6.3 矢量 Green 张量与 Maxwell 方程

对真实的电磁场问题, 我们需要处理的是矢量波动方程。此时 Green 函数推广为一个二阶张量 $\overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}')$ (也称为 dyadic Green's function), 满足:

$$\nabla \times \nabla \times \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') - k_0^2 \varepsilon(\mathbf{r}) \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = \overline{\mathbf{I}} \delta(\mathbf{r} - \mathbf{r}'), \quad (28.38)$$

其中 $\overline{\mathbf{I}}$ 是单位并矢 (unit dyad)。电场与电流源的关系为:

$$\mathbf{E}(\mathbf{r}) = i\omega\mu_0 \int \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}(\mathbf{r}') d^3 r'. \quad (28.39)$$

注意这里的因果关系: 偶极子源 $\mathbf{J}$ 产生场 $\mathbf{E}$, 中间的传递函数就是 Green 张量。

28.6.4 周期结构中的 dyadic Green 函数与 Bloch-Floquet 展开

对 PCSEL 这样的周期结构, Green 张量也必须满足 Bloch 定理。具体来说, 可以将其展开为:

$$\overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}'; \mathbf{k}) = \sum_n \frac{\mathbf{E}_n^R(\mathbf{r}; \mathbf{k}) \otimes \mathbf{E}_n^L(\mathbf{r}'; \mathbf{k})}{2\tilde{\omega}_n(\mathbf{k})[\tilde{\omega} - \tilde{\omega}_n(\mathbf{k})]}, \quad (28.40)$$

其中 $\mathbf{E}_n^{R/L}$ 分别是右模和左模, $\tilde{\omega}_n$ 是复本征频率。这个表达式有几个关键特征:

1. 分母中的极点结构清楚地表明：Green 函数的奇点对应系统的本征模；
2. 分子中的双线性形式反映了非厄米系统的双正交特性；
3. 求和覆盖了所有离散模和连续谱贡献（后者在某些情况下需要写成积分）。

式 (28.40) 的严格推导需要小心处理收敛性和完备性问题。对开放系统，由于 QNM 的空间增长特性，这个展开在某些区域可能条件收敛甚至发散。实用的解决方案包括使用 PML 正则化或采用适当的解析延拓技术 [60, 34]。

28.6.5 Green 函数极点与准正则模的对应关系

现在我们回到之前留下的问题：为什么说 QNM 是 Green 函数的极点？从式 (28.40) 可以看出，当探测频率 $\tilde{\omega}$ 趋近某个本征值 $\tilde{\omega}_n$ 时，对应的项会发散。在复频率平面上，这些发散的点就是极点（pole）。

更严格地说，可以在极点附近做 Laurent 展开：

$$\overline{\mathbf{G}}(\tilde{\omega}) \approx \frac{\text{Res}_n}{\tilde{\omega} - \tilde{\omega}_n} + \text{regular part}, \quad (28.41)$$

其中留数 Res_n 正好正比于 $\mathbf{E}_n^{\text{R}} \otimes \mathbf{E}_n^{\text{L}}$ 。这就建立了两种语言之间的精确对应：

QNM 语言	Green 函数语言
本征频率 $\tilde{\omega}_n$	极点位置
模场 $\mathbf{E}_n^{\text{R/L}}$	留数的因子分解
归一化条件	留数的归一化

28.6.6 Purcell 因子的 Green 函数表达

Green 函数方法的一个重要应用是计算自发辐射增强因子（Purcell factor）。考虑一个位于 $\mathbf{r}_0$ 的电偶极子 $\mathbf{p}$ ，其辐射功率可通过 Green 张量表示为：

$$P = \frac{\omega}{2} \Im [\mathbf{p}^* \cdot \mathbf{E}(\mathbf{r}_0)] = \frac{\omega^2 \mu_0}{2} |\mathbf{p}|^2 \hat{\mathbf{p}} \cdot \Im [\overline{\mathbf{G}}(\mathbf{r}_0, \mathbf{r}_0)] \cdot \hat{\mathbf{p}}, \quad (28.42)$$

其中 $\hat{\mathbf{p}}$ 是偶极矩方向的单位矢量。相对于真空中的辐射功率 P_0 ，Purcell 因子定义为：

$$F_P(\mathbf{r}_0, \hat{\mathbf{p}}, \omega) = \frac{P}{P_0} = \frac{6\pi c}{\omega} \hat{\mathbf{p}} \cdot \Im [\overline{\mathbf{G}}(\mathbf{r}_0, \mathbf{r}_0)] \cdot \hat{\mathbf{p}}. \quad (28.43)$$

这个公式的强大之处在于：它不需要显式地知道任何本征模，只需要计算源点处的 Green 张量即可。当然，如果将式 (28.40) 代入，也能得到用 QNM 展开的形式：

$$F_P(\mathbf{r}_0, \hat{\mathbf{p}}, \omega) \approx \sum_n \Re \left[\frac{A_n}{\omega - \tilde{\omega}_n} \right], \quad (28.44)$$

其中

$$A_n = \frac{6\pi c}{\omega} \frac{(\hat{\mathbf{p}} \cdot \mathbf{E}_n^{\text{R}}(\mathbf{r}_0))(\hat{\mathbf{p}} \cdot \mathbf{E}_n^{\text{L}}(\mathbf{r}_0))}{2\tilde{\omega}_n \langle \mathbf{E}_n^{\text{L}}, \mathbf{E}_n^{\text{R}} \rangle}$$

是由左右模重叠与归一化共同决定的复残数。仅在单模、弱重叠、高 Q 近似下，式 (28.44) 才可进一步退化为 Lorentzian 峰形；一般开放系统可出现非 Lorentz 线形和模间干涉项。

可能出错处

有些初学者在使用式 (28.44) 时会忘记两点：第一，这个求和通常需要包含足够多的模才能收敛；第二，对强开放模，QNM 的空间增长特性可能导致式 (28.40) 在某些位置不适用。在这些情况下，直接使用全波仿真提取 $\Im[\bar{G}]$ 反而更可靠。

28.7 伴随模、双正交与广义归一化

一旦算子不再自伴，右模本身通常不足以完成投影与微扰分析。对第一次接触这套语言的读者，最稳妥的直觉是：先把左模看成做投影时的测试函数，而不要先把它想成“另一支也要单独观察到的物理场”。

伴随算子是什么？选定一个内积 $\langle \cdot, \cdot \rangle$ 后，伴随算子由“分部积分后的移项”定义：

$$\langle u, \hat{\mathcal{L}}v \rangle = \langle \hat{\mathcal{L}}^A u, v \rangle + \text{边界项} \quad (28.45)$$

若边界项为零，则 $\hat{\mathcal{L}}^A$ 就是把 $\hat{\mathcal{L}}$ “搬到左边”得到的算子；封闭厄米问题中更进一步有 $\hat{\mathcal{L}}^A = \hat{\mathcal{L}}$ 。

例子

数值矩阵里的最小构造。若仿真离散后得到矩阵本征问题 $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ ，并选用内积 $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\dagger \mathbf{M} \mathbf{v}$ ($\mathbf{M}$ 常为能量权重矩阵或 FEM 质量矩阵)，则其伴随矩阵为

$$\mathbf{A}^A = \mathbf{M}^{-1} \mathbf{A}^\dagger \mathbf{M},$$

左本征矢满足 $\mathbf{A}^A \mathbf{y} = \lambda^* \mathbf{y}$ 。在最常见的标准内积 $\mathbf{M} = \mathbf{I}$ 下， $\mathbf{A}^A = \mathbf{A}^\dagger$ ；若问题是 complex-symmetric（常见于互易无损系统的某些离散形式），则可进一步退化为 $\mathbf{A}^A = \mathbf{A}^\top$ 。这就是“伴随算子 = 离散矩阵的伴随”的最小可操作版本。

为什么开放系统必须区分左右模？开放边界使边界项不再自动消失，因此左模与右模必须分开定义。更一般的做法是同时引入右模与伴随左模：

$$\hat{\mathcal{L}}(\tilde{\omega}_n) \mathbf{E}_n^R = 0, \quad \hat{\mathcal{L}}^A(\tilde{\omega}_n) \mathbf{E}_n^L = 0. \quad (28.46)$$

其中 $\hat{\mathcal{L}}^A$ 是伴随算子， $\mathbf{E}_n^L$ 是左模。对一般非厄米问题，真正起作用的是左模和右模形成的双正交结构，而不是右模与自身的厄米内积。也就是说，开放系统中的“正交”更接近

$$\langle \mathbf{E}_m^L, \mathbf{E}_n^R \rangle \propto \delta_{mn}, \quad (28.47)$$

而不是封闭腔里熟悉的 $\langle \mathbf{E}_m, \mathbf{E}_n \rangle_\varepsilon = \delta_{mn}$ 。

对 Maxwell 开放问题，进一步的细节取决于介质是否互易、是否色散，以及数值上是否通过 PML 处理开放边界。对许多互易系统，一个常用的广义 QNM 归一化形式是

$$\langle \Psi_m, \Psi_n \rangle = \int_{\Omega+\text{PML}} \left[\mathbf{E}_m \cdot \frac{\partial(\omega\varepsilon)}{\partial\omega} \mathbf{E}_n - \mathbf{H}_m \cdot \frac{\partial(\omega\mu)}{\partial\omega} \mathbf{H}_n \right] dV, \quad (28.48)$$

其中 $\Psi_n = (\mathbf{E}_n, \mathbf{H}_n)$ ，且积分区域包含物理区与 PML [60, 34]。这个表达式的意义不在于要求读者把它背下来，而在于明确三件事：

1. 广义归一化里经常不出现普通的复共轭；
2. 材料色散会直接进入归一化，而不是只出现在“材料模型”章节；
3. PML 并不只是吸收边界的数值技巧，在 QNM 归一化里它还承担把开放问题转成可积配对的数学角色。

在封闭、无损、无色散极限下，式 (28.48) 会退回更熟悉的能量型归一化；但对真实 PCSEL 的开放模，若继续坚持只用

$$\langle \mathbf{E}_m, \mathbf{E}_n \rangle_\varepsilon = \int_V \varepsilon(\mathbf{r}) \mathbf{E}_m^*(\mathbf{r}) \cdot \mathbf{E}_n(\mathbf{r}) dV = \delta_{mn}. \quad (28.49)$$

往往就会在归一化、模耦合和微扰分母上埋下系统错误。也正因此，本书正文只有在明确处于封闭或弱开放近似下，才会继续使用式 (28.49) 这类厄米归一化。

这里有两条最容易被忽略、却又最关键的原因。第一，对纯出射 QNM 取复共轭后，会把出射边界条件翻成入射边界条件，因此 $\mathbf{E}^*$ 本身通常不是同一个开放本征问题的合法测试函数。第二，开放模在无穷远处的空间延拓并不平方可积，若仍执意把它塞进普通厄米内积，分子、分母和归一化常数都可能一起失真。正因为这两点，开放系统一阶微扰不能机械照搬封闭腔公式。

这意味着一个非常重要的实务结论：若两个不同求解器采用了不同归一化方式，不能直接比较它们输出的“场强大小”，必须比较与归一化无关的不变量，或者先把它们转换到同一内积和归一化框架下。正文中大量出现的重叠因子之所以用分式定义，也是为了让归一化常数在分子分母中相消。

例如，量子阱重叠因子常写成

$$\Gamma_{\text{QW}} = \frac{\int_{V_{\text{QW}}} \varepsilon |\mathbf{E}|^2 dV}{\int_V \varepsilon |\mathbf{E}|^2 dV}, \quad (28.50)$$

它之所以稳健，就在于不依赖模整体振幅的选取。

28.8 复频率、衰减率与 Q ：不要混淆振幅寿命和能量寿命

开放腔模常写成复频率形式

$$\tilde{\omega} = \omega_r - i\gamma, \quad \gamma > 0. \quad (28.51)$$

代回时间因子后，场的时间依赖为

$$e^{-i\tilde{\omega}t} = e^{-i\omega_r t} e^{-\gamma t}. \quad (28.52)$$

因此场振幅的衰减时间常数是 $1/\gamma$ 。但能量与场振幅平方成正比，所以腔内能量按

$$U(t) \propto e^{-2\gamma t} \quad (28.53)$$

衰减, 对应的能量寿命为

$$\tau_p = \frac{1}{2\gamma}. \quad (28.54)$$

这时品质因子定义为每个周期储能与损耗的比值, 可写成

$$Q = \omega_r \tau_p = \frac{\omega_r}{2\gamma}. \quad (28.55)$$

这个推导之所以需要在附录里明确写出, 是因为很多实际文稿会把 $1/\gamma$ 直接称为“光子寿命”, 或者把振幅衰减率和能量衰减率混用。对被动光学问题, 这种混淆可能只差一个因子 2; 但一旦进入阈值条件、线宽估计和速率方程接口, 这个因子 2 就会在不同模块之间制造系统偏差。

28.9 互易性: 为什么许多耦合公式其实来自同一个定理

正文中的耦合模理论、微扰公式和灵敏度分析, 背后都有一个共同工具, 即 Lorentz 互易性。设两组场 $(\mathbf{E}_1, \mathbf{H}_1)$ 和 $(\mathbf{E}_2, \mathbf{H}_2)$ 在同一线性、时不变、局域介质中分别满足

$$\varepsilon = \varepsilon^T, \quad \mu = \mu^T, \quad (28.56)$$

所对应的 Maxwell 方程。式 (28.56) 是这一节最关键的适用范围前提: 它排除了磁光、旋光、显式时变调制和其他非互易介质。只有在这一前提下, 下面的 Lorentz 互易性推导才按最熟悉的对称形式成立。

本节默认介质张量满足式 (28.56)。若器件包含磁光偏置、时空调制或其他非互易机制, 则互易关系必须改写, 不能直接把本节公式平移过去。

$$\nabla \times \mathbf{E}_1 = +i\omega\mu\mathbf{H}_1, \quad \nabla \times \mathbf{H}_1 = \mathbf{J}_1 - i\omega\varepsilon\mathbf{E}_1, \quad (28.57)$$

$$\nabla \times \mathbf{E}_2 = +i\omega\mu\mathbf{H}_2, \quad \nabla \times \mathbf{H}_2 = \mathbf{J}_2 - i\omega\varepsilon\mathbf{E}_2. \quad (28.58)$$

对第一式与第二组场做混合乘积, 再用向量恒等式

$$\begin{aligned} \nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) &= \mathbf{H}_2 \cdot (\nabla \times \mathbf{E}_1) - \mathbf{E}_1 \cdot (\nabla \times \mathbf{H}_2) \\ &\quad - \mathbf{H}_1 \cdot (\nabla \times \mathbf{E}_2) + \mathbf{E}_2 \cdot (\nabla \times \mathbf{H}_1), \end{aligned} \quad (28.59)$$

可得到

$$\nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) = \mathbf{E}_2 \cdot \mathbf{J}_1 - \mathbf{E}_1 \cdot \mathbf{J}_2. \quad (28.60)$$

对体积积分并使用散度定理, 就得到互易性形式

$$\int_{\partial V} (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) \cdot \hat{\mathbf{n}} dS = \int_V (\mathbf{E}_2 \cdot \mathbf{J}_1 - \mathbf{E}_1 \cdot \mathbf{J}_2) dV. \quad (28.61)$$

式 (28.61) 的价值在于, 它为“某个局域扰动为什么会以某种重叠积分方式耦合到目标模”提供了统一解释。很多看起来不同的耦合系数公式, 本质上都能追溯到互易性与边界条件。

互易性并非无条件成立。若介质存在显著非互易效应、时变调制、磁光非对称张量或非局域响应, 则式 (28.61) 需要修改, 甚至可能失效。将其直接套用到所有有源介质问题上是不严谨的。

28.10 一阶微扰公式：什么时候它是定理，什么时候它只是工程近似

现在可以回到正文中最关键的开放系统数学缺口：非厄米一阶扰动到底应怎样写。若先从最一般的角度出发，把本征问题写成

$$\hat{\mathcal{L}}(\tilde{\omega}, p) \mathbf{E}^R = 0, \quad (28.62)$$

其中 p 是几何、材料或工作点参数，那么对小扰动 $p \rightarrow p + \delta p$ 做一阶展开，可得

$$\left[\frac{\partial \hat{\mathcal{L}}}{\partial \omega} \right]_n \delta \tilde{\omega} + \delta \hat{\mathcal{L}} \mathbf{E}_n^R + \hat{\mathcal{L}}_n \delta \mathbf{E}_n^R = 0. \quad (28.63)$$

再引入左模投影，就得到开放系统的一阶扰动骨架

$$\delta \tilde{\omega}_n = - \frac{\left\langle \mathbf{E}_n^L, \delta \hat{\mathcal{L}} \mathbf{E}_n^R \right\rangle}{\left\langle \mathbf{E}_n^L, \frac{\partial \hat{\mathcal{L}}}{\partial \omega} \mathbf{E}_n^R \right\rangle}. \quad (28.64)$$

这就是开放系统里真正普适的一阶公式。它比封闭腔公式更“硬”，原因不在于形式更复杂，而在于它保留了两件开放系统中不能被省略的信息：**左/右模的双正交结构，以及算子对频率的显式依赖。**

若系统退回封闭、无损、弱开放极限，且左模与右模在厄米内积下重合，同时忽略显式色散，那么式 (28.64) 就会退化为熟悉的频移公式

$$\frac{\delta \omega}{\omega} \approx - \frac{1}{2} \frac{\int_V \delta \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}{\int_V \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}. \quad (28.65)$$

它的物理意义仍然非常直观：若正的电场能量主要落在折射率增大的区域，则共振频率通常下降。正文中关于热漂移、材料修正和几何微扰的许多定性判断，都可用这一公式快速给出方向感。

但这里必须准确区分三种层级。

第一层，**严格定理层级**：式 (28.64)。只要左模、右模、伴随算子和归一化框架是自洽的，它就是开放系统一阶扰动的正确起点。

第二层，**弱开放工程层级**：式 (28.65)。当泄露较弱、模式局域较强、色散不主导，且扰动主要发生在腔内高场区域时，它通常是非常有用的趋势公式。

第三层，**超出微扰适用层级**：近简并模强混合、模式身份交换、扰动直接重构辐射边界、材料色散和吸收显著、或电热回写已经把问题从“局部小改动”推成“空间分布重构”。此时无论厄米还是非厄米一阶公式都不应继续扛主问题，必须回到完整本征问题或全耦合重算。

若扰动是附加小吸收 $\delta \varepsilon'' > 0$ ，则同样的骨架会给出复频率虚部修正，也就是额外损耗，而非实部折射率。于是，从数学上看，“热漂移改频率”和“掺杂吸收改线宽”其实是同一套开放系统微扰公式的不同投影：前者更看重 $\Re \delta \tilde{\omega}$ ，后者更看重 $\Im \delta \tilde{\omega}$ 。

若某个 PCSEL 候选模已经处于 quasi-BIC 附近，那么轻微孔形扰动往往首先重排的是辐射通道相消条件，而不是腔内平均折射率。此时若只用式 (28.65) 看实部频移，可能会得到“改动很小”的错觉；而真正决定阈值排序的，恰恰是虚部损耗的放大变化。这里就必须至少回到式 (28.64) 的层级，甚至直接做全波重算。

易错点

用一阶微扰公式解释已经明显发生模排序翻转、模形重构或辐射通道重组的现象，通常会导致错误结论。只要扰动大到足以改变近简并模之间的耦合关系，或者大到使左/右模自身的身份发生变化，就应回到完整本征问题，而不是继续依赖一阶近似。

28.11 这些数学工具在 PCSEL 仿真中怎么实际使用

附录的价值，最终还是要回到建模实践。对本书关心的 PCSEL 问题，这些工具至少有四类直接用途。

第一，用时间因子和复频率关系检查损耗与增益符号是否在不同模块间保持一致。若 RCWA、FDTD 和器件增益模型对介质虚部的正负含义不一致，必须立即停下来修正。

第二，用 Fourier-Bloch 展开理解数值方法之间的联系。这样在比较 PWEM、RCWA 和 CWT 时，读者就知道它们差别主要在截断和投影，而不是在“互相矛盾的物理定律”。

第三，用归一化无关的重叠积分比较不同求解器结果，避免把场强数值本身的差异误判为物理差异；若必须比较开放模的绝对模幅或灵敏度，则应先确认所用的是同一种左/右模与广义归一化框架。

第四，用复频率与一阶微扰公式快速做 sanity check。例如，温度升高若使波导折射率整体增加，那么目标共振一般应向长波漂移；若仿真却给出相反趋势，就要优先检查材料模型和符号约定。

第五，对强开放候选模，不要只记录 ω_r 或 Q ，而应同时记录使用的归一化、是否采用 QNM / PML 框架、以及微扰判断使用的是厄米代理还是非厄米左/右模公式。否则后续章节即便复用了同一个符号，也未必在说同一层数学对象。

28.12 从 CWT 本征矢量提取远场偏振：方位角、椭圆率与 Stokes 参量

本节给出一组可直接写进代码的公式：如何从第十章、第十一章的四波 CWT 本征矢量读出远场偏振态。在实际器件设计中，除了关心模式的辐射常数和品质因子，还需要直接从 CWT 解中提取远场的偏振特性。设第 j 个超模的本征矢量为

$$\mathbf{V}_j = \begin{bmatrix} c_{x,j} & s_{x,j} & c_{y,j} & s_{y,j} \end{bmatrix}^T, \quad (28.66)$$

其中下标 x, y 分别表示沿 $\pm x$ 和 $\pm y$ 方向传播的基本波对。在最小四波模型中，远场主要由前向波 $c_{x,j}$ 和 $c_{y,j}$ 贡献（后向波 $s_{x,j}, s_{y,j}$ 主要形成驻波反馈）。因此，远场电场的横向分量可近似写为

$$\mathbf{E}_{\text{ff}} \approx c_{x,j} \hat{\mathbf{x}} + c_{y,j} \hat{\mathbf{y}}. \quad (28.67)$$

由此可直接定义三个关键偏振参数：

1. 偏振方位角 (azimuth angle) 线偏振主轴相对于 x 轴的夹角为

$$\psi_j = \frac{1}{2} \operatorname{atan2}(2 \operatorname{Re}[c_{x,j} c_{y,j}^*], |c_{x,j}|^2 - |c_{y,j}|^2), \quad \psi_j \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right]. \quad (28.68)$$

特别地, 当相位差接近零时, 式 (28.68) 会退化到熟悉的幅值比公式; 当 $|c_{x,j}| = |c_{y,j}|$ 且相位差为零时, $\psi_j = 45^\circ$, 对应对角线偏振; 当 $c_{y,j} = 0$ 时, $\psi_j = 0^\circ$, 对应纯 x 偏振。

2. 椭圆率 (ellipticity) 若考虑复振幅的相对相位, 先定义椭圆角 χ_j :

$$\chi_j = \frac{1}{2} \arcsin\left(\frac{2 \operatorname{Im}[c_{x,j}^* c_{y,j}]}{|c_{x,j}|^2 + |c_{y,j}|^2}\right), \quad e_j = \tan \chi_j, \quad e_j \in [-1, 1], \quad (28.69)$$

其中 $e_j = 0$ 对应完全线偏振, $|e_j| = 1$ 对应圆偏振。

3. 斯托克斯参量 (Stokes parameters) 更一般地, 可以用归一化斯托克斯参量完整刻画偏振态:

$$\begin{aligned} s_0 &= |c_{x,j}|^2 + |c_{y,j}|^2, \\ s_1 &= \frac{|c_{x,j}|^2 - |c_{y,j}|^2}{s_0}, \\ s_2 &= \frac{2 \operatorname{Re}(c_{x,j}^* c_{y,j})}{s_0}, \\ s_3 &= \frac{2 \operatorname{Im}(c_{x,j}^* c_{y,j})}{s_0}. \end{aligned} \quad (28.70)$$

这四个参量满足 $s_1^2 + s_2^2 + s_3^2 = 1$ (完全偏振光), 且在庞加莱球上构成一个单位向量。结合第八章中的对称性分析, 可以预测不同模式在不同对称破缺下的轨迹演化。

实际应用提示

在优化算法中, 可以直接将式 (28.68) 和式 (28.69) 作为目标函数的约束项。例如, 若要设计 45° 线偏振基模, 可在损失函数中加入惩罚项

$$\mathcal{L}_{\text{pol}} = w_\psi \left(\psi_j - \frac{\pi}{4}\right)^2 + w_e |e_j|^2, \quad (28.71)$$

其中权重 w_ψ, w_e 控制偏振纯度要求。这种方法比单纯依靠几何猜测更高效, 尤其在高阶模抑制和多孔耦合结构中。

28.13 群论工具箱: C_{4v} 点群完整参考

本节把第八章只按需引用的群论事实集中成一份完整参考: 生成元、特征标表、直积表、选择规则与 $C_{4v} \rightarrow C_{2v}$ 分支律。它服务于查阅而非线性阅读。

生成元与群元素: C_{4v} 由两个生成元生成:

- C_4 : 绕 z 轴旋转 90° (逆时针方向为正)

- σ_v : 关于 xz 平面 (即 $y = 0$ 平面) 的镜面对称

全部 8 个群元素为: $\{E, C_4, C_4^2 = C_2, C_4^3, \sigma_{v,x}, \sigma_{v,y}, \sigma_{d,xy}, \sigma_{d,-xy}\}$ 。其中:

- E : 恒等操作
- C_4, C_4^3 : $\pm 90^\circ$ 旋转 (这两者属于同一类)
- $C_2 = C_4^2$: 180° 旋转
- $\sigma_{v,x}, \sigma_{v,y}$: 垂直镜面 (分别关于 yz 和 xz 平面)
- $\sigma_{d,xy}, \sigma_{d,-xy}$: 对角镜面 (分别关于 $y = x$ 和 $y = -x$ 平面)

特征标表回顾与应用:

C_{4v}	E	$2C_4$	C_2	$2\sigma_v$	$2\sigma_d$	典型基函数
A_1	1	1	1	1	1	$z, x^2 + y^2, f(r)$
A_2	1	1	1	-1	-1	R_z
B_1	1	-1	1	1	-1	$x^2 - y^2$
B_2	1	-1	1	-1	1	xy
E	2	0	-2	0	0	$(x, y), (R_x, R_y)$

表中每一行的物理含义: 当对该 irrep 的基函数施加相应的对称操作时, 函数乘以对应的特征标值 (+1 保持不变, -1 改变符号)。例如, A_1 表示在所有 8 个操作下都不变——这是“完全对称”的表示; 而 E 是二维表示, 其特征标为 0 意味着两个基函数在该操作下互相交换或线性组合。

PCSEL 中的模式分类实操步骤:

1. 提取近场分布: 从仿真中获得 $|\mathbf{E}(x, y)|^2$ 或特定分量的实部/虚部
2. 应用对称操作: 对场分布依次施加 C_4 旋转和四个镜面反射
3. 记录响应: 每次操作后, 判断场是 (a) 不变 (+1)、(b) 反相 (-1)、还是 (c) 变成另一个正交模式 (二维表示的特征)
4. 查表比对: 将得到的特征标序列与特征标表逐行对比, 确定归属

快捷技巧: 实际上只需要测试三个关键操作即可唯一确定大多数情况: C_4 (区分单重态/双重态)、 σ_v (区分 A/B 型)、 σ_d (区分下标 1/2)。

基于上述流程, PCSEL 模式可分为两类:

- **单重态** (A_1, A_2, B_1, B_2): 非简并模式, 电场具有确定的宇称变换性质。这些模式往往是 BIC 候选 (若不与辐射通道 irrep 匹配)。
- **双重态** (E): 二重简并模式, 对应一对正交偏振态。在无扰动情况下能量严格相等, 是典型的亮模候选。

选择规则的群论表述：考虑任意微扰算符 $\hat{O}$ （可以是介电常数扰动、表面形变、电流注入等）诱导的两个模式 $|\psi_i\rangle$ 和 $|\psi_f\rangle$ 之间的耦合矩阵元：

$$M_{fi} = \langle \psi_f | \hat{O} | \psi_i \rangle. \quad (28.72)$$

该矩阵元非零的必要条件是三者的直积包含全对称表示 A_1 ：

$$\Gamma_i \otimes \Gamma_O \otimes \Gamma_f \supset A_1, \quad (28.73)$$

其中 Γ 表示所属的 irrep。利用 C_{4v} 的直积表可以快速判断：

$\otimes$	A_1	A_2	B_1	B_2	E
A_1	A_1	A_2	B_1	B_2	E
A_2		A_1	B_2	B_1	E
B_1			A_1	A_2	E
B_2				A_2	E
E					$A_1 \oplus A_2 \oplus B_1 \oplus B_2$

实例分析 1：四极矩辐射的选择规则

考虑四极矩辐射算符的分量 $Q_{xy} \propto xy$ 。从特征标表可知 xy 属于 B_2 表示。现在判断几组可能的跃迁：

- $A_1 \leftrightarrow B_2$ 跃迁： $A_1 \otimes B_2 \otimes B_2 = A_1 \otimes A_1 = A_1$ （对称性允许）
- $A_1 \leftrightarrow A_1$ 跃迁： $A_1 \otimes B_2 \otimes A_1 = B_2 \not\supset A_1$ （对称性禁戒）
- $B_1 \leftrightarrow E$ 跃迁： $B_1 \otimes B_2 \otimes E = A_2 \otimes E = E \not\supset A_1$ （禁戒）

这解释了为什么某些对称性保护的 BIC 无法向自由空间辐射——是因为对称性从根本上禁止了这种耦合，而非因为“损耗太小”。

实例分析 2：法向辐射通道的 irrep

在 Γ 点，法向辐射对应的是横向均匀平面波，其两支线偏振 $(\hat{x}, \hat{y})$ 构成 E 表示的基。因此：

- 若腔模属于 E 表示，则 $\Gamma_{\text{mode}} \otimes \Gamma_{\text{rad}} = E \otimes E = A_1 \oplus A_2 \oplus B_1 \oplus B_2 \supset A_1$ ——允许辐射（亮模）
- 若腔模属于 A_1 表示，则 $A_1 \otimes E = E \not\supset A_1$ ——禁止法向辐射（BIC 候选）

因此，高阶带边模式（通常是 E 表示）可表面发射，而基模（通常是 A_1 ）可能成为暗模。

对称性破缺与不可约表示分解：当结构发生某种变形或引入各向异性时，对称性会降低。设原始对称群为 G ，破缺后的子群为 $H \subset G$ 。此时 G 的每个 irrep 按照一定规则“分解”为 H 的若干个 irrep 的直和。这个过程称为**分支律** (branching rule)。

对于 PCSEL 常见的 $C_{4v} \rightarrow C_{2v}$ 破缺（例如引入椭圆形孔或矩形晶格），分解规则如下：

$$\begin{aligned} A_1(C_{4v}) &\rightarrow A_1(C_{2v}), \\ A_2(C_{4v}) &\rightarrow A_2(C_{2v}), \\ B_1(C_{4v}) &\rightarrow A_1(C_{2v}), \\ B_2(C_{4v}) &\rightarrow A_2(C_{2v}), \\ E(C_{4v}) &\rightarrow B_1(C_{2v}) \oplus B_2(C_{2v}). \end{aligned} \quad (28.74)$$

关键的物理后果：

- **双重态分裂**: E 表示分解为两个不同的单重态 B_1 和 B_2 。这意味着原本简并的一对模式会发生能量劈裂，劈裂大小正比于对称性破缺强度。
- **偏振锁定**: 分裂后的两个模式分别继承原 E 表示的两个正交基，对应固定的线偏振方向。这在实验中表现为输出光的偏振方向被”钉扎”在某个晶轴方向。
- **选择规则放宽**: 由于 C_{2v} 的特征标表更简单（只有 4 个一维 irrep），原来在 C_{4v} 中被禁戒的某些跃迁可能在 C_{2v} 中变得允许。这可以用来解释为何轻微的结构不对称会导致 BIC 泄漏出有限的辐射。

实际应用建议 在设计 PCSEL 时，可以主动利用对称性工程：

1. 若要实现稳定的单线偏振输出，故意打破 C_{4v} 对称性（如采用椭圆孔或矩形占空比）
2. 若要追求最高品质因子的 BIC，保持完美的 C_{4v} 并确保工作点是 $A_1/A_2/B_1/B_2$ 之一
3. 若要实现双波长或双偏振激光器，可以利用接近简并但略微分裂的准- E 态

28.14 关键量的首次定义位置索引

为便于读者在正文各章之间快速定位，本节给出全书最重要的派生量与合成量在何处首次引入或定义的参考索引。注意：基础物理量（如 $\mathbf{E}$ 、 ω 、 $\mathbf{k}$ 等）不列，只列出由多个基础量组合而成、或由特定物理机制引入的派生量。

符号/量	首次定义章节及位置
Γ_{opt}	第七章，式 (7.5)：光学约束因子，定义为模式与有源区的能量重叠。
$\alpha_{\text{rad}}, \alpha_{\text{int}}$	第七章，式 (7.7)：辐射损耗与内部损耗的等效空间系数。
g_{th}	第七章，式 (7.12)：阈值增益， $\alpha_{\text{rad}} + \alpha_{\text{int}}$ 与 Γ_{opt} 之比。
κ_1, κ_2	第十章，式 (10.10)：CWT 中的耦合系数，来自主导几何 Fourier 分量与纵向模重叠。
$\mathbf{C}_{1\text{D}}$	第十一章，式 (11.37)：一维反馈耦合矩阵。
$\mathbf{C}_{\text{rad}}$	第十一章，式 (11.39)：辐射耦合矩阵，非厄米，来自开放通道。
$\mathbf{C}_{2\text{D}}$	第十一章，式 (11.41)：二维高阶耦合矩阵。
v_{en}	第七章，式 (7.10)：能量速度，将空间增益/损耗映射为时间增长率的参数。
$\tilde{n}^2 = n^2 + 2ik_0 n\tilde{\alpha}$	第十一章，式 (11.5)：含增益/损耗的复折射率。
$\xi_{m,n}(z)$	第十一章，式 (11.7)：光子晶体层的 Fourier 系数。
$\phi_0(z)$	第十一章，式 (11.15)：背景层状波导的基模纵向剖面。

符号/量	首次定义章节及位置
$\mathbf{u}_{A_1}, \mathbf{u}_{B_1}, \mathbf{u}_{E,x/y}$	第十章, 式 (10.14)至式 (10.16): 四波子空间在 C_{4v} 下的 irrep 基底。
$G(z, z')$	第十一章, 式 (11.22): 辐射波的纵向 Green 函数 (近似局域均匀背景)。
$G_{m,n}(z, z')$	第十一章, 式 (11.34): 高阶波在光子晶体层内的倏逝 Green 函数。
$\delta = \beta - \beta_0$	第十一章, 式 (11.17): 偏离 Bragg 条件的传播常数偏移。
$\alpha_{\text{rad},j}$	第十一章, 式 (11.50): 第 j 模的辐射常数, $2 \text{Im}(\lambda_j)$ (亦可写为 $-2 \text{Im}(\tilde{\omega}_j)/v_{g,j}$)。
η_{N_z}	第十一章, 式 (11.4): 纵向模截断能量占比。
$\text{rad}(\omega)$	第十一章, 式 (11.53): 辐射通道引入的自能项。
Γ_{PC}	第十三章, 式 (13.26): 光子晶体层的光学限制因子, 用于二维等效介电常数修正。
$\tilde{\epsilon}_a, \tilde{\epsilon}_b$	第十三章, 式 (13.27): 送入二维 PWEM 的修正等效介电常数。
$\alpha_{\text{rad}} = 2 \text{Im}(\lambda) = -2 \text{Im}(\tilde{\omega})/v_g$	第十一章, 式 (11.50): 辐射常数与 CWT 本征值虚部 (及复频率虚部) 之间的换算关系。
I_{th}	第七章, 式 (7.15): 阈值电流, 是器件级电-热-光结果, 需经电注入模型换算。
ΔG_{*j}	第七章: 目标模与第 j 竞争模之间的模鉴别裕量。
$\partial\omega_m/\partial T, \partial\gamma_m/\partial T$	第十一章, 式 (11.61): 三维 CWT 输出的温度灵敏度导数, 用于热耦合接口。

这条索引的设计目的只有一个: 让读者在遇到某个量时, 能在 30 秒内找到它在主线上的定义位置, 而不是在各章符号之间来回猜测。本索引不追求穷举, 只覆盖”跨越两个以上章节使用”的派生量。基础符号的定义应在各章首次使用时已经明确, 此处不再重复。

回看本章

数学附录的作用是统一全书的语法, 而非堆公式。时间因子决定损耗与增益的符号, 傅里叶卷积解释了周期介质中的耦合矩阵, QNM / resonant states 说明开放模为何不是普通束缚态, 左/右模与广义归一化说明为什么开放系统不能直接照搬厄米内积, 复频率给出 Q 与光子寿命的精确定义, 互易性与一阶微扰则为耦合和灵敏度分析提供了共同骨架。只要这些约定统一, 正文中的光学、器件和仿真章节就能共享同一套数学接口; 若这些约定混乱, 后面的一切“高级模型”都可能只是表面连贯。

练习题

1. [基础] 从式 (28.1) 出发，推导在本书时间因子约定下被动介质为何对应正的 $\Im \varepsilon$ 。
2. [基础] 从能量衰减定义出发，重新推导式 (28.55)，并说明为什么振幅寿命和能量寿命差一个因子 2。
3. [进阶] 利用式 (28.13)，解释为什么保留有限个倒格矢后，本征问题会变成矩阵本征值问题。
4. [进阶] 从式 (28.64) 出发，解释为什么开放系统一阶微扰需要左模，而不能只用右模与自身的复共轭关闭投影。
5. [综合]（思路提示：先定层级，再列关键式）给出一个一阶微扰公式明显失效的 PCSEL 场景，并说明失效来自哪一条前提被破坏。

延伸阅读

- 下一章将把这些数学语法压缩成器件级建模的符号链，特别是从材料增益到阈值电流的映射。阅读时应同时核对本章变量在后续模型中的定义和适用范围。
- 若想进一步打牢开放系统与扰动理论，可对照 [28, 57] 中关于周期介质与模式扰动的附录阅读。阅读时应重点核对开放系统归一化、QNM 与非厄米微扰的标准假设。
- 若想系统补强开放系统归一化、QNM 与非厄米微扰，可继续阅读 [39, 38, 48, 60, 34, 3]。阅读时应重点核对开放系统归一化、QNM 与非厄米微扰的标准假设。

表 28.2: 本附录数学工具在正文中的应用位置

数学工具	正文应用	具体用途
时间因子约定	第 3,7,17 章	统一损耗/增益符号
Bloch-Floquet 展开	第 4,5,12 章	PWEM/RCWA 的平面波基底
QNM 双正交归一化	第 11 章	非厄米微扰公式
互易定理	第 8,13 章	远场 - 近场关联
留数定理	第 13 章	RCWA 散射矩阵极点分析
卷积矩阵构造	第 12,13 章	周期介质的傅里叶因子化
复频率- Q 关系	第 3,7,8b 章	BIC/quasi-BIC 的 Q 值定义

第二十九章 半导体激光器与器件建模附录

读完本章，你能

- 把正文中分散出现的“材料增益、模增益、阈值条件、阈值电流、热回写”整理成一条可审计的器件符号链。
- 掌握 Poisson、漂移-扩散、连续性方程和热方程在 PCSEL 器件建模中的角色分工，以及它们与光学模型的接口变量。
- 学会区分可直接测得参数、文献参数和拟合参数，并据此建立校准顺序和不确定性传播框架。

读前准备 建议已掌握 第**十七**章至第**二十**章 中关于外延层光学、量子阱增益、电注入和热-电-光耦合的内容。

阅读主线 本章所处模型层级：**L3**→**L4**

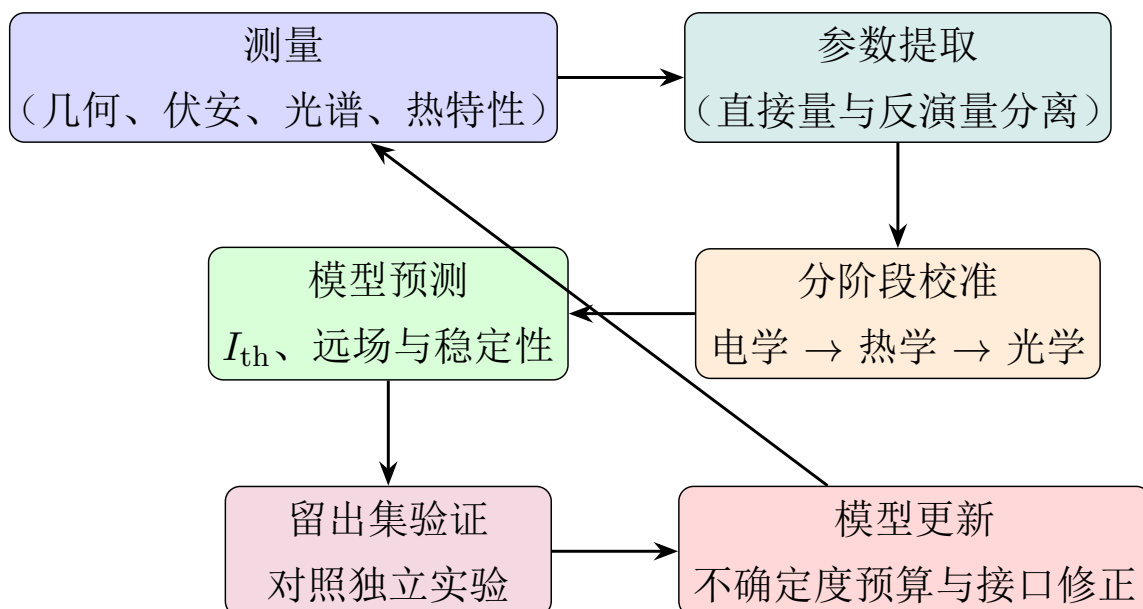


图 29.1: 器件级建模的“校准-验证”闭环：从测量和参数提取出发，经分阶段校准与预测，再由留出集验证回写模型不确定度。

建议先用图29.1对齐“测量-提取-校准-预测-回验”的闭环，再逐节阅读具体方程与接口定义。

正文第十七章至第二十章已经把器件物理分开讲过，但真正开始做器件模型时，读者仍然很容易卡在一个地方：每一章单独看都懂，一旦想把它们连起来，就不知道变量应该怎样流动。本附录就是为了解决这个问题。我们先建立从材料增益到阈值电流的符号链，说明每个箭头背后对应的是哪一类模型；随后给出最小器件方程组，并明确区分哪些式子属于光学模型，哪些属于电学模型，哪些属于热-电-光耦合模型；再整理常见边界条件、参数来源和校准顺序；然后把“测量 → 参数提取 → 分层校准 → 预测 → 回验”写成一条可执行的实验闭环；最后给出不确定性传播与结果报告方式，帮助读者把“能完成运行”提升为“能解释、能复核”。

29.1 为什么器件建模最容易在“接口”上出错

许多 PCSEL 项目在单个模块内看起来都没有大问题。光学模型给出了目标模和阈值增益，电学模型给出了注入分布，热学模型给出了温升。但一旦这些模块连在一起，结果却变得难以解释。最常见的原因是模块之间的接口没有定义清楚，而非某个模块方程错。比如，光学模块输出的是阈值增益 g_{th} ，而电学模块求的是载流子分布 $N(\mathbf{r})$ ；若没有说明用什么权重把 $N(\mathbf{r})$ 映射到模增益，就无法判断器件是否达到阈值。又比如，热学模块输出的是温度场 $T(\mathbf{r})$ ；若没有指定它如何进入折射率、增益峰位和吸收，那么温升就无法真正影响模式排序。

因此，本附录关注的重点是把接口变量和变量流向写死，而非再写更多方程。只有这样，器件模型才不会退化成“几个求解器串起来，但谁也说不清谁在驱动谁”。

直觉框

器件建模不是把所有物理都堆进一个大模型里就自动完成了。主要困难在于明确定义每个模块的输入、输出和适用边界，使整条链既能前向计算，也能反向审计。

29.2 从材料增益到阈值电流：先把符号链写全

对 PCSEL 器件而言，最核心的一条链可以写成

$$g_{mat}(N, T, \omega) \longrightarrow g_{mod,i} \longrightarrow g_{net,i} \longrightarrow \text{threshold for mode } i \longrightarrow I_{th}. \quad (29.1)$$

这里每一个箭头都代表一个独立的建模步骤，而不是单一公式的直接代换。

第一步是从材料增益到模增益。这一步需要第十七章给出的纵向与空间重叠信息，因此它已经是光学-有源区耦合问题，而非纯材料问题。

第二步是从模增益到净模增益。这里要减去的不仅是辐射损耗，还包括自由载流子吸收、重掺杂层吸收、金属邻近损耗等内部损耗，因此它同时依赖光学模式和器件层栈。

第三步是阈值判据。对每个候选模 i ，只有当净模增益刚好补偿总损耗时，才谈得上其达到光学阈值。

第四步才是从阈值所要求的载流子条件反推阈值电流。这一步不能只靠冷腔和增益模型，必须进入 Poisson 与漂移-扩散方程，因为真实器件注入下的载流子分布往往不是均匀的。

这样写的好处是，任何时候只要结果出错，都能定位到底是哪一个箭头出了问题，而不是把所有误差都模糊地归咎于“器件模型不准”。

29.3 第一步：材料增益怎样变成模增益

若先采用最简单的均匀载流子近似，把量子阱区域中的材料增益写成 $g_{\text{mat}}(N, T, \omega)$ ，则第 i 个模的模增益可近似写成

$$g_{\text{mod},i}(N, T, \omega) \approx \Gamma_{\text{QW},i} g_{\text{mat}}(N, T, \omega), \quad (29.2)$$

其中 $\Gamma_{\text{QW},i}$ 是量子阱重叠因子。式 (29.2) 属于**光学-有源区耦合模型**，不是纯电学也不是纯材料公式。

但只要注入开始明显不均匀，式 (29.2) 就只能作为一致性核查。更合理的写法应采用空间加权平均，例如

$$g_{\text{mod},i} = \frac{\int_{V_{\text{act}}} g_{\text{mat}}(N(\mathbf{r}), T(\mathbf{r}), \omega_i) W_i(\mathbf{r}) dV}{\int_{V_{\text{act}}} W_i(\mathbf{r}) dV}, \quad (29.3)$$

其中 $W_i(\mathbf{r})$ 是与模场有关的非负权函数，工程上常取与局域电场能量密度成正比的形式。式 (29.3) 的意义非常重要：一旦采用真实器件分布，阈值所需的就不再是单一数字 N_{th} ，而是一个满足加权阈值判据的空间分布。

29.4 第二步：从模增益到净模增益，再到阈值

对目标模 i ，净模增益写成

$$g_{\text{net},i} = g_{\text{mod},i} - \alpha_{\text{int},i} - \alpha_{\text{rad},i}, \quad (29.4)$$

其中 $\alpha_{\text{int},i}$ 表示内部损耗， $\alpha_{\text{rad},i}$ 表示辐射损耗。若采用复频率语言，也可以把这些损耗等价写入模衰减率或光子寿命。阈值条件就是

$$g_{\text{net},i} = 0. \quad (29.5)$$

这一步看似简单，真正容易被误用的地方在于：式 (29.5) 给出的仍然只是**光学阈值条件**，并没有自动给出阈值电流。因为要把它变成 I_{th} ，还必须回答“怎样的注入分布才能实现这个净模增益”。

在均匀近似下，人们常把式 (29.5) 先解成一个等效阈值载流子密度 N_{th} 。这可以作为设备级仿真的初值或一致性核查，但绝不能忘记：一旦器件中存在显著电流拥挤、热非均匀或泄漏通道，这个 N_{th} 就不再是器件级的完整答案，而只是后续输运模型应当达到的目标条件之一。

“阈值增益”与“阈值电流”的区别，不是单位不同这么简单。前者属于模态能量平衡问题，后者属于输运和复合驱动下的器件工作点问题。只要输运和温升没有显式进入模型，就只能宣称前者，不能宣称后者。

29.5 第三步：速率方程的一致性核查

虽然完整器件结论需要漂移-扩散模型，但速率方程近似仍然有价值，因为它给出了一条快速、透明的量纲链。

若暂时把有源区看成一个等效体积 V_{act} ，并用注入效率 η_{inj} 表示进入有效有源区的电流比例，则稳态下常见的 ABC 近似写成

$$\frac{\eta_{\text{inj}} I}{q V_{\text{act}}} = AN + BN^2 + CN^3 + R_{\text{st}}, \quad (29.6)$$

其中 A, B, C 分别代表 SRH、辐射和 Auger 复合系数， R_{st} 表示受激辐射等效项。阈值附近由于受激辐射刚开始出现，常把阈值电流近似写成

$$I_{\text{th}} \approx \frac{q V_{\text{act}}}{\eta_{\text{inj}}} (AN_{\text{th}} + BN_{\text{th}}^2 + CN_{\text{th}}^3). \quad (29.7)$$

式 (29.7) 不是完整器件模型，它只是**器件级一致性核查**。它的作用在于：当漂移-扩散模型给出某个 I_{th} 区间时，读者可以迅速判断这个结果是否与等效有源区体积、注入效率和复合量级大体相容。若二者相差几个数量级，优先应怀疑参数、边界或单位，而不是急着解释“这是某种新物理”。

若光学阶段估计目标模所需 N_{th} 并不高，但完整器件模拟却给出异常高的 I_{th} ，最先应检查的是注入效率、漏电路径和电流拥挤，而不是先去怀疑增益模型本身。因为式 (29.7) 告诉我们，高阈值电流并不一定意味着高阈值增益，它也可能只是电流没有有效送进目标有源区。

29.6 第四步：完整器件方程组是什么，各自属于哪类模型

当需要超越速率方程近似时，就必须进入器件偏微分方程模型。最常见的一组方程由 Poisson 方程、电子和空穴连续性方程、漂移-扩散电流关系以及热方程组成。它们可以写成

$$\nabla \cdot (\epsilon_s \nabla \phi) = -q(p - n + N_D^+ - N_A^-), \quad (29.8)$$

$$\frac{\partial n}{\partial t} = \frac{1}{q} \nabla \cdot \mathbf{J}_n + G_n - R_n, \quad (29.9)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \nabla \cdot \mathbf{J}_p + G_p - R_p, \quad (29.10)$$

$$\mathbf{J}_n = q\mu_n n(-\nabla \phi) + qD_n \nabla n, \quad (29.11)$$

$$\mathbf{J}_p = q\mu_p p(-\nabla \phi) - qD_p \nabla p, \quad (29.12)$$

$$\rho C_{\text{th}} \frac{\partial T}{\partial t} = \nabla \cdot (k_{\text{th}} \nabla T) + Q_{\text{Joule}} + Q_{\text{nr}} + Q_{\text{opt}}. \quad (29.13)$$

其中光学热源 Q_{opt} 包含多种物理机制：

$$Q_{\text{opt}} = \underbrace{\alpha_{\text{FCA}}(N, T) I_{\text{opt}}}_{\text{自由载流子吸收}} + \underbrace{\alpha_{\text{interband}} I_{\text{opt}}}_{\text{带间吸收}} + \underbrace{\alpha_{\text{metal}} I_{\text{opt}}}_{\text{金属邻近吸收}} + \underbrace{\alpha_{\text{scatter}} I_{\text{opt}}}_{\text{散射吸收}}. \quad (29.14)$$

这里 I_{opt} 是局域光强， α_{FCA} 由第十七章的自由载流子吸收公式给出， $\alpha_{\text{interband}}$ 在光子能量超过带隙时显著， α_{metal} 描述金属接触或等离子体增强结构导致的吸收， α_{scatter} 则来自粗糙

侧壁和缺陷引起的散射损耗。在典型 PCSEL 工作条件下, 自由载流子吸收通常是主导项, 但在特定结构 (如表面等离子激元增强或超薄膜器件) 中, 金属吸收和带间吸收可能不可忽略。

为了避免概念混乱, 必须立刻说明这些方程的类别。

式 (29.8) 至式 (29.12) 属于**电学输运模型**。它们回答的是电势、载流子和电流怎样在器件中分布。

式 (29.13) 属于**热模型**。它回答的是温度场怎样由 Joule 热、非辐射复合热和光吸收热驱动出来。

而把这些结果写回折射率、材料增益与吸收, 即

$$\Delta n = \Delta n(N, T), \quad \Delta g = \Delta g(N, T, \omega), \quad \Delta \alpha = \Delta \alpha(N, T), \quad (29.15)$$

则属于**热-电-光耦合模型**, 因为它把电学和热学结果重新送回光学模块。

29.7 边界条件不是实现细节, 而是器件物理的一部分

对器件方程来说, 边界条件几乎和控制方程一样重要。一个非常实用的写法, 是在建模笔记中把每一条边界都写成“边界类型 + 数学式 + 物理含义”的三元组。例如:

对电学模型, 欧姆接触通常意味着电势或准费米能级在接触处被固定, 或者总电流被指定; 绝缘边界则对应法向电流为零。对热模型, 理想热沉可用固定温度近似, 对流边界可写成

$$-k_{\text{th}} \nabla T \cdot \hat{n} = h(T - T_{\infty}), \quad (29.16)$$

其中 h 是等效换热系数。若界面热阻不可忽略, 还需要写成温度跳变与热流之间的关系。对光学模型, 开放边界不能用理想反射边界替代, 而应使用 PML、辐射边界或端口条件。

教材上最容易被忽略的一点是: 边界条件并非“软件设置页最后再填的选项”, 而是物理模型本身的一部分。没有边界条件, 就没有完整模型。更重要的是, 边界条件往往对应整个器件设计中最不确定的一部分, 例如接触质量、封装散热和外界反射环境。因此, 它们同时也是不确定性预算的重要来源。

29.8 参数来源要分层管理: 测量值、文献值和拟合值不是同一种可信度

器件模型参数很多, 但它们的可信度并不相同。最稳妥的做法, 是把参数分成三层。

第一层是**直接测量参数**, 例如层厚、几何尺寸、某些 I-V 或热阻数据。这类参数优先级最高, 因为它们与当前器件最接近。

第二层是**文献参数**, 例如某一材料体系在相近掺杂和温度区间下的迁移率模型、SRH 寿命经验范围或折射率色散参数。这些参数可用于起始建模, 但必须附带文献环境信息, 因为材料体系和工艺差异可能很大。

第三层是**反演拟合参数**, 例如接触电阻率、界面热阻或某些难以直接测得的增益压缩系数。它们是必须明确标记, 并做鲁棒性扫描, 防止把“为了拟合当前曲线引入的自由度”误当成可外推的材料常数, 而非不能用。

表 29.1: 器件模型中常见参数家族及其主要来源。表中给出的是参数类型与建模角色，而不是可直接引用的数据库数值。

参数家族	常见来源	在模型中的主要作用
层厚、孔径、电极尺寸	工艺测量、截面表征	决定纵向模、单胞辐射、注入路径与热扩散路径
折射率、色散、吸收	文献 + 光谱拟合	决定模式频率、重叠、辐射与内部损耗
增益参数、透明密度、微分增益	文献 + 增益模型拟合	决定从载流子到模增益的映射
迁移率、复合系数、接触参数	文献 + I-V / L-I 校准	决定电流扩展、串联电阻与阈值电流来源
热导率、界面热阻、换热系数	文献 + 热阻或温升校准	决定稳态温升和热漂移反馈

这张表之后最需要补的一句，是参数出处的优先级说明：几何真值优先于名义设计表，直接测量参数优先于文献平均值，文献平均值优先于反演拟合值。若某个参数只能来自反演拟合，就应在正文或附表中明确写出它依赖哪些观测量被校准出来，以及它是否只在当前器件结构与温度区间内有效。否则，读者很容易把“为解释一组曲线而引入的自由度”误当成普适材料常数。

29.9 校准顺序为什么通常是先电学，再热学，最后光学回写

当模型参数不完备时，校准顺序会极大影响结论可解释性。一个经验上非常稳健的顺序是：先校准电学，再校准热学，最后做光学回写。原因是，如果一开始就试图用光学参数去吸收电学和热学的误差，很容易出现“拟合上去了，但变量物理意义全乱了”的情况。

更具体地说，首先应利用 I-V 数据、接触结构和几何信息，把串联电阻、注入效率和漏电路径调到合理范围。接着再用热阻、光谱热漂移或热像信息去约束热边界和热源强度。只有在电和热都大致满足基本约束后，才适合利用阈值漂移、共振波长漂移和输出特性去微调增益或折射率温度系数。

这个顺序来自变量依赖关系：电学决定热源分布，热学决定折射率和增益回写，而光学响应受前两者共同驱动。若倒置这个顺序，光学拟合参数往往会吸收本应由电学或热学参数解释的误差。

29.10 测量 → 参数提取：先区分直接观测量与反演量

实验闭环的第一步，是把手头数据按“直接观测量”和“需要模型反演的量”分开，而非立刻拟合。直接观测量包括截面层厚、孔径、刻蚀深度、电极尺寸、I-V 曲线、某些热阻数据、远场角宽、偏振纯度和谱峰位置。它们应优先直接进入参数账本，因为这些量几乎不依赖模型假设。反过来，接触电阻率、界面热阻、有效注入效率、非辐射复合系数和某些增益压缩参数，则往往是通过模型解释实验时才被反演出来的量，而非直接测量量。

把这两类量混在一起，是器件建模最常见的失控点。因为一旦把反演量伪装成直接测量量，后续所有不确定性都会被低估。更稳妥的做法是明确写出

$$\mathcal{M}_{\text{obs}} \longrightarrow \mathcal{P}_{\text{direct}} \cup \mathcal{P}_{\text{infer}}, \quad (29.17)$$

其中 $\mathcal{M}_{\text{obs}}$ 是原始观测集合， $\mathcal{P}_{\text{direct}}$ 是可直接进入模型的参数集合， $\mathcal{P}_{\text{infer}}$ 是必须在校准过程中反演的参数集合。式 (29.17) 的意义并不在于数学新奇，而在于它强迫读者先回答：“我现在到底是在用实验给模型喂确定信息，还是在用模型解释实验来推回未知参数？”

表 29.2: 器件建模实验闭环中的典型“测量量 $\rightarrow$ 参数提取”映射。重点不在于把每个参数都直接测出来，而在于明确哪些量是观测量、哪些量是反演量。

测量量	优先约束的参数或子模型	说明
截面 SEM / AFM / 工艺记录	层厚、孔径、刻蚀深度、电极几何	这是几何真值的第一来源，优先级高于名义设计表。
I-V 曲线与串联电阻斜率	接触边界、漏电路径、spreading layer 电阻	主要用于校准电学子模型，而不是拿来“顺便”调光学参数。
L-I 曲线与阈值漂移	注入效率、复合参数区间、热回写强度	只能在电学与热学已基本满足基本约束后进入。
谱峰位置及其温漂	有效折射率温度系数、增益谱偏移	主要约束热回写到光学的接口。
远场、偏振和近远场对照	候选模身份、偏振选择、辐射通道权重	更适合作为回验量，而不是一开始就用于大自由度拟合。
热像或热阻测试	界面热阻、换热边界、热源分布代理	用于锁定热学边界，而不是直接替代器件内部温度场。

若这类表在具体项目中进一步扩展成材料参数表，还应在表后附一段“数据来源与适用条件”。最稳妥的写法是明确注明：折射率数据来自哪类数据库或文献、适用于哪一温度和掺杂范围，而非只报一个数字；增益参数基于哪类量子阱结构；热导率是室温低掺杂参考值还是已包含重掺杂退化。任何材料参数表都应至少同时带“来源、温度、掺杂、平台”四个标签。只有把这些前提写出来，后面的校准和回验才不会把参考值误读成跨平台、跨温区的普适常数。

材料参数的适用边界 重要提示：表中数值若只是室温、低掺杂、名义结构下的参考量级，就必须直说。尤其是折射率温度系数、热导率和增益参数，通常都随温度、掺杂和应变明显变化。把它们当成无条件常数，是器件附录里最容易埋下的后续失真源。

热导率的温度依赖性

半导体材料的热导率 $\kappa(T)$ 强烈依赖温度。对 GaAs、InP 等 III-V 化合物，在室温及以上温区可采用以下经验公式 [18, 1]:

$$\kappa_{\text{GaAs}}(T) \approx \kappa_{300} \left(\frac{T}{300 \text{ K}} \right)^{-\alpha}, \quad \kappa_{\text{InP}}(T) \approx \kappa_{300} \left(\frac{T}{300 \text{ K}} \right)^{-\beta}, \quad (29.18)$$

其中 κ_{300} 是 300 K 时的热导率，指数 $\alpha \approx 1.2\text{--}1.3$ (GaAs), $\beta \approx 1.0\text{--}1.1$ (InP)。这说明：

- 温度从 300 K 升至 400 K 时，GaAs 热导率下降约 25%–30%
- 温度从 300 K 降至 200 K 时，GaAs 热导率上升约 40%–50%
- 重掺杂会进一步降低热导率 (声子 - 杂质散射)，典型退化因子 0.5–0.8

对高精度热仿真，建议在材料参数表中明确 $\kappa(T)$ 的函数形式而非单一常数。若只使用室温值 κ_{300} ，需评估由此引入的误差：对 $\Delta T \approx 50 \text{ K}$ 的温升，忽略温度依赖性会导致热阻低估约 10%–15%。

若某个设计的阈值电流明显高于冷腔和增益链条的预期，正确做法是先检查 I-V 与接触结构是否说明了额外串联电阻或漏电路径，而非立刻去改增益模型；若这些都合理，再看热阻与谱漂移是否提示温升过高；最后才轮到怀疑增益模型本身。这就是“先测量分层，再决定谁该被校准”的含义。

29.11 预测 → 回验：如何判断模型真的具备外推能力

一个器件模型是否可靠，不能只看它能否重现实验中已经参与拟合的曲线。关键的是：在校准阶段没有直接使用过的观测量，模型是否仍能解释。于是，实验闭环至少应分成两批量：

1. 校准量：用于锁定几何、电学、热学和部分光学接口；
2. 回验量：保留出来，不参与拟合，只用于检查模型是否具备预测能力。

对 PCSEL 来说，一个很自然的分层闭环是：

$$\begin{aligned} &\text{测量} \rightarrow \text{参数提取} \rightarrow \text{电学校准} \rightarrow \text{热学校准} \\ &\rightarrow \text{光学回写} \rightarrow \text{回验未参与拟合的远场、偏振、阈值漂移或工作窗口}. \end{aligned} \quad (29.19)$$

式 (29.19) 的实践意义很直接。若模型只会重现已经用于拟合的 I-V 与某一条 L-I 曲线，却无法解释同一器件的谱漂移、远场主瓣变化或偏振窗口，那么它就还不具备“器件预测模型”的资格。相反，若模型在未参与拟合的观测量上仍能给出正确趋势和合理量级，即便数值上还有误差，也说明它抓住了主导物理。

因此，在报告器件建模结果时，最值得坚持的一条写法是：明确声明哪些实验量被用于校准，哪些实验量被保留为回验。只要这一步没有写清楚，“模型与实验一致”这句话就没有足够的信息量。

29.12 仿真-实验对标协议：如何让数字与测量对话

仿真的终极价值不在于自治，而在于能够预测和解释实验。然而，大量 PCSEL 仿真工作在迈向实验对标的第一步就失败了：要么比较的对象不对等，要么忽略了关键的不确定度来源，要么用事后拟合冒充事前预测。本节给出一套严格的仿真 - 实验对标协议，确保每一个“吻合良好”的宣称都经得起推敲。

29.12.1 对标前的层级对齐检查

在将仿真结果与实验数据放在一起比较之前，必须先完成以下五项层级对齐检查：

实验对标要求

仿真 - 实验对标的五项前置检查

1. **几何保真度检查**：仿真中的几何参数（晶格常数 a 、孔径 r/a 、刻蚀深度 h_{etch} 、外延层厚）是否与 SEM/TEM 实测一致？若使用名义设计值而非实测值，必须说明理由并评估敏感度。
2. **材料模型真实性检查**：折射率和增益谱是来自实测椭偏/PL 数据，还是仅靠文献经验公式？对有源区，是否包含了载流子浓度依赖的增益压缩和折射率变化？
3. **边界条件等价性检查**：仿真的是无限周期单胞、有限尺寸阵列还是完整器件？实验测得的是面发射主瓣、边缘发射还是积分球总功率？两者是否在相同的物理量层面进行比较？
4. **工作点对齐检查**：仿真的是阈值处、阈值以上固定倍数还是最大功率点？实验数据是在 CW 还是脉冲条件下测得？若实验未明确热沉温度和驱动条件，仿真不应直接声称“绝对阈值电流吻合”。
5. **不确定度量化检查**：仿真侧是否提供了收敛误差估计？实验侧是否给出了仪器精度、样品间涨落和重复性误差？若任何一方缺失不确定度信息，比较时应降级为“定性趋势对比”而非“定量验证”。

可能出错处

最常见的对标失效案例是：仿真给出的是冷腔 Q 值换算的“阈值增益”，实验给出的是 LIV 曲线上的转折点对应的“注入电流”。这两个量虽然相关，但中间隔着载流子输运效率、内量子效率、串联电阻和非辐射复合等多个未独立标定的环节。直接将二者画在同一张图上并声称“验证了阈值模型”，属于典型的范畴错误。

29.12.2 三级对标体系：从定性到定量

根据可用信息的完整程度和对标目标的差异，建议采用三级对标体系：

Level 1: 定性趋势对标（最低门槛） 这一级只要求仿真正确预测参数变化的方向，不要求幅度准确。典型问题包括：

- 孔径增大时，阈值是先降后升还是单调变化？
- 刻蚀加深时，远场主瓣变窄还是变宽？
- 温度升高时，激射波长红移速率是多少 nm/K？

通过标准：仿真预测的趋势符号与实验一致，且不存在反例。

报告规范

Level 1 对标的规范表述示例”如图 X 所示，仿真预测随着占空比从 0.3 增加到 0.6，A 模的辐射效率单调上升，这与 PL 淬灭校正后的 EL 强度趋势一致（Pearson 相关系数 $r = 0.87$ ）。需强调这是定性趋势验证，不涉及绝对效率标定。”

Level 2: 半定量范围对标（推荐门槛） 这一级要求仿真给出的数值落在实验测量的不确定度范围内，或至少在同一数量级。典型问题包括：

- 仿真阈值电流密度 $J_{th,sim}$ 是否处于实验值的 $\pm 15\%$ 以内？[行业标准] 对成熟工艺平台， $\pm 30\%$ 通常被认为是前期设计的可接受窗口，但发表级对标建议达到 $\pm 15\%$
- 远场发散角的仿真值与 CCD 拟合值偏差是否小于 5° ？
- 模式波长差的仿真预测与光谱仪分辨率相比是否显著？对波长绝对值，建议要求 $\Delta\lambda < \pm 0.5 \text{ nm}$ [行业标准]

通过标准：关键量的相对偏差 $\delta = |X_{sim} - X_{exp}|/X_{exp}$ 应满足：

- 阈值电流 I_{th} : $\delta < 15\%$ (发表级) 或 $\delta < 30\%$ (前期设计)
- 激射波长 λ : $\delta < 0.5 \text{ nm}$ 或 $\delta < 0.1\%$
- 斜率效率 η_d : $\delta < 20\%$
- 远场发散角 θ : $\delta < 5^\circ$ 或 $\delta < 10\%$

所有偏差均应可由已知不确定度源 (材料参数误差、几何尺寸公差、温度漂移等) 合理解释。

实验对标要求

Level 2 对标的 uncertainty 预算模板若仿真值为 $X_{sim} \pm \sigma_{sim}$ ，实验值为 $X_{exp} \pm \sigma_{exp}$ ，则兼容性判据为：

$$|X_{sim} - X_{exp}| \leq k \sqrt{\sigma_{sim}^2 + \sigma_{exp}^2},$$

其中 k 是覆盖因子。对近似正态、双侧区间，工程上常取 $k \approx 2$ ；严格 95% 对应 $k = 1.96$ 。本书将“ $k \approx 2$ ”标记为 [工程经验] 近似，若需发表级 uncertainty 报告应明确采用的统计分布与置信定义。若上述不等式不成立，则需要检查是否存在系统性偏差（如材料参数

失配或未建模物理机制)。

Level 3: 定量预测对标 (严格门槛) 这一级要求在没有事后拟合的前提下, 仿真预测与实验数据在误差棒内完全吻合。这通常需要:

- 所有输入参数均来自独立表征 (SEM、椭偏、Hall 效应、DLTS 等)
- 仿真模型包含所有已知的一阶和二阶物理机制
- 进行了盲测验证: 用一部分实验数据校准模型, 然后用另一部分数据进行预测检验

通过标准: reduced chi-square $\chi^2_{\nu} = \frac{1}{N-p} \sum_i \frac{(y_{\text{sim},i} - y_{\text{exp},i})^2}{\sigma_i^2}$ 接近 1 (理想范围 0.8–1.2), 其中 N 是数据点数, p 是拟合参数个数。

可能出错处

Level 3 对标的常见陷阱是“过度调参”: 通过调整多个自由参数 (如有源区掺杂浓度、缺陷态密度、接触电阻等) 使得仿真曲线与实验点完美重合, 但这些参数的物理合理性未经独立验证。这种做法得到的“吻合”只是数学拟合, 不具备预测能力。正确的做法是报告每个可调参数的物理约束范围和敏感性分析。

29.12.3 关键物理量的对标技术规范

针对 PCSEL 特有的几类关键测量, 以下是具体的对标技术要点:

29.12.3.1 LIV 曲线的对标

LIV (Light-Current-Voltage) 特性是激光器最基本的宏观指标。对标时要注意:

- **阈值提取方法一致性:** 实验上常用拐点法、线性拟合法或二次微分法提取 I_{th} , 仿真上也应采用相同算法处理 $L-I$ 曲线, 而不是直接用冷腔 Q 值推算。
- **热滚降 (thermal rollover) 建模:** 若实验观察到高温或大电流下的功率饱和/下降, 仿真必须包含自加热效应和温度依赖的效率下降机制 (如 Auger 复合、载流子泄漏)。
- **串联电阻校准:** $V-I$ 曲线的斜率应与实验 extracted 的串联电阻 R_s 一致。若偏差超过 20%, 需要检查接触模型和掺杂分布。

29.12.3.2 光谱特性的对标

激光波长和边模抑制比 (SMSR) 的对标要点:

- **波长温度漂移:** 实验测得的 $d\lambda/dT$ 应与仿真中材料色散 + 热光效应的综合预测一致。[文献值] 对 InGaAsP/InP 相关器件, 常见量级约为 0.05–0.1 nm/K, 但需绑定工作电流与温区条件 [78]。

- **SMSR 的定义统一：**实验 SMSR 通常定义为主模与最强边模的功率比 (dB)，仿真也应在相同带宽和积分时间内计算，避免因频谱分辨率不同导致的人为高估。
- **动态展宽效应：**若实验光谱仪分辨率有限（如 $> 0.1 \text{ nm}$ ），仿真应考虑线宽展宽机制（相位噪声、模式跳动），否则可能出现“仿真线宽远窄于实验”的假性不符。

29.12.3.3 远场图案的对标

远场辐射图案的对标最具挑战性，因为其对几何误差极为敏感：

- **角度坐标系统一：**确认实验使用的是立体角 (θ, ϕ) 还是直角坐标 (x, y) 探测器平面，并进行适当的 Jacobian 变换。
- **背景扣除：**实验远场常包含杂散光和基底辐射，对标前应减去暗背景和自发辐射背景。
- **有限尺寸效应：**若仿真用的是无限周期近似，而实验器件只有几十个周期的阵列，衍射旁瓣的位置和强度会有显著差异。此时应改用有限尺寸全波仿真或直接使用 Fraunhofer 衍射公式修正。

交叉验证协议

远场对标的快速诊断流程

1. 比较主瓣 FWHM：若仿真值明显小于实验值，可能是阵列尺寸过大或孔径均匀性假设过理想
2. 比较旁瓣位置：若旁瓣角度系统性偏移，可能是晶格常数或波长输入有误
3. 比较偏振消光比：若实验 PER 远低于仿真值，可能存在应力双折射或加工不对称
4. 旋转平均 vs 方位分辨：确认实验是否做了角度平均，避免将各向异性仿真与各向同性实验直接比较

29.12.4 当仿真与实验不符时的诊断树

即使完成了上述所有检查，仿真与实验出现分歧仍是常态。这时应避免两种极端反应：一是盲目怀疑仿真，二是轻易否定实验。建议采用如下系统化诊断流程：

1. **复查输入参数：**逐一核对所有几何和材料参数，特别注意单位制转换 ($\text{nm}/\mu\text{m}/\text{eV}/\text{meV}$)、温度条件和批次间工艺变异。
2. **简化测试：**构建一个极度简化的基准案例（如单层平板波导、理想 PEC 边界），验证求解器本身是否正确配置。
3. **网格/基函数收敛再确认：**有时初始收敛测试不够充分，特别是在引入新物理场（如热耦合）后需要重新验证。

4. **敏感性分析**: 对可疑参数做扫描, 观察是否存在某个参数的微小变化就能解释仿真 - 实验差距。这可能揭示未被充分认识的关键机制。
5. **独立工具复核**: 用另一种数值方法 (如 FDTD vs FEM) 或另一款软件重算同一结构, 排除单一工具的 bugs 或实现细节影响。
6. **请求原始数据**: 若实验数据来自合作组, 索要原始测量文件和校准记录, 有时数据处理过程中的平滑、滤波或归一化会引入人为 artifacts。

建立一个”失败日志”文档, 详细记录每一次仿真 - 实验不符的案例、排查过程和最终原因。长期积累下来, 你会发现某些类型的偏差反复出现 (如始终低估阈值电流 20%), 这往往指向某个被忽视的系统性误差源 (如接触电阻模型过于理想)。

29.12.5 对标结果的报告伦理

最后, 必须强调对标报告的学术诚信底线:

学术诚信提醒

仿真 - 实验对标的红线准则

- **禁止选择性呈现**: 不得只展示吻合良好的数据点而隐藏明显偏离的点。若某些条件下的对标失败, 应如实报告并讨论可能原因。
- **禁止事后调参冒充预测**: 若使用了实验数据来确定模型参数, 必须声明这是”拟合”而非”预测”, 并用独立的测试集进行验证。
- **禁止模糊不确定度**: 不得用光滑曲线连接稀疏的实验点制造”完美跟踪”的视觉错觉。应清晰标示误差棒和数据点密度。
- **禁止跨层级比较**: 不得将冷腔仿真结果直接与器件级 LIV 曲线放在同一纵轴上比较, 除非明确说明了中间的转换假设。

29.13 不确定性传播: 不要只给单值, 要给证据范围

一旦模型里存在文献参数和拟合参数, 就不应只给出单一预测值, 而应报告一个带条件的区间。最常见的一阶传播写法是

$$\delta y^2 \approx \nabla_p y^T \nabla_p y, \quad (29.20)$$

其中 $\mathbf{p}$ 是参数向量, $\mathbf{p}$ 是其协方差矩阵, y 可以是阈值电流、波长漂移、光束偏角或温升等输出量。若缺乏完整协方差, 也应至少进行区间扫描和灵敏度排序, 明确指出哪些参数真正主导了结论。

从报告角度看, 这意味着结论最好写成”在给定参数区间与边界条件假设下, 某输出落在什么范围”, 而不是只报一个最优单值。对器件设计而言, 这种写法更诚实, 也更有工程意义, 因为它直接体现设计对工艺和参数漂移的鲁棒性。

易错点

将一个在狭窄拟合条件下获得的参数值直接外推到不同温度、不同几何或不同注入区间，是器件模型最常见的失真来源之一。尤其当某参数本来只是“补偿误差的拟合旋钮”时，它更不应被当作材料常数传播到其他工作点。

29.14 回到 PCSEL：可交付器件结论的组成

到这里，可以把器件建模的真正交付物说清楚。一个合格的 PCSEL 器件模型不应只交付阈值电流数值，还应给出可审计的证据链：

它应当明确给出外延层与目标模式之间的重叠信息；明确材料增益如何映射成模增益；明确哪一类损耗主导了阈值；明确电流如何分布到有源区；明确热源与热点在哪里；明确温升如何回写到折射率、增益和吸收；最后明确在这些回写后，目标模是否仍在工作窗口内保持优势。

只要这条证据链缺少其中一环，结论就应主动降级。反过来说，若这条链完整，即便最终结果只是“某设计方案在当前参数不确定度下不够鲁棒”，它仍然是高质量器件建模，因为它解释了失败发生在哪里。

回看本章

器件建模的核心，是把“材料增益 -> 模增益 -> 净模增益 -> 阈值条件 -> 阈值电流 -> 热回写”组织成一条变量流向清晰的证据链，而非把更多方程堆在一起。速率方程近似适合做一致性核查，完整器件结论则必须依赖 Poisson、漂移-扩散和热方程。边界条件是模型本身的一部分，参数来源必须分层管理，校准顺序应优先稳住电学和热学，再做光学回写；而一套真正可交付的器件模型，还必须把“测量 -> 参数提取 -> 分层校准 -> 预测 -> 回验”写成显式闭环。真正有用的器件模型，不只给一个数值，还要告诉读者这个数值是怎样来的、由哪些参数主导、在什么条件下会失效。

练习题

1. [基础] 结合式 (29.1)，写出你当前研究对象中每一个箭头对应的输入、输出和近似条件。
2. [基础] 说明为什么式 (29.7) 只能作为器件级一致性核查，而不能替代完整漂移-扩散模型。
3. [进阶] 为一个电注入 PCSEL 设计一套参数分层表，区分直接测量、文献引用和拟合反演三类参数。
4. [进阶] 假设热回写后目标模排序翻转，写出你会优先检查的三个接口变量，并说明它们分别属于光学、电学还是耦合模型。

5. [综合]（思路提示：先定层级，再列关键式）为一个你熟悉的器件案例写出一条“测量 -> 参数提取 -> 分层校准 -> 预测 -> 回验”的最小实验闭环，并指出哪一项观测量最适合保留为回验量。

延伸阅读

- 下一章将把本附录中的变量链变成数值验证和发布门槛，也就是从“能建模”走向“能证明模型可信”。阅读时应同时核对本章变量在后续模型中的定义和适用范围。
- 若想进一步对照半导体激光器器件建模的一般框架，可结合 [10, 80, 20, 74] 中关于增益、器件设计与实验对接的内容一起阅读。阅读时应重点核对相关模型假设、参数范围与文献适用条件。

第三十章 仿真附录：检查清单与流程表

读完本章，你能

- 理解为什么“收敛、守恒、交叉验证和版本记录”本身就是物理结论的一部分，而不是报告末尾的形式主义附录。
- 掌握 RCWA、FDTD 和 FEM 三类数值方法各自最关键的验证逻辑，并知道它们在 Lumerical 与 COMSOL 风格 workflow 中的典型落点。
- 学会把仿真结果分成“探索性、已验证、可发布”三个层级，用明确门槛控制结论升级。

读前准备 建议已掌握 第十四章、第十五章、第二十五章和第二十七章 中关于方法分工、完整 workflow 和常见失败模式的内容。

阅读主线 本章所处模型层级：L1→L4

前文已经讲清了 PCSEL 应该怎样建模，但要让模型成为可信证据，还需要最后一层工作，即验证与发布门槛。本附录先说明为什么验证是物理论证本身，而非清单化装饰；随后分别讨论 RCWA、FDTD 和 FEM 的验证逻辑，重点解释每一项检查到底在防什么误判；再把这些方法层面的检查提升到 workflow 层面，形成从冷腔到器件级的阶段门槛；最后给出一个通用排错决策树，帮助读者在结果异常时按正确顺序排查，而不是一开始就盲目调网格。

30.1 为什么验证不是额外工作，而是物理论证本身

对 PCSEL 这类开放、多尺度、多物理场问题来说，一次仿真从来不只是“给出一个数值”。它同时还暗含了大量前提：边界是否真正开放，网格是否足够细，时间窗是否足够长，参数版本是否一致，模型是否仍处在近似适用范围内。若这些前提没有被验证，那么输出的数值就不能算物理证据，只能算求解器在某组设置下给出的一个候选答案。

这也是为什么本附录不把验证写成孤立 checklist。需要掌握的是每一项检查对应的物理误差来源。例如，能量守恒检查并非为了让表格更完整，而是为了防止把人工边界反射误判成高 Q ；阵列尺寸收敛是为了确认你研究的确实是有限器件物理，而不是一个偶然选中的尺寸共振点，而非为了“多跑几轮显得更认真”。

直觉框

仿真验证的目标是说明当前证据链支持的物理结论在常见数值误差和建模误判下仍保持不变，而非证明求解器能够运行。

30.2 通用验证逻辑：先问收敛，再问不变量，再问模型边界

无论使用哪一种求解器，验证的顺序都不应是随意的。最稳妥的做法是按三层推进。

第一层是**数值收敛**。也就是检查当离散尺度继续加密时，关键输出是否趋于稳定。这一层防止的是“纯离散误差被误读成物理趋势”。

第二层是**不变量或守恒量检查**。例如无源周期散射中的功率平衡、模排序对无关归一化的稳健性、以及同一物理结论在不同激励方式下是否保持。这一层防止的是“设置看起来收敛，但其实在解错误问题”。

第三层是**模型边界检查**。也就是说，即便数值已经收敛，也要问当前模型的物理前提是否仍然成立。比如，二维有效折射率模型对某结构也许已经数值收敛，但若纵向强矢量效应开始主导，它仍然不能支撑最终三维器件结论。

只有这三层都站住，结果才有资格升级。若只通过第一层而没过第二层或第三层，结果通常仍然只能算探索性结果。

30.3 RCWA 验证：验证对象是单胞开放散射，而不是某条窄谱线

RCWA 的优势在于它能高效处理周期单胞的开放散射，但这也意味着它最容易被误读成“只要谱线足够尖就说明模式足够好”。正确的验证逻辑应当围绕四件事展开。

第一，傅里叶截断是否收敛。也就是说，随着保留倒格矢数目增加，关键量例如共振频率、上下出光比例和目标模与竞争模的排序是否稳定。若排序本身仍会随截断阶数变化，那么任何后续结论都应当暂停。

第二，层状切片和几何逼近是否足够准确。对有刻蚀深度、斜侧壁或复杂层栈的结构，RCWA 常需把纵向轮廓分层近似。此时不仅傅里叶阶数要收敛，纵向切片数也要收敛。否则你看到的谱线变化，可能只是几何逼近误差。

第三，功率平衡是否成立。对无源结构，反射、透射与吸收应满足功率守恒。若不守恒，优先应怀疑截断、材料模型或归一化，而不是先解释谱特征。

第四，谱特征是否真和目标模对应得上。一个高质量 RCWA 结果，不应只给一张反射谱，还应能说明某个谱特征与哪类 Gamma 点候选模相对应，以及该特征是否与前面的 PWE/CWT 结论同向。

若把这些检查映射到软件路径上，在 Lumerical 的 RCWA 工作流里，重点通常是傅里叶阶数、入射角/偏振设置和材料层切片一致性；在 COMSOL 的单胞频域 Wave Optics 路径里，虽然实现框架不是 RCWA，但本质上也要检查“单胞周期散射是否随网格、端口和相位周期边界稳定”。求解器名称不同，验证逻辑并没有变。

若某个单胞设计在 RCWA 中显示一条非常窄的反射特征，但将傅里叶阶数从 N_G 提高到

$N_G + 2$ 后，竞争模与目标模的排序发生翻转，那么最合理的结论是“当前单胞结论尚未收敛，不能进入下一阶段”，而非“结构具有很强模式选择”。

30.4 FDTD 验证：时间窗、PML 和激励源决定你看到的是模还是伪影

FDTD 的优势在于它直接处理有限阵列、宽带响应和近远场变换，因此在 PCSEL 工作中尤其适合做有限器件验证。但它也有典型的风险：若时间窗太短、PML 不稳或激励方式不合适，就会把数值伪影误读成模式特征。

第一，网格与时间步必须以关键输出为准，而不是以“默认稳定条件”为准。稳定只意味着算法不会发散，不意味着提取出的远场、 Q 或模排序已经可信。必须显式检查近场、远场主瓣和共振位置对网格是否收敛。

第二，PML 不只是一个吸收层参数，而是开放边界定义的一部分。若 PML 距离器件太近、厚度不足或参数与出射角谱不匹配，就会产生人工反射，进而污染衰减率和远场。很多表面上很高的 Q ，本质上只是边界没吸干净。

第三，激励源决定了你容易看到什么模。若源位置、偏振或频带过于偏向某一类模式，就可能让目标模看起来“天然更强”，但这并不等价于它在自发起振条件下也一定更有优势。因此，在有限阵列 FDTD 中，建议至少用两类激励方式交叉检查，例如脉冲偶极源与更接近目标模对称性的定制源。

第四，若要从时域衰减提取 Q ，就必须同时报告拟合时间窗和残差。因为 FDTD 输出常包含多个模的叠加，拟合区间选得不同，得到的衰减率可能明显变化。

从软件映射看，FDTD 主要对应 Lumerical FDTD 或 MEEP 一类时域工具。COMSOL 的主力并非 FDTD，而是频域 FEM。因此，如果你的工作流以 COMSOL 为主，就不应强行用“时域直觉”来解释所有频域结果，而应转向本附录后面的 FEM 验证逻辑。

30.5 FEM 验证：收敛的不只是网格，还有问题本身的弱形式实现

频域 FEM 在 PCSEL 建模中很有价值，特别是处理复杂几何、材料色散和多物理场接口时。但 FEM 的验证重点与 FDTD 不同。对 FEM 来说，真正要小心的除了网格大小，还有单元阶次、PML 形式和弱形式实现中的材料定义。

第一，必须区分 h -收敛和 p -收敛。也就是说，不仅要问网格细化是否稳定，还要问提高单元阶次后结果是否保持一致。对高对比度电磁边界和强曲率几何，区分这两类收敛尤其重要。

第二，开放系统中的本征频率问题常依赖 PML。若 PML 设计不当，本征模可能带入明显伪影或被错误地局域化。因此，PML 厚度、拉伸参数和位置都必须扫一轮，而不是一次设置后就默认可靠。

第三，要尽可能交叉比较本征问题与驱动问题。若某一共振在本征频率研究中存在，那么在对应频率附近的驱动频域响应中也应能看到相容特征。若二者长期对不上，往往说明边界、材料或模识别出了问题。

第四，弱形式中的材料色散和损耗写法必须和正文约定一致。特别是在包含复介电常数、温度依赖材料和各向异性时，只要实现细节稍有不一致，数值就可能“稳定但错误”。

在工程软件上，这一路径最典型的是 COMSOL 的 Wave Optics 频域或本征频率研究。其优势是与 Semiconductor 和 Heat Transfer 共几何、共网格策略更自然；代价则是验证工作不能偷懒，因为复杂耦合越多，弱形式和边界实现的错误就越不容易被肉眼发现。

对 FEM 本征频率问题，出现一个数值上平滑、形状上也像“模式”的场分布，并不自动说明它就是你要的物理模。只要 PML、端口、材料或对称边界设置不当，求解器同样可以给出自洽但无物理意义的伪模。

30.6 收敛性检查的目标：不是让数字不变，而是让结论不变

很多人做收敛分析时，只盯着一个绝对数值，例如共振波长变化不到多少就算通过。这个标准并非无用，但对 PCSEL 来说还不够，因为真正影响设计决策的通常是结论本身是否稳定，而非单个数值。例如：目标模和竞争模的排序是否稳定，远场主瓣是否仍朝同一方向，热回写后工作窗口是否仍然存在。

因此，更成熟的收敛检查应当至少包含两层阈值。一层是数值量收敛阈值，例如频率、 Q 、温升等变化小于某个比例；另一层是结论阈值，即排序、趋势和设计选择不发生翻转。若第一层过了而第二层没过，那么从工程角度看，结果仍然不够稳健，因为它还不能支撑设计决策。

30.7 从冷腔到器件级：需要怎样的阶段门槛

前文 第二十五章 把 workflow 分成了 L1 到 L4 四层。本附录则把这四层进一步转成“升级门槛”。表 30.1 总结了一个实用版本。

表 30.1: 从冷腔到器件级的阶段门槛。只有当前一层证据满足门槛，结果才应升级到下一层结论。

层级	主要输出	至少应满足的门槛
L1 结构候选层	候选模、对称性、被动产损耗趋势	单胞或简化模型收敛；功率平衡与边界一致；候选模可与理论图像对上
L2 光学阈值层	阈值增益、有限阵列远场、模排序	有限尺寸结果对阵列尺度和开放边界稳定；目标模相对竞争模排序不翻转
L3 电学映射层	载流子分布、J-V、阈值电流形成机制	电学边界与参数来源清楚；器件结论不再依赖均匀增益假设
L4 热-电-光工作层	工作窗口、热漂移、模式稳定性	温度场回写后结论仍稳健；工作窗口而非单点最优被明确给出

若要把“多物理场已收敛”写得更可复核，还应给出连续两轮闭环迭代的量化差分，例如

$$\frac{|\Delta I_{\text{th}}|}{I_{\text{th}}} < 2\%, \quad |\Delta T_{\text{max}}| < 1 \text{ K}, \quad |\Delta \lambda| < 0.02 \text{ nm},$$

并说明这些门槛对应的是冷腔层、电热层还是完整回写层。否则，“已收敛”仍然只是一个没有门槛的形容词。

这个表最重要的作用，是防止结论越级。只要尚未通过 L3 门槛，就不能输出器件级阈值电流；尚未通过 L4 门槛，就不能输出热稳定工作窗口。这样做看似保守，实际是在节约研究时间，因为它能避免把错误层级的结果带进后续更昂贵的仿真和实验中。

30.8 收敛标准不应只写“已收敛”，而应写成门槛

对真正可复用的 PCSEL 建模记录来说，“结果已收敛”这句话几乎没有信息量。更有用的写法，是先规定当前层最重要的 2–3 个观测量，再规定它们的允许变化范围。例如，单胞 RCWA 层可以把目标模频率、上下辐射比和阈值代理增益作为门槛量；有限阵列全波层可以把目标模频率、 Q 、主瓣角度和主瓣半高全宽作为门槛量；电热光层则可把阈值电流、峰值温升和工作频率漂移作为门槛量。

表 30.2: 多物理场链中的最小收敛门槛示例。

层级	建议观测量	收敛写法示例
RCWA 单胞层	候选模频率、反射谷位置、上下辐射比	增加傅里叶阶数后，三者相对变化均小于预设容差
FDTD/FEM 冷腔层	目标模频率、 Q 、远场主瓣角度、主瓣半高全宽	网格/PML/阵列尺寸继续加密时，结论不翻转且变化进入平台区
电学层	J–V 曲线拐点、载流子峰值位置、串联电阻估计	网格和边界再细化时，关键电学量变化不再显著
热学层	峰值温升、热点位置、热阻提取	热网格与热边界变化后，热点位置与热阻进入稳定区
电热光闭环层	阈值电流、工作频率漂移、目标模相对竞争模裕量	回写再细化后，目标模身份与窗口结论不翻转

30.9 网格收敛最好固定成五步流程

若每次都凭直觉决定“再加密一点”，很容易把收敛检查做成不可复现的个人习惯。更稳妥的最小流程是五步：第一，先选一个当前最关心的可观测量，例如目标模频率、 Q 或主瓣角度；第二，只改一个离散参数，如网格尺寸、单元阶次或时间步长；第三，记录两到三个连续加密层级下的变化；第四，确认变化趋势进入平台区，而不是仅偶然变小一次；第五，再切到下一个离散参数。这个顺序的意义，在于避免多个数值旋钮同时变化后无法归因。

易错点

多物理场链里最常见的伪收敛，是“光学网格收敛了，但电热回写后结果又翻转”。这是层级升级后可观测量本身变了，而非收敛失败。因此，任何声称“器件级结果已收敛”的说法，都必须写明它是在冷腔层、电学层、热学层还是闭环层收敛。

30.10 排错决策树：结果异常时按什么顺序查

当结果异常时，不宜同时修改多个设置并混合归因。更有效的做法是固定一棵决策树。图30.1 给出一条适合 PCSEL 建模的最小顺序：先查版本和几何，再查边界与单位，再查收敛，最后才讨论是否可能存在真正的新物理。

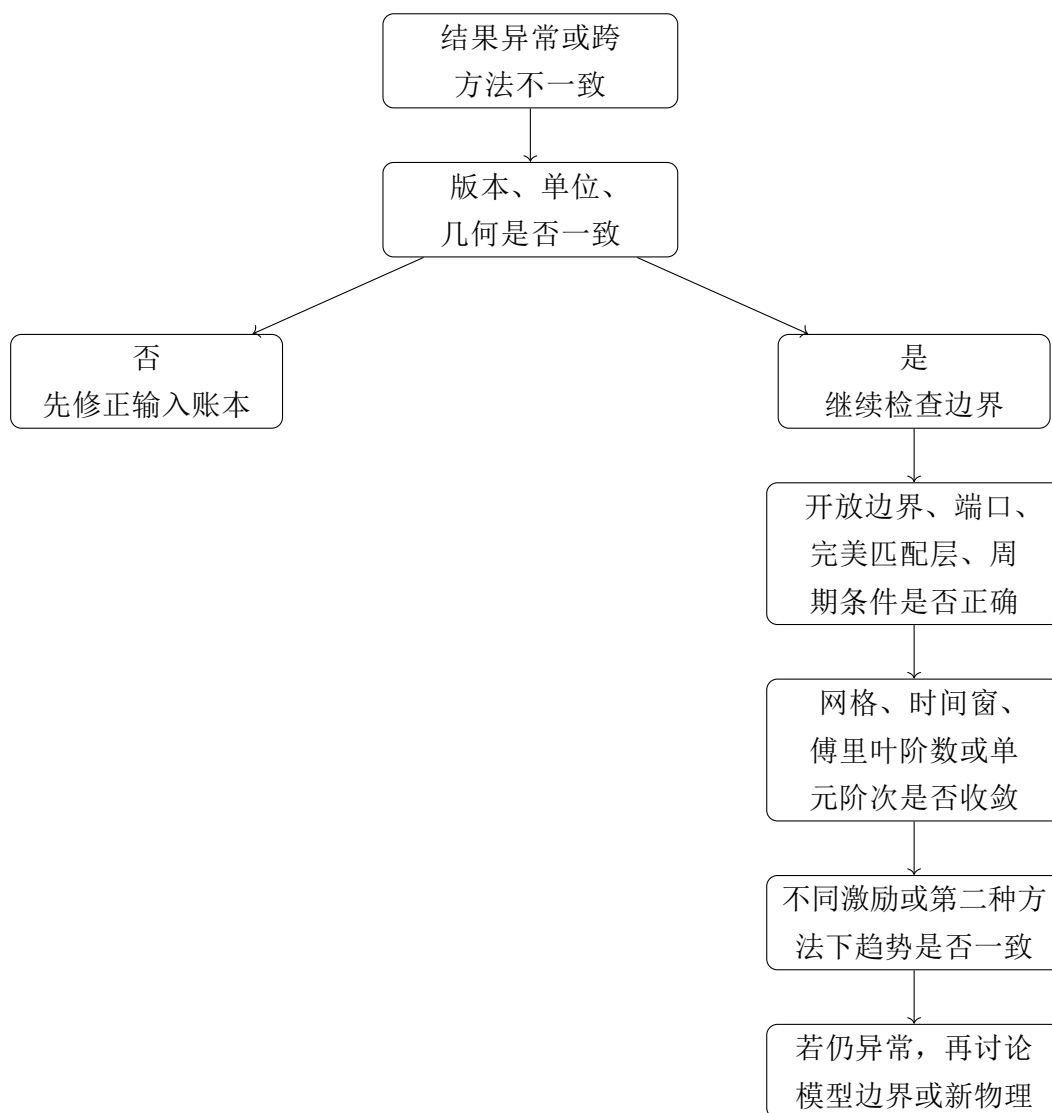


图 30.1: PCSEL 建模的最小排错决策树。

这棵树的意义不在于它是否覆盖所有细节，而在于它强制研究者避免一个高频错误：还没确认输入和边界是否一致，就急着把异常结果解释成模式分裂、非线性或热反馈新现象。对开放系统而言，这种过早解释尤其危险，因为数值边界伪影本来就很容易伪装成模态物理。

只保存“最直观的一次结果图”，而不保存失败设置、收敛扫描和版本差异记录，会让团队在几周后完全失去判断力：既不知道当前结果是不是偶然命中，也不知道过去排除过哪些伪影路径。对 PCSEL 这类多阶段工作流，这种信息丢失的代价极高。

30.11 怎样把验证结果写成可发布报告

当所有验证都完成后，最好的报告方式并非把几十张收敛图全部塞进正文，而是把它们整理成三层证据：

第一层是“当前结论依赖什么模型和近似”。这让读者知道结论的物理边界。

第二层是“当前结论通过了哪些数值门槛”。这让读者知道结果不只是理想图，而是经得起反证的结果。

第三层是“当前结论仍受哪些不确定因素限制”。这让读者知道结果是否已达到可发布级别，还是仍然停留在探索性级别。

对软件路径而言，也应采取同样写法。例如，不必写成“在某软件里点了哪些选项”，而应写成“在周期单胞开放散射模型中已验证傅里叶/网格收敛，在有限阵列开放全波模型中已验证 PML 与阵列尺寸收敛，在器件模型中已验证电热回写后结论不翻转”。这样写既保持与求解器实现解耦，又保留了足够的技术可追溯性。

30.12 数值不确定度量化与报告规范

基于 第十六章 中的误差分析理论，本节给出可直接用于项目实践的数值不确定度量化流程和报告模板。

30.12.1 各方法的主要误差源

多方法互证的另一个关键维度是理解并量化每种方法的数值误差来源。只有清楚地知道每个数字背后的不确定度，才能做出可靠的工程判断。本节系统梳理各类方法的主要误差源、收敛行为和交叉验证策略。

先统一本章收敛门槛口径（探索级/本书算例/发布级） 为避免把“调参初筛阈值”误读成“发布级硬标准”，本章后续阈值统一按三层门槛解释：探索级用于快速筛查方向；[本书算例]用于本教材 benchmark 对齐；发布级用于论文或项目验收。若项目硬件、几何或误差预算变化，应重新定义门槛，而不是机械照搬。

表 30.3: 本章收敛门槛的三层口径（方法学文献 + 本书 benchmark 约定）。

指标	探索级（快速筛查）	本书算例（章节默认）	发布级（建议）
PWEM 频率收敛	相邻 N_G 频率漂移 < 1%	相邻 N_G 频率漂移 < 0.5%	相邻 N_G 频率漂移 < 0.2% 并给收敛曲线
RCWA 共振峰漂移	阶数扫描后峰位变化 < 2 nm	峰位变化 < 0.5 nm	峰位变化 < 0.2 nm 且功率守恒误差 < 10^{-3}
FDTD/FEM 共振频率一致性	同数量级一致	相对偏差 < 1%	相对偏差 < 0.5% 并给网格与时间窗敏感性
模式排序一致性	目标模不翻转	目标模不翻转 + 竞争模式次序稳定	目标模不翻转并给最坏样本与误差条

30.12.1.1 PWEM 的误差来源与控制

PWEM 作为本征模求解器，其主要误差来自三个方面：

平面波截断误差 PWEM 将介电常数分布展开为傅里叶级数，截断阶数 N_G 直接影响精度。典型的收敛行为是：随着 N_G 增加，本征频率按幂律收敛。对高对比度结构（如空气孔/半导体界面），需要更高的 N_G 才能达到与低对比度结构相同的精度。

数值误差控制

PWEM 收敛性检查清单

- **频率收敛：** 目标模式的归一化频率 $\omega a/(2\pi c)$ 随 N_G 的变化应小于 10^{-4}
- **带隙位置收敛：** 带隙边缘的频率漂移应小于带隙宽度的 1%
- **场分布收敛：** 连续两次 N_G 增加的场均方根差异应小于 1%
- **对称性保持：** 简并模的频率分裂应远小于它们与邻近模式的间距

典型经验值：对空气桥型 PC， $N_G \geq 11 \times 11$ ；对衬底支撑型 PC， $N_G \geq 15 \times 15$ 。

k 空间采样误差 对色散曲线的绘制， $\mathbf{k}$ 空间的采样密度会影响带结构的平滑度和极值点的准确性。建议在布里渊区高对称点附近加密采样，以确保带边位置的精度优于 0.5%。

有效折射率近似误差 二维 PWEM 通常用有效折射率 n_{eff} 代替真实的三维波导结构。这会引入系统性误差，特别是对垂直损耗和模式体积的计算。定量估计时，应将二维 PWEM 结果与三维 RCWA 或 FEM 进行基准比对，典型偏差范围在 5% – 15% 之间。

30.12.1.2 CWT 的模型降阶误差

CWT 通过将无穷维 Maxwell 方程投影到少数主导波上来实现模型降阶。其误差主要来自 truncation 和耦合系数近似。

主导波选择的完备性 CWT 的精度取决于所选基底是否能充分描述真实物理。一个实用的检验方法是逐步增加波的数目，观察关键输出（如阈值增益、辐射效率）的相对变化：

$$\varepsilon_N = \left| \frac{X_{N+1} - X_N}{X_N} \right|, \quad (30.1)$$

其中 X_N 是用 N 个波计算的目标量。当 $\varepsilon_N < 1\%$ 时可认为收敛。

耦合系数的计算精度 CWT 中的耦合系数 κ 通常通过对单胞积分获得：

$$\kappa = \frac{\omega^2}{2V} \int_V \Delta\varepsilon(\mathbf{r}) \mathbf{E}_i^*(\mathbf{r}) \cdot \mathbf{E}_j(\mathbf{r}) d^3r. \quad (30.2)$$

数值积分时的网格密度和几何逼近会直接影响 κ 的精度。建议做法是：先用高精度全波方法（RCWA/FEM）计算单胞场，再用这些场校准 CWT 的耦合系数。

交叉验证协议

CWT 与 RCWA 的交叉验证协议

1. 用相同单胞几何分别运行 CWT 和 RCWA
2. 比较两者的共振频率偏差： $|\lambda_{\text{CWT}} - \lambda_{\text{RCWA}}|/\lambda_{\text{RCWA}} < 2\%$
3. 比较辐射效率的趋势一致性（不必逐点相同）
4. 比较参数扫描的灵敏度符号（如对孔径变化的响应方向）

若以上任一条件不满足，应返回检查 CWT 的主导波选择或 RCWA 的傅里叶阶数。

30.12.1.3 RCWA 的傅里叶收敛与 Gibbs 现象

RCWA 的核心是将电磁场和介电函数展开为 Floquet-Bloch 级数。其主要误差源包括：

傅里叶阶数不足导致的 Gibbs 振荡 对具有高对比度界面的结构，介电函数的傅里叶级数在不连续处会出现 Gibbs 振荡。这会导致：

- 共振峰位置的系统性偏移（通常向长波方向）
- 虚假的旁瓣或共振分裂
- 功率守恒 violation（反射 + 透射 $\neq 1$ ）

数值误差控制

RCWA 收敛性检查清单

- **傅里叶阶数扫描**：从 $N_G = 10$ 开始，每次增加 5，直到共振波长变化 $< 0.5 \text{ nm}$
- **功率守恒**： $|R + T + A - 1| < 10^{-3}$ （无源结构）
- **reciprocity 检查**：正向和反向入射应给出相同的透射谱
- **层厚收敛**：对有刻蚀深度的结构，纵向切片数应使得每片厚度 $< \lambda/(20n)$

对深孔结构（纵横比 $> 5:1$ ），可能需要 $N_G > 30$ 才能收敛。

坐标变换技术的必要性 对圆柱形空气孔等曲面边界，标准的阶梯近似会引入显著的几何误差。此时应采用因子化规则（factorization rules）或曲线坐标系 RCWA 来提高精度。一个简单的诊断是：若增加 N_G 后共振波长仍在单调漂移（而非震荡收敛），则很可能需要改进几何表述。

30.12.1.4 FDTD 的时间域误差累积

FDTD 通过在时间和空间上离散 Yee 网格来演化 Maxwell 方程。其误差特性与其他方法显著不同。

数值色散误差 Yee 网格中的波传播速度依赖于波长与网格尺寸之比。这会导致：

$$v_{\text{num}} = c \left[1 - \frac{1}{24}(k\Delta x)^2 + O((k\Delta x)^4) \right], \quad (30.3)$$

其中 $k = 2\pi/\lambda$ 。为将相速度误差控制在 1% 以内，需要 $\Delta x < \lambda/(10n)$ 。对 PCSEL 这种多波长干涉的结构，建议使用更保守的标准： $\Delta x < \lambda/(15n)$ 。

时间窗长度与 Q 值提取误差 从高 Q 谐振器的衰减曲线提取 Q 值时，时间窗长度 T_{win} 必须满足：

$$T_{\text{win}} > \frac{10Q}{\omega_0}, \quad (30.4)$$

否则拟合得到的衰减速率会有显著偏差。一个常见的症状是：延长仿真时间后，提取的 Q 值持续增大而不饱和。

PML 反射引起的伪影 完美匹配层（PML）的理论反射率为零，但数值实现中会有微小反射。对超高 Q 模式（ $Q > 10^6$ ），即使是 10^{-6} 量级的 PML 反射也会污染结果。缓解策略包括：

- 增加 PML 层数（典型值：12-16 层）
- 使用渐变电导率剖面（polynomial grading）
- 将 PML 放置在远离强场区域的位置（至少 $\lambda/2$ ）
- 若问题本身是无限周期单胞，可在横向采用 periodic boundary + Bloch 条件；若是有限器件，应保留开放边界（PML），并通过增大计算域、提高 PML 厚度/阶次与参数扫描抑制伪反射

交叉验证协议

FDTD 与 FEM 的交叉验证协议

1. 在同一有限尺寸器件上分别运行 FDTD（时域）和 FEM（频域）
2. 比较共振频率：偏差应 $< 1\%$
3. 比较 Q 值：在对数尺度上一致即可（如两者都在 10^4 量级）
4. 比较近场图案的重叠积分： $|\langle E_{\text{FDTD}} | E_{\text{FEM}} \rangle|^2 > 0.9$
5. 比较远场主瓣方向和角宽：峰值角度偏差 $< 2^\circ$

注意：由于 FDTD 是宽带激励而 FEM 是单频求解，前者可能激发更多杂散模，因此近场重叠略低是可接受的。

30.12.1.5 FEM 的网格自适应与弱形式误差

FEM 通过变分原理将微分方程转化为线性代数问题。其误差估计最为成熟。

h- refinement 与 p-refinement FEM 提供两种收敛途径：

- **h-精化**：减小网格尺寸 Δh ，固定单元阶次 p
- **p-精化**：提高单元多项式阶次 p ，固定网格

对光滑解，p-精化的指数收敛速度远超 h-精化。但对具有奇点（如尖锐金属边角）的问题，h-精化更为稳健。

残差型后验误差估计 现代 FEM 软件通常提供基于残差的误差指示器：

$$\eta_K = \|R(u_h)\|_{L^2(K)} + \|J(u_h)\|_{L^2(\partial K)}, \quad (30.5)$$

其中 $R(u_h)$ 是单元内部残差， $J(u_h)$ 是通量跳跃。adaptive mesh refinement (AMR) 会根据 η_K 自动加密高误差区域。

数值误差控制

FEM 收敛性检查清单

- **自由度扫描**：绘制目标量 vs 自由度数，确认已进入渐近平台区
- **单元阶次提升**：从二次单元升到三次单元，目标量变化应 $< 1\%$
- **PML 敏感性**：移动 PML 边界或改变 PML 参数，本征频率漂移应 $< 0.1\%$
- **能量范数收敛**： $\|u - u_h\|_E / \|u\|_E$ 应随自由度增加而单调下降

对本征频率问题，还需额外检查：虚部（损耗）的收敛速度通常慢于实部（频率）。

30.12.2 最小必要收敛扫描组合

对 PCSEL 多方法工作流，不建议对所有参数都做全面扫描（那样成本过高），而是聚焦于对最终结论最敏感的三个维度：

表 30.4: 各方法的最小必要收敛扫描组合。

方法	必扫参数	通过标准
PWEM	平面波截断数 N_G	目标模频率变化 $< 0.5\%$ ，简并模分裂 $<$ 邻近模间距的 10%
RCWA	傅里叶阶数 + 纵向切片数	共振波长漂移 $< 0.5 \text{ nm}$ ，功率不平衡 $< 10^{-3}$
FDTD	网格密度 + 时间窗长度	Q 值变化 $< 5\%$ ，远场主瓣角度漂移 $< 2^\circ$
FEM	网格自由度 + 单元阶次	本征频率实部变化 $< 0.1\%$ ，虚部变化 $< 10\%$
CWT	主导波数量	阈值增益相对变化 $< 2\%$

实际操作时，建议采用“两步粗扫 + 一步精扫”策略：先在较大步长下确定大致平台区，然后在平台区内做一次精细扫描，最后用 Richardson 外推或直接取平台区的平均值作为最佳估计。这样可以避免陷入局部平台或误判收敛趋势。

30.12.3 不确定度的合成与传递

对多步骤 workflow，前一级仿真输出的不确定度会成为下一级输入的误差源。设第 i 级输出的相对不确定度为 δ_i ，则最终结果的总相对不确定度可用平方和根 (RSS) 估算：

$$\delta_{\text{total}} = \sqrt{\sum_{i=1}^N w_i^2 \delta_i^2}, \quad (30.6)$$

其中 w_i 是第 i 级对最终结果的灵敏度权重。对缺乏严格灵敏度分析的情况，保守做法是取 $w_i = 1$ 。

数值误差控制

典型 PCSEL 工作流的不确定度预算示例 假设一个完整的冷腔 → 阈值电流预测链条包含以下步骤：

- PWEM 带结构： $\delta_1 \approx 0.5\%$ （频率定位）
- CWT 耦合系数： $\delta_2 \approx 3\%$ （辐射损耗提取）
- 速率方程阈值： $\delta_3 \approx 5\%$ （载流子输运模型）
- 热回写迭代： $\delta_4 \approx 2\%$ （温度依赖折射率）

按 RSS 合成，总不确定度约为 $\delta_{\text{total}} \approx \sqrt{0.5^2 + 3^2 + 5^2 + 2^2}\% \approx 6.2\%$ 。这里应明确该结果属于 [本书算例]，不应直接外推到任意器件。它提醒我们：即使每一步都已“收敛”，最终阈值电流预测仍可能存在数个个百分点级别的不确定度。因此，声称“仿真阈值为 150 mA”不如写成“仿真阈值为 150 ± 10 mA”更符合报告规范。

可能出错处

一个常见但危险的误区是：在前几步用较粗糙的网格快速筛选，期望在最后一步用精细网格“纠正”所有误差。这种做法忽略了误差的非线性放大效应——早期阶段的定性误判（如模式排序错误）无法通过后期的高精度计算挽回。正确的策略是：在每一层都确保数值误差不至于改变定性结论。

30.12.4 交叉验证的可接受偏差范围

当使用两种独立方法进行交叉验证时（如 FDTD vs FEM），需要事先定义什么是“一致”。表 30.5 给出的阈值是 [本书建议门槛]，用于教学和工程筛查，不代表社区统一标准。

这些门槛的来源可追溯到 RCWA/FDTD/FEM 的经典收敛文献与有限尺寸敏感性讨论 [40, 45, 52]；但具体项目应按器件波段、几何复杂度和实验误差预算重定义后再用于发布级结

表 30.5: 跨方法交叉验证的可接受偏差范围（[本书建议门槛]，可按项目重定义）。

待比较量	宽松容差 (探索级)	严格容差 (发布级)
共振波长/频率	< 5 nm 或 < 3%	< 1 nm 或 < 0.5%
Q 值	同数量级即可（[探索级]）	相对偏差 < 20%（[发布级]）
远场主瓣角度	< 5°	< 2°
近场重叠积分	> 0.8	> 0.9
阈值增益	< 10%	< 3%
模式排序	完全一致	完全一致（硬性要求）

论。

可能出错处

一个常见误区是用“宽松容差”级别的吻合去支撑“发布级”强结论。例如，若两种方法的 Q 值只在同一数量级（如 10^4 vs 2×10^4 ），就只能归类为 [工程经验]/探索级一致，而不能宣称 [文献值]/发布级定量可靠。正确做法是先声明采用的容差等级，再限定结论强度。

30.12.5 可重复性报告的最小信息集

为确保他人能够复现你的数值结果，建议在论文或技术报告的附录中包含以下信息：

报告规范

PCSEL 数值仿真的最小可重复信息清单

- 1. 几何参数：晶格常数、孔径/占空比、刻蚀深度、层厚（附公差或加工变异范围更佳）
- 2. 材料模型：折射率数据来源（实测/文献/拟合）、色散公式类型（Cauchy/Sellmeier/Drude-Lorentz）、是否包含虚部
- 3. 离散设置：网格类型（结构化/非结构化）、关键区域网格尺寸、单元阶次（FEM）、时间步长（FDTD）
- 4. 边界条件：PBC/PML/PEC 的位置和参数、PML 层数和吸收剖面
- 5. 收敛证据：至少提供一组关键量的收敛曲线（可在 Supplemental Material 中）
- 6. 软件信息：求解器名称、版本号、关键插件或自定义脚本
- 7. 硬件资源：CPU/GPU 型号、内存、单次仿真耗时（有助于他人评估复现成本）

示例表述（可放入 Methods 部分）：“RCWA 仿真使用 S4 求解器 (v1.8)，傅里叶阶数 $N_G = 21$ ，纵向切片数 12 层。网格收敛测试见 Supplementary Fig. SX，共振波长标准偏差为 0.6 nm。”

若期刊不允许长篇方法细节，可将完整收敛数据和脚本上传至公开代码仓库（如 GitHub、Zenodo 或 figshare），并在正文中提供 DOI 链接。这已成为计算光子学领域的最佳实践，也是许多高水平期刊的强制要求。

回看本章

仿真验证是物理论证的一部分，而非报告末尾的形式性清单。RCWA 需要证明单胞开放散射的截断和功率平衡可靠，FDTD 需要证明时间窗、PML 和激励源没有把伪影当成模，FEM 需要证明网格、单元阶次和弱形式实现不会制造伪模。更高一层地说，所有这些方法验证最终都要服务于工作流门槛：只有当 L1 到 L4 的证据依次满足基本约束，结论才应升级。可靠的 PCSEL 建模，不只是“求解器完成运行”，而是“结论有门槛、异常有决策树、结果有可发布级别”。

练习题

1. [基础] 为一个你当前正在研究的 PCSEL 结构，分别写出 RCWA、FDTD 或 FEM 中最关键的三项验证门槛，并说明每一项各自防止什么误判。
2. [基础] 结合 表30.1，说明为什么一个已经通过 L2 的结构仍然不能直接宣称器件级阈值电流优越。
3. [进阶] 试着把 图30.1 改写成适合你自己项目的排错流程，并加入你最常遇到的一类异常症状。
4. [综合]（思路提示：先定层级，再列关键式）设计一份一页纸验证报告模板，使其同时适用于以 Lumerical 为主和以 COMSOL 为主的工作流。

延伸阅读

- 建议与 第二十五章、第二十七章 对照阅读，把本附录直接转化成项目级发布门槛和调试清单。阅读时应同时核对本章变量在后续模型中的定义和适用范围。
- 建议与 第二十六章、第二十七章 对照阅读，把本附录中的验证门槛、例题中的执行步骤和失败模式中的反向诊断锁成一条完整证据链。阅读时应同时核对本章变量在后续模型中的定义和适用范围。

第三十一章 术语表与符号总表

读完本章，你应能

- 把全书中最关键的术语、符号和层级区分压缩成可反复查阅的一份统一对照表。
- 避免在阅读 第9章、第二十三章和二十五至二十七 等跨模块章节时发生语义漂移。
- 建立一套适合研究记录、仿真汇报和论文写作的统一命名习惯。

读前准备 本章是查阅型索引，没有先修门槛：从任何一章跳转过来查词都可以直接使用，遇到层级可疑的表述时随时翻开即可。

阅读主线 本章所处模型层级：跨层索引（L1→L4）

本章先给出核心术语表，再给出符号总表，随后整理最常见的混淆对照，最后说明本章在工作流、例题与失败模式诊断中的使用位置。为了帮助读者快速建立对全书符号体系的整体认知，图31.1以树状图的形式展示了所有关键符号的分类架构。该图将数百个分散在各章的物理量系统地归纳为三大族群：基础物理（蓝色）、材料几何（绿色）和数值计算（橙色），并在每个类别下进一步细分为若干子类，标注出最具代表性的符号实例。建议初次接触本书时先将此图浏览一遍，后续遇到陌生符号时可迅速定位其所属范畴以减少翻阅时间。

直觉框

术语表是为了让证据链不失真，而非为了把词翻译得“直观”。PCSEL 研究跨越开放电磁系统、半导体器件和多物理场建模，同一个词若在不同章节里含义漂移，结论就会在不知不觉中越级。

31.1 如何使用本章

本章最适合在三种情境下反复打开。第一，读到一个看似熟悉、但层级可能不同的词时，例如“threshold”“mode”“stability”。第二，准备把某个中间结果写进报告或论文时，用它检查自己是否把代理模型结论写成了器件结论。第三，团队讨论中出现“大家都在用同一个词，但好像在说不同东西”时，用本章统一语义。

如果需要一个最直接的用法，可以把本章与三张表配套使用：表25.1、表26.3和表26.4。前两张表告诉你应该走哪条证据链，本章则保证链条上的每一个词和符号都指向同一层物理含义。

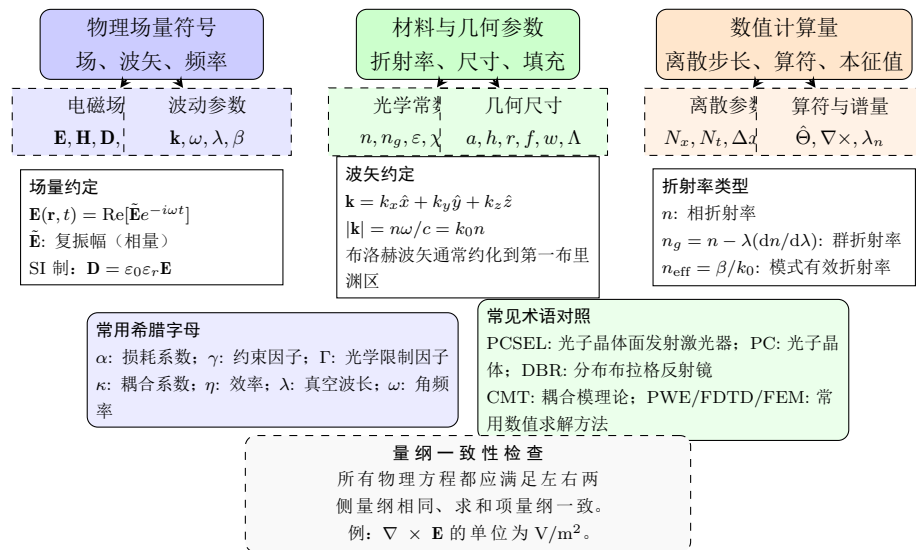


图 31.1: 本书主要符号与术语的分类索引。图中按场量、材料几何和数值变量三类汇总常见记号, 并给出场量约定、波矢约定与量纲一致性提示。

31.2 核心术语表

术语	教材化定义
布洛赫模 (Bloch mode)	周期结构中满足布洛赫相位条件的本征场。它是无限周期分析的基本对象, 但不自动等于有限器件中的工作模。
Γ 点模	布里渊区中心附近的模式家族。它与垂直辐射、偏振选择和面发射最直接相关, 因此是 PCSEL 的核心候选模区域。
带边模 (band-edge mode)	色散曲线局部斜率很小、群速度趋近于零附近的模式。它是被动色散概念, 不自动等于最终激射模。
被动共振 (passive resonance)	在无增益、无注入、无热回写条件下, 由几何和材料分布决定的开放系统共振。
有源阈值 (active threshold)	在给定增益、损耗、注入和工作点条件下, 净增益刚好补偿总损耗的起振条件。
阈值增益 (threshold gain)	线性光学层级下, 为补偿模式总损耗所需的最小模增益。它是光学量, 不等于器件阈值电流。
阈值电流 (threshold current)	器件层级下, 经过电注入、载流子输运、热回写和损耗共同决定的起振电流。
导模共振 (guided resonance)	本来受波导约束、但因周期扰动耦合到辐射连续谱而可向外泄露的开放模式。
连续谱束缚态 (BIC)	频率落在辐射连续谱中, 却因对称性或干涉机制而不向外泄露的理想束缚态。

术语	教材化定义
对称性保护 (symmetry-protected BIC)	由于模式对称性与可辐射通道不匹配, 导致耦合严格为零的 BIC。它常出现在 Γ 点附近。
偶然 BIC (accidental BIC)	是由不同辐射通道之间的振幅和相位抵消形成的 BIC, 而非由严格对称性。
准 BIC (quasi-BIC)	当理想 BIC 因有限尺寸、扰动、无序或倾斜激发而获得有限泄露后形成的高 Q 共振。它在真实 PCSEL 中更常见。
辐射通道 (radiation channel)	模式可以向外界泄露能量的远场通道, 包括不同偏振、角度和上下半空间的辐射分量。
偏振选择	结构对不同偏振模式的辐射、损耗或增益施加不对称约束, 从而导致某一偏振更容易成为稳定工作模。
远场工程	通过改变晶格、扰动、有限阵列边界或纵向层栈, 主动塑造主瓣、旁瓣和偏振纯度的设计行为。
有效折射率近似	用等效二维折射率替代真实三维层状结构的代理模型。它适合快速筛选, 但不能替代最终三维矢量验证。
光学约束因子 (optical confinement factor)	模场与有源区或损耗区的重叠权重。它把材料层的性质映射到模层级。
材料增益 (material gain)	材料本身在给定载流子密度和频率下提供的受激辐射增益。
模增益 (modal gain)	材料增益经过模场重叠加权后作用于某一模式的有效增益。
增益压缩 (gain compression)	随光子密度增大而导致微分增益下降的非线性效应。它直接进入动态响应与调制带宽分析。
微分增益 (differential gain)	材料或模式增益对载流子密度的导数, 是决定调制响应和弛豫振荡频率的重要量。
Henry 因子 (linewidth enhancement factor)	载流子引起的折射率变化与增益变化之间的耦合因子, 常记为 α_H 。它把振幅噪声耦合到相位噪声。
弛豫振荡 (relaxation oscillation)	光子数与载流子数在平衡点附近的耦合振荡, 是小信号调制响应中的核心特征频率。
小信号调制响应	对工作点附近微小驱动扰动的线性频率响应, 用于描述调制速度和动态稳定性。
3 dB 调制带宽	小信号调制响应幅度下降到低频值的 $1/\sqrt{2}$ 时对应的频率, 常作为速度指标。
鲁棒性窗口	在规定工艺误差、热漂移和注入偏差范围内, 器件仍保持目标性能的参数窗口。
容差分析	系统评估几何、材料、注入和热边界误差如何传播到模式排序、远场和器件性能的分析流程。
验证闭环	从测量到参数提取、分层校准、模型预测再到回验的闭环过程, 用于检验模型是否具备外推能力。

术语	教材化定义
回验 (back-validation)	使用未参与校准的观测量去检验模型预测能力的步骤。若只会拟合不会回验, 模型就还不可靠。

31.3 符号总表

符号	所属层级	含义
$\omega, \tilde{\omega}$	电磁	实频角频率与复频率; 复频率虚部表征损耗或增长。
Q	电磁	品质因子, 常由复频率或功率损耗定义。高 Q 不自动等于最低阈值电流。
α_{rad}	电磁	辐射损耗系数, 反映模式向自由空间泄露的强弱。
α_{int}	电磁/器件	内部损耗系数, 包括自由载流子吸收、掺杂损耗、金属邻近损耗等。
Γ_{opt}	光学/器件	模场在有源区中的光学约束因子。
Γ_{QW}	器件	模场与量子阱区域的重叠因子。
Γ_{loss}	器件	模场与主要损耗区域的重叠因子。
g_{mat}	器件	材料增益。
g_{mod}	光学/器件	模增益, 常近似写为 $\Gamma_{\text{opt}}g_{\text{mat}}$ 或更精细的空间积分形式。
g_{th}	光学	阈值增益, 即补偿模式总损耗所需的最小模增益。
$N(\mathbf{r})$	器件	载流子密度空间分布。
$J(\mathbf{r})$	器件	电流密度空间分布。
$T(\mathbf{r})$	热/器件	温度空间分布。
$\Delta n, \Delta g$	耦合模型	热或载流子回写到折射率和增益的变化量。
τ_p	动态学	光子寿命。
τ_n	动态学	有效载流子寿命。
S 或 P	动态学	腔内光子数或与之成比例的模式强度变量。
β_{sp}	动态学	自发辐射耦合因子。
α_H	动态学/噪声	Henry 线宽增强因子。
ω_R 或 f_R	动态学	弛豫振荡角频率或频率。
$f_{3\text{dB}}$	动态学	小信号调制响应的 3 dB 带宽。
ε_g	动态学	增益压缩系数。
μ_n, μ_p	电学	电子与空穴迁移率。
D_n, D_p	电学	电子与空穴扩散系数。

符号	所属层级	含义
R_{SRH}	电学/器件	Shockley–Read–Hall 复合率。
R_{Auger}	电学/器件	Auger 复合率。
Q_{Joule}	热/器件	Joule 热源项。
k_{th}	热学	热导率。
R_{th}	热学/实验接口	热阻，用于连接耗散功率与温升。
κ_1, κ_2	半解析模型	CWT 或玩具模型中的耦合系数。
γ_i, γ_r	半解析模型	玩具模型中的内部损耗与辐射损耗。

31.4 多义符号与上下文辨识

符号	常见语境	使用提醒
Q	冷腔品质因子、辐射 Q 、 加载 Q	若不说明损耗来源，单写 Q 很容易把辐射损耗、内部损耗和外部耦合混成同一量。
Q_c	导带偏移系数	这是材料能带分配系数，不是品质因子 Q 的下标变体；出现在外延和异质结章节时应优先按材料参数理解。
Γ	布里渊区中心点、光学 约束因子、量子阱重叠 因子、某些降阶模型中的 矩阵记号	章节切换时必须靠上下文或下标区分； Γ 点与 Γ_{opt} 绝不是同一类对象，写成矩阵时也不应和约束因子混读。
β	slab 传播常数、自发辐射 耦合因子 β_{sp}	在器件和动态学章节尤其容易混淆，应尽量带下标或直接写全名。
g	材料增益、模增益、阈值 增益、微分增益	若没有下标，必须立即说明它属于材料层、模层还是导数层。
α	损耗系数、Henry 因子 α_H	前者是损耗量，后者是线宽增强因子；两者物理意义完全不同。
N	载流子密度、样本数、阵列 周期数	数值章节和器件章节中应避免只写孤立的 N 而不加说明。

使用建议。遇到多义符号时，先看它第一次出现在哪一章；再看它是否带下标；最后判断它属于电磁、器件、动态学还是 workflows 语境。对全书最常见的几个多义符号，最稳妥的写法是不要让 Q 、 Γ 、 β 、 g 脱离上下文孤立出现。

全书中凡出现 “threshold” “stability” “mode selection” 这类词，默认都必须带层级。若没有说明它属于被动电磁层级、简化器件代理层级，还是器件级工作点层级，那么表述就是不完整的。

31.5 高频混淆对照

易混淆对	教材化区分句
带边模 vs 激光模	前者是被动色散对象，后者是增益、损耗、注入和热回写共同决定的工作结果。
被动共振 vs 有源阈值	前者描述结构本身的开放电磁行为，后者描述增益补偿后的起振条件。
阈值增益 vs 阈值电流	前者是光学量，后者是器件量；前者可以来自冷腔和增益代理，后者必须经过电热回写。
BIC vs 高 Q 共振	BIC 是理想耦合严格为零的束缚态，高 Q 共振只说明泄露很小，不说明耦合严格为零。
准 BIC vs 低阈值模	准 BIC 可能提供较低辐射损耗，但若注入、热和偏振选择不利，仍不一定成为最低阈值器件模。
亮模 vs 先激光模	亮模说明向外辐射强，不说明净阈值最低。亮模甚至可能因为辐射损耗大而更晚起振。
无限周期 vs 有限器件	前者给候选模式池和局部色散信息，后者决定边缘泄露、远场、尺寸收敛和真实工作模。
均匀增益 vs 电注入增益	前者是代理模型，后者由 $J(\mathbf{r})$ 、 $N(\mathbf{r})$ 和温度场共同决定。
冷腔频率 vs 工作频率	工作频率还会受载流子、温升和非线性效应拖动。
校准量 vs 回验量	校准量用于拟合模型参数，回验量用于检验模型外推能力；两者不能混用。
名义值最优 vs 鲁棒性最优	前者只在单点参数上最好，后者要求在误差和漂移存在时仍保持性能窗口。
仿真一致 vs 物理正确	两个求解器给出相近数字只是必要条件，不是充分条件；仍需检查边界、层级和守恒关系。

合规写法示例：“该结构在被动开放电磁模型下显示出接近准 BIC 的高 Q 候选模，因此在光学层级上具有较低辐射损耗潜力；其器件级阈值电流与线宽仍需结合 第二十三章、第二十九章 中的动态学、注入与热回写模型评估。”

易错点

把“存在一个高 Q 的 Γ 点候选模”直接改写成“器件一定会以单模低阈值工作”，是全书最需要避免的术语越级。类似地，把“调制响应更快”写成“线宽一定更窄”也是把动态学结论和噪声结论混在了一起。

31.6 如何把本章用回 workflow、例题与失败模式

第二十五章 负责说明建模链应该怎样向前推进；第二十六章 完整展开这条链；第二十七章 则把同一条链倒过来，检查哪里最容易断。本章的作用，是确保这三条阅读路径使用的是同一套词汇和符号。

更具体地说：

- 当你在第九章 里看到 BIC、quasi-BIC、radiation channel 时，应回到本章确认它们与高 Q 、亮模、远场工程的边界。
- 当你在第二十三章 里看到 α_H 、 ω_R 、 f_{3dB} 时，应回到本章确认这些量属于动态学和噪声层级，而不是静态阈值层级。
- 当你在第二十六章 里整理中间结果时，可直接用本章检查每一个输出量属于“候选模证据”“有限尺寸证据”还是“器件代理证据”。
- 当你在第二十七章 中排错时，可用本章快速判断自己究竟是符号混用、术语越级，还是模型层级本身已经错位。

回看本章

本章把全书中最容易漂移的术语和符号压缩成一份统一工具表。有效的用法，是在写公式、做 workflow、整理例题中间结果和诊断失败模式时，持续用它检查自己是否发生了层级混用、概念越级或符号漂移，而非偶尔查一个词。

练习题

1. [基础] 从第九章、第二十三章 中各找出两个你最容易混淆的术语，分别写出一句不会越级的定义。
2. [基础] 用本章的混淆对照表，重写一句你在研究笔记里曾经写过的“高 Q 因此低阈值”式表述。
3. [进阶] 以第二十六章 的一张中间结果表为对象，给出每个中间量的层级标签：被动光学、有限尺寸、器件代理或动态学。
4. [综合]（思路提示：先定层级，再列关键式）说明为什么在实验校准中，回验量不能再反过来参与同一轮参数拟合。

延伸阅读

- 若你在整理理论与辐射主线，建议与第九章、第二十八章 对照阅读。阅读时应同时核对本章变量在后续模型中的定义和适用范围。

- 若你在整理调制、线宽与噪声主线，建议与 第二十三章、第二十九章 对照阅读。阅读时应同时核对本章变量在后续模型中的定义和适用范围。
- 若你在整理项目执行链，建议与 chapters 二十五至二十七、第三十章 对照阅读。阅读时应同时核对本章变量在后续模型中的定义和适用范围。

参考文献

- [1] Sadao Adachi. GaAs, AlAs, and Al_xGa_{1-x}As: Material parameters for use in research and device applications. *Journal of Applied Physics*, 58(3):R1–R29, 1985.
- [2] Doyeol Ahn. Theoretical study of strained InGaP quantum-well lasers. *Applied Physics Letters*, 66(5):628–630, 1995.
- [3] Filippo Alpeggiani, Nikhil Parappurath, Ewold Verhagen, and L. Kuipers. Quasinormal-mode expansion of the scattering matrix. *Physical Review X*, 7(2):021035, 2017.
- [4] Takeshi Aoki, Shota Ishimura, Takuya Inoue, Hiroyuki Yoshinaga, Makoto Ogasawara, Yuhki Itoh, Kosuke Fujii, Yusuke Sawada, Shun Kimura, Naoki Fujiwara, Hideki Yagi, Masaki Yanagisawa, Hidenori Takahashi, Takehiro Tsuritani, Ryohei Morita, Masahiro Yoshida, Menaka De Zoysa, Kenji Ishizaki, and Susumu Noda. Directly modulated c-band high-power photonic-crystal surface-emitting lasers for free-space optical communications. *Journal of Lightwave Technology*, pages 1–9, 2026.
- [5] Takeshi Aoki, Yuhki Itoh, Kosuke Fujii, Hiroyuki Yoshinaga, Naoki Fujiwara, Makoto Ogasawara, Yusuke Sawada, Rei Tanaka, Hideki Yagi, Masaki Yanagisawa, Masahiro Yoshida, and Takuya Inoue. High-power, stable single-mode CW operation of 1550 nm wavelength InP-based photonic-crystal surface-emitting lasers. *Applied Physics Express*, 17(4):042004, 2024.
- [6] Takeshi Aoki, Yuhki Itoh, Makoto Ogasawara, Kosuke Fujii, Shota Kimura, Yusuke Sawada, Hiroyuki Yoshinaga, Naoki Fujiwara, Hideki Yagi, Masaki Yanagisawa, Masahiro Yoshida, Takuya Inoue, Menaka De Zoysa, Kenji Ishizaki, and Susumu Noda. Narrow-linewidth 1.55- μ m-wavelength InP-based photonic-crystal surface-emitting lasers with near-diffraction-limited beam quality. *Optics Express*, 33(20):42480, 2025.
- [7] Dogukan Apaydin, Hjalmar Andersson, Lukas Uhlig, Sarina Graupeter, Joachim Ciers, Giulia Cardinali, Erik Strandberg, Tim Wernicke, Michael Kneissl, Ulrich Theodor Schwarz, Philippe Tassin, and Asa Haglund. Deep-uv photonic crystal surface-emitting lasers. *Laser & Photonics Reviews*, 20(6), 2025.
- [8] Jean-Pierre Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114(2):185–200, 1994.

- [9] Zijun Bian, Xingyu Zhao, Jingzhao Liu, Daehyun Kim, Adam F. McKenzie, Stephen Thoms, Paul Reynolds, Neil D. Gerrard, Aye S. M. Kyaw, James Grant, Katherine Rae, Jonathan R. Orchard, Calum H. Hill, Connor W. Munro, Pavlo Ivanov, David T. D. Childs, Richard J. E. Taylor, and Richard A. Hogg. Resonator embedded photonic crystal surface emitting lasers. *npj Nanophotonics*, 1(1), 2024.
- [10] Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović. *Diode Lasers and Photonic Integrated Circuits*. Wiley, 2 edition, 2012.
- [11] Francesco G. Della Corte, Giuseppe Cocorullo, Mario Iodice, and Ivo Rendina. Temperature dependence of the thermo-optic coefficient of InP, GaAs, and SiC from room temperature to 600 K at the wavelength of 1.5 μm . *Applied Physics Letters*, 77(11):1614–1616, 2000.
- [12] R. D. Esman and D. L. Rode. Semiconductor-laser thermal time constant. *Journal of Applied Physics*, 59(2):407–409, 1986.
- [13] Shanhui Fan and J. D. Joannopoulos. Analysis of guided resonances in photonic crystal slabs. *Physical Review B*, 65(23):235112, 2002.
- [14] Shanhui Fan, Weimin Suh, and J. D. Joannopoulos. Temporal coupled-mode theory for the fano resonance in optical resonators. *Journal of the Optical Society of America A*, 20(3):569–572, 2003.
- [15] V. G. Garcia and M. Farzaneh. Transient thermal imaging of a vertical cavity surface-emitting laser using thermoreflectance microscopy. *Journal of Applied Physics*, 119(4):045105, 2016.
- [16] C. Gautam, M. Pan, Y. Chen, T. J. Rotter, G. Balakrishnan, and W. Zhou. Laterally coupled photonic crystal surface emitting laser arrays. *Journal of Applied Physics*, 135(19), 2024.
- [17] Chhabindra Gautam, Mingsen Pan, Subhashree Seth, Thomas J. Rotter, Ming Zhou, Bradley J. Thompson, Ricky Gibson, Shanhui Fan, Ganesh Balakrishnan, and Weidong Zhou. Mode distribution impact on photonic crystal surface emitting laser performance. *APL Photonics*, 9(7), 2024.
- [18] C. J. Glassbrenner and Glen A. Slack. Thermal conductivity of silicon and germanium from 3 k to the melting point. *Physical Review*, 134(4A):A1058–A1069, 1964.
- [19] Charles Henry. Theory of the linewidth of semiconductor lasers. *IEEE Journal of Quantum Electronics*, 18(2):259–264, 1982.

- [20] Katsunori Hirose, Ying-Chung Liang, Xiangheng Kuang, Susumu Sakai, Kenji Iwahashi, and Susumu Noda. Watt-class high-power, high-beam-quality photonic-crystal lasers. *Nature Photonics*, 8(5):406–411, 2014.
- [21] K. M. Ho, C. T. Chan, and C. M. Soukoulis. Existence of a photonic gap in periodic dielectric structures. *Physical Review Letters*, 65(25):3152–3155, 1990.
- [22] Chia Wei Hsu, Bo Zhen, Jeongwon Lee, Song-Liang Chua, Steven G. Johnson, John D. Joannopoulos, and Marin Soljačić. Observation of trapped light within the radiation continuum. *Nature*, 499(7457):188–191, 2013.
- [23] Chia Wei Hsu, Bo Zhen, A. Douglas Stone, John D. Joannopoulos, and Marin Soljačić. Bound states in the continuum. *Nature Reviews Materials*, 1:16048, 2016.
- [24] Masaya Imada, Susumu Noda, Akihiko Chutinan, Mitsuru Mochizuki, Tamotsu Tanaka, and Hiroyuki C. Kurokawa. Coherent two-dimensional lasing action in surface-emitting laser with triangular-lattice photonic crystal structure. *Applied Physics Letters*, 75(3):316–318, 1999.
- [25] T. Inoue, M. Yoshida, M. D. Zoysa, K. Ishizaki, and S. Noda. Design of photonic-crystal surface-emitting lasers with enhanced in-plane optical feedback for high-speed operation. *Optics Express*, 28(4):5050, 2020.
- [26] Takuya Inoue, Ryohei Morita, Shota Ishimura, Shuei Nakano, Hidenori Takahashi, Takehiro Tsuritani, Menaka De Zoysa, Kenji Ishizaki, Masatoshi Suzuki, and Susumu Noda. Frequency-modulated high-power photonic-crystal surface-emitting lasers for long-distance coherent free-space optical communications. *Nature Photonics*, 19(12):1330–1335, 2025.
- [27] Jian-Ming Jin. *The Finite Element Method in Electromagnetics*. Wiley, 3 edition, 2015.
- [28] John D. Joannopoulos, Steven G. Johnson, Joshua N. Winn, and Robert D. Meade. *Photonic Crystals: Molding the Flow of Light*. Princeton University Press, 2 edition, 2008.
- [29] Steven G. Johnson and Joan D. Joannopoulos. Block-iterative frequency-domain methods for maxwell’s equations in a planewave basis. *Optics Express*, 8(3):173–190, 2001.
- [30] Ben King, Hans Wenzel, Eduard Kühn, Mindaugas Radziunas, and Paul Crump. Design of very-large area photonic crystal surface emitting lasers with an all-semiconductor photonic crystal. *Optics Express*, 32(25):44945, 2024.
- [31] Ashok Kodigala, Thomas Lepetit, Qing Gu, Babak Bahari, Yeshaiah Fainman, and Boubacar Kanté. Lasing action from photonic bound states in continuum. *Nature*, 541(7636):196–199, 2017.

- [32] Herwig Kogelnik and C. V. Shank. Coupled-wave theory of distributed feedback lasers. *Journal of Applied Physics*, 43(5):2327–2335, 1972.
- [33] Kirill Koshelev, Sergey Lepeshov, Mingkai Liu, Andrey Bogdanov, and Yuri Kivshar. Asymmetric metasurfaces with high-Q resonances governed by bound states in the continuum. *Physical Review Letters*, 121(19):193903, 2018.
- [34] Philip Trøst Kristensen, Rong-Chun Ge, and Stephen Hughes. Normalization of quasinormal modes in leaky optical cavities and plasmonic resonators. *Physical Review A*, 92(5):053810, 2015.
- [35] U. Kruger and K. Petermann. Dependence of the linewidth of a semiconductor laser on the mode distribution. *IEEE Journal of Quantum Electronics*, 26(12):2058–2064, 1990.
- [36] Ben Lang, Phillip D. Sewell, Ana Vuković, Karl Boylan, Samir Rihani, David Moodie, Richard Spalding, Nannicha Hattasan, and Graham Berry. Assessment and extension of rapid design tools for modeling photonic crystal surface-emitting lasers. *Journal of the Optical Society of America B*, 42(5):1123, 2025.
- [37] K. M. Leung and Y. F. Liu. Full vector wave calculation of photonic band structures in face-centered-cubic dielectric media. *Physical Review Letters*, 65(21):2646–2649, 1990.
- [38] P. T. Leung, S. Y. Liu, S. S. Tong, and K. Young. Time-independent perturbation theory for quasinormal modes in leaky optical cavities. *Physical Review A*, 49(4):3068–3073, 1994.
- [39] P. T. Leung, S. Y. Liu, and K. Young. Completeness and orthogonality of quasinormal modes in leaky optical cavities. *Physical Review A*, 49(4):3057–3067, 1994.
- [40] Lifeng Li. Use of fourier series in the analysis of discontinuous periodic structures. *Journal of the Optical Society of America A*, 13(9):1870–1876, 1996.
- [41] Ying-Chung Liang, Xue-Hua Peng, Susumu Sakai, Kenji Iwahashi, and Susumu Noda. Three-dimensional coupled-wave model for square-lattice photonic crystal lasers with TE polarization. *Optics Express*, 20(14):15945–15961, 2012.
- [42] Ying-Chung Liang, Xue-Hua Peng, Susumu Sakai, Kenji Iwahashi, and Susumu Noda. Three-dimensional coupled-wave model for triangular-lattice photonic crystal lasers with transverse-electric-like polarization. *Optics Express*, 21(8):10319–10335, 2013.
- [43] Takashi Matsubara, Shota Saito, Masanobu Yoshimoto, Yoshinori Tanaka, Toshiyuki Sato, and Susumu Noda. Gan photonic-crystal surface-emitting laser at blue-violet wavelengths. *Science*, 319(5871):445–447, 2008.

- [44] M. G. Moharam and T. K. Gaylord. Rigorous coupled-wave analysis of planar-grating diffraction. *Journal of the Optical Society of America*, 71(7):811–818, 1981.
- [45] M. G. Moharam, T. K. Gaylord, Eric B. Grann, and Drew A. Pommet. Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings. *Journal of the Optical Society of America A*, 12(5):1068, 1995.
- [46] David Moodie, Karl Boylan, Nannicha Hattasan, Samir Rihani, Stuart Pearce, Lu Qi, Sukhjiban Dosanjh, Eva Repiso Menendez, Matteo Silva, Richard Spalding, Stephen Burlinson, Martin Gillanders, David Turner, and Graham Berry. 1.55 μm wavelength band photonic crystal surface emitting laser with n-side photonic crystal and operation at up to 85 $^{\circ}\text{C}$. *Optics Express*, 32(6):10295, 2024.
- [47] Ryohei Morita, Shota Ishimura, Takuya Inoue, Kosuke Nishimura, Hidenori Takahashi, Takehiro Tsuritani, Menaka De Zoysa, Kenji Ishizaki, Masatoshi Suzuki, and Susumu Noda. High-speed high-power free-space optical communication via directly modulated watt-class photonic-crystal surface-emitting lasers. *Optica*, 11(7):971, 2024.
- [48] E. A. Muljarov, W. Langbein, and R. Zimmermann. Brillouin-wigner perturbation theory in open electromagnetic systems. *EPL (Europhysics Letters)*, 92(5):50010, 2010.
- [49] Steven C. Nicholes, James W. Raring, Erik J. Norberg, Chad S. Wang, Matthew M. Dummer, Steven P. DenBaars, and Larry A. Coldren. Highly polarized single-chip ELED sources using oppositely strained MQW emitters and absorbers. *IEEE Photonics Technology Letters*, 20(14):1267–1269, 2008.
- [50] Susumu Noda, Takuya Inoue, Masahiro Yoshida, John Gellera, Menaka De Zoysa, and Kenji Ishizaki. High-power and high-beam-quality photonic-crystal surface-emitting lasers: a tutorial. *Advances in Optics and Photonics*, 15(4):977–1032, 2023.
- [51] Susumu Noda, Masahiro Yoshida, Takuya Inoue, Menaka De Zoysa, Kenji Ishizaki, and Ryoichi Sakata. Photonic-crystal surface-emitting lasers. *Nature Reviews Electrical Engineering*, 1(12):802–814, 2024.
- [52] Lars Persson, Moritz Riedel, Asa Haglund, and Ulrich Theodor Schwarz. Finite-size effects in photonic-crystal surface-emitting lasers: Critical discussion of different approximations. *Optics Express*, 33(25):53098, 2025.
- [53] Xin Qi, Hai Huang, Chenyu Li, Kewei Shi, and Zhaoyu Zhang. POST: photonic swin transformer for automated and efficient prediction of PCSEL. *Nanophotonics*, 14(22):3599–3610, 2025.
- [54] Haoyu Qin, Weixuan Zhang, Shaohu Chen, Huizhen Zhang, Ruhao Pan, Junjie Li, Lei Shi, Jian Zi, and Xiangdong Zhang. Quasi-bound flat bands in the continuum. *Nature Communications*, 16(1), 2025.

- [55] Mindaugas Radziunas, Hans Wenzel, Ben King, Paul Crump, and Eduard Kühn. Dynamical simulations of single-mode lasing in large-area all-semiconductor PCSELs. *Optics Letters*, 50(6):1953, 2025.
- [56] Matthew N. Robinson, Stephen J. Sweeney, and Richard A. Hogg. Two-dimensional coupled wave theory for triangular lattice TM-polarised photonic crystal surface emitting lasers. *IEEE Journal of Selected Topics in Quantum Electronics*, 31(2):1–11, 2025.
- [57] Kazuaki Sakoda. *Optical Properties of Photonic Crystals*. Springer, 2 edition, 2005.
- [58] Fahrettin Sarcan, Yue Wang, Thomas F. Krauss, Tulin Erucar, and Ayse Erol. Dilute nitride resonant-cavity light emitting diode. *Optics & Laser Technology*, 122:105888, 2020.
- [59] Robert P. Sarzała and Łukasz Piskorski. Comparative analysis of lasing performance of oxide-confined and proton-implanted vertical-cavity surface-emitting diode lasers. *Applied Physics A*, 102(2):359–366, 2010.
- [60] Christophe Sauvan, Jean-Paul Hugonin, I. S. Maksymov, and Philippe Lalanne. Theory of the spontaneous optical emission of nanosize photonic and plasmon resonators. *Physical Review Letters*, 110(23):237401, 2013.
- [61] Anthony E. Siegman. *Lasers*. University Science Books, 1986.
- [62] Richard A. Soref and B. R. Bennett. Electrooptical effects in silicon. *IEEE Journal of Quantum Electronics*, 23(1):123–129, 1987.
- [63] Weimin Suh, Zheng Wang, and Shanhui Fan. Temporal coupled-mode theory and the presence of non-orthogonal modes in lossless multimode cavities. *IEEE Journal of Quantum Electronics*, 40(10):1511–1518, 2004.
- [64] Allen Taflove and Susan C. Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Artech House, 3 edition, 2005.
- [65] Natsuo Taguchi, Akinori Iwai, Masahiro Noguchi, Hiroaki Takahashi, Atsuo Michiue, Menaka De Zoysa, Takuya Inoue, Kenji Ishizaki, and Susumu Noda. Green-wavelength GaN-based photonic-crystal surface-emitting lasers. *Applied Physics Express*, 17(1):012002, 2024.
- [66] Hakan E. Türeci, A. Douglas Stone, and Li Ge. Theory of the spatial structure of nonlinear lasing modes. *Physical Review A*, 76(1):013813, 2007.
- [67] Hakan E. Türeci, A. Douglas Stone, Li Ge, Stefan Rotter, and Robert J. Tandy. Ab initio self-consistent laser theory and random lasers. *Nonlinearity*, 22(1):C1–C18, 2009.

- [68] Mohammad Fazel Vafadar and Songrui Zhao. Ultralow threshold surface emitting ultraviolet lasers with semiconductor nanowires. *Scientific Reports*, 13:6633, 2023.
- [69] Ziyi Wang, Xia Liu, Pinyao Wang, Huanyu Lu, Bo Meng, Wei Zhang, Lijie Wang, Yanjing Wang, and Cunzhu Tong. Continuous-wave operation of 1550 nm low-threshold triple-lattice photonic-crystal surface-emitting lasers. *Light: Science & Applications*, 13(1):44, 2024.
- [70] Hans Wenzel, Eduard Kühn, Ben King, Paul Crump, and Mindaugas Radziunas. Theory of the linewidth–power product of photonic–crystal surface–emitting lasers. *IEEE Journal of Quantum Electronics*, 61(1):1–14, 2025.
- [71] Jingxiao Xu, Douglas McCulloch, and Martin D. B. Charlton. Modeling full PCSELs and VCSELs using modified rigorous coupled-wave analysis. *Optics Express*, 32(13):22169, 2024.
- [72] W. Yang, Xue-Hua Peng, Ying-Chung Liang, Susumu Sakai, Kenji Iwawashi, and Susumu Noda. Three-dimensional coupled-wave analysis of TM-like polarization in two-dimensional square-lattice photonic crystal slab structures. *Optics Express*, 22(10):12326–12342, 2014.
- [73] Yi Yang, Chao Peng, Yong Liang, Zhengbin Li, and Susumu Noda. Analytical perspective for bound states in the continuum in photonic crystal slabs. *Physical Review Letters*, 113(3):037401, 2014.
- [74] Koji Yoshida, Menaka De Zoysa, Kento Asano, Bongshik Song, Yasutomo Ota, Kenji Ishizaki, Yoshinori Tanaka, and Susumu Noda. High-brightness scalable continuous-wave single-mode photonic-crystal laser. *Nature*, 618(7965):727–732, 2023.
- [75] Masahiro Yoshida, Menaka De Zoysa, Kenji Ishizaki, Yoshinori Tanaka, Masato Kawasaki, Ranko Hatsuda, Bongshik Song, John Gellela, and Susumu Noda. Double-lattice photonic-crystal resonators enabling high-brightness semiconductor lasers with symmetric narrow-divergence beams. *Nature Materials*, 18(2):121–128, 2019.
- [76] Cheng Zhang, Huijie Li, and Dong Liang. Antireflective vertical-cavity surface-emitting laser for LiDAR. *Nature Communications*, 15(1):1105, 2024.
- [77] Xiaoyu Zhang, Xuefan Yin, Kangjiang Yang, Zixuan Zhang, Peishen Li, Feifan Wang, Junyi Gu, Zihao Chen, Hongbin Li, and Chao Peng. Electrically driven heterogeneous III-V/Si photonic crystal surface-emitting laser. *Optics Letters*, 50(22):7083, 2025.
- [78] Y. M. Zhang, R. Mirin, A. M. Sergent, J. B. Schlager, B. E. Hammons, C. W. Wilmsen, K. L. Lear, S. P. Kilcoyne, R. P. Schneider, and K. D. Choquette. Cryogenic performance of double-fused 1.5- μm vertical cavity lasers. *Journal of Lightwave Technology*, 17(3):503–508, 1999.

- [79] Bo Zhen, Chia Wei Hsu, Ling Lu, A. Douglas Stone, and Marin Soljačić. Topological nature of optical bound states in the continuum. *Physical Review Letters*, 113(25):257401, 2014.
- [80] Wei Zhou and Shuming Pan. Photonic-crystal surface-emitting lasers: Review and introduction of modulated structures. *Applied Physics Letters*, 123(14):140501, 2023.