基于垂直腔面发射激光器的光无线系统中三维覆盖的Q学习

Hossein Safi，IEEE会员，Rizwana Ahmad，Iman Tavakkolnia，IEEE高级会员，以及Harald Haas，IEEE会士

arXiv:2602.03526v1 [math.OC] 3 Feb 2026

摘要—光束发散角控制是在接收器高度变化时维持室内光无线通信（OWC）系统可靠覆盖的关键因素。传统系统采用固定发散角，由于光功率集中与空间扩散之间的非凸权衡，会导致显著的覆盖性能下降。本文提出一种基于强化学习的框架，用于基于垂直腔面发射激光器（VCSEL）的光无线通信网络中的动态发散适应。通过与环境持续交互，强化学习智能体自主学习接收器高度与光束发散之间的近最优映射，从而无需解析建模或计算密集型穷举搜索。仿真结果表明，所提方法在较低接收器高度下可实现高达92%的覆盖，并在挑战性条件下保持稳健性能，从而为下一代光无线系统中密集垂直腔面发射激光器阵列部署实现可扩展、实时且高能效的光束控制。

索引词——自适应覆盖控制、波束发散优化、光无线通信、强化学习、垂直腔面发射激光器阵列。

I. 引言

6G 及后续技术对高容量、低时延连接的需求不断增长，正推动无线技术创新。在新兴接入技术中，光无线通信（OWC）作为射频（RF）系统的一种有前景的补充，凭借免许可频谱、增强的物理层安全性，以及实现远超 100 Gb/s 数据速率的潜力，受到了广泛关注 [1],[2]。近期，垂直腔面发射激光器（VCSEL）被提出作为高性能 OWC 接入点（AP）的关键使能技术，这得益于其高调制带宽、高能效以及可扩展的紧凑阵列设计 [3], [4]。垂直腔面发射激光器阵列能够产生多个窄而高度定向的光束，同时服务于室内环境中的不同区域，从而使总数据速率超过每秒太比特级别 [4]。其圆形光束对称性和晶圆级可制造性，使其非常适合构建具有独立寻址发射器的结构化密集网格阵列 [5]。

作者就职于英国剑桥大学工程系 LiFi 研发中心，地址：英国剑桥；电子邮箱：{hs905,ra714, it360, huh21}@cam.ac.uk。

尽管具有这些优势，实际的基于垂直腔面发射激光器的光无线通信系统仍面临若干与空间动态和光束几何形状相关的关键挑战。现有阵列设计通常依赖固定的光学配置，例如焦距恒定的透镜，并针对特定接收平面高度进行优化 [5]–[7]。因此，当用户垂直移动或接收平面发生变化时，相邻光束的重叠和发散将变得不理想，导致覆盖性能严重下降、光束间干扰以及信号功率损耗。此外，激活的垂直腔面发射激光器数量及其光束宽度通常是按照单一高度下保证覆盖的要求来选择的，这会造成不必要的能耗，并降低其他高度下的效率。在密集部署中，这些问题会进一步加剧，因为光束发散必须在两个相互冲突的目标之间取得平衡：最大化覆盖面积，同时尽量减少光学干扰。

为解决这些局限，动态光束发散角控制已成为未来光无线通信网络的一项关键设计目标。自适应光学元件（如电控可调液晶透镜和变焦光学器件）的最新进展，使实时调节每个垂直腔面发射激光器光束的发散角成为可能 [8],[9]。然而，确定最优发散配置仍是一个非凸问题，发射器几何结构、传播损耗和空间干扰之间存在强耦合。传统的解析方法或基于梯度的优化方法并不适用于此类情形，因为目标函数是分段的、不可微的，并且高度依赖环境 [10]。

受这些挑战的启发，本文提出一种由人工智能（AI）驱动的方法，利用 Q学习自主优化光束发散角，从而在基于垂直腔面发射激光器的光无线通信系统中实现动态三维（3D）覆盖。具体而言，我们采用强化学习（RL）来构建光束选择与激活问题。强化学习是一类序贯决策方法，智能体通过与环境交互学习最优策略，以最大化长期奖励。不同于静态配置或预先计算的查找表，所提出的 Q-learning 框架能够学习自适应的状态-动作策略，直接将环境条件（例如接收器高度）映射到最优发散角设置。这种无模型表述无需显式

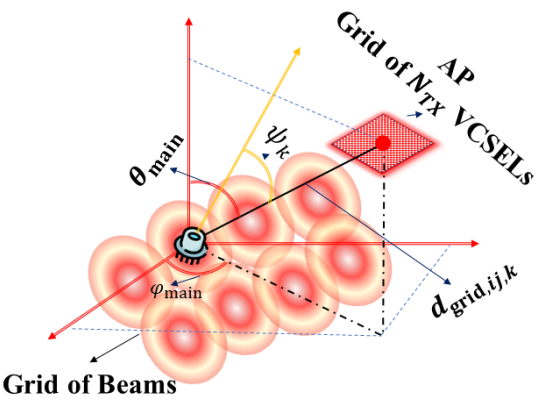

图1. 基于垂直腔面发射激光器的下行传输配置

信道建模；与穷举搜索或迭代优化方法相比，它还能以较低的推理复杂度实现近乎实时的自适应。

通过全面的仿真验证了所提出方法的有效性和实用性，结果表明，该方法在不同接收器高度和部署场景下均能实现近最优覆盖并快速收敛。具体而言，仿真结果表明，所提出的方法具有近最优的覆盖性能：在较低接收器高度下覆盖率最高可达 92%，在具有挑战性的条件下也可达 50%；与穷举优化相比，该方法将在线计算成本降低了数个数量级。这些结果证实，基于人工智能的学习能够变革光无线系统中的波束管理，为未来 6G环境中的智能、可扩展、自优化垂直腔面发射激光器接入网络铺平道路。

本文其余部分安排如下。第二节介绍系统模型和光束传播特性。第三节建立波束发散优化问题，并阐述采用自适应学习的必要性。第四节讨论基于人工智能的学习框架。第五节介绍仿真设置和数值结果。最后，第六节总结全文并展望未来研究方向。

II. 系统模型

A. 几何配置与发射器阵列

系统在由三维空间 $L \times W \times H$ 所界定的矩形室内环境中运行。天花板中央安装了一个集中式接入点（AP），用于支持下行通信。该 AP 采用由 N_{TX} 垂直腔面发射激光器 [6],[10], 组成的结构化网格，从而在接收平面上形成相应的光束网格，如图 1 所示。第 k 个垂直腔面发射激光器光束由其源位置 $(x_{\mathrm{source},\,k},\ y_{\mathrm{source},\,k},\ z_{\mathrm{source},\,k})$ 主动转向。因此，主光束方向由极角 θ_{main} 和方位角 ϕ_{main} 确定，形成波束方向向量$\mathbf{d}_{\mathrm{beam},\,k}$，其表示为

$$\mathbf{d}_{\mathrm{beam},k} = \begin{bmatrix} \sin(\theta_{\mathrm{main}})\cos(\phi_{\mathrm{main}}) \\ \sin(\theta_{\mathrm{main}})\sin(\phi_{\mathrm{main}}) \\ \cos(\theta_{\mathrm{main}}) \end{bmatrix}. \tag{1}$$

接收器面板上的感兴趣区域被离散为网格点 (x_i, y_j)，分辨率分别为 Δ_x和 Δ_y。从第 k个发射器到网格点 (i, j) 的空间向量定义为

$$\mathbf{d}_{\mathrm{grid},ij,k} = \begin{bmatrix} x_i - x_{\mathrm{source},k} \\ y_j - y_{\mathrm{source},k} \\ z_{ij} - z_{\mathrm{source},k} \end{bmatrix}, \tag{2}$$

其中 $d_{ij,k} = ||\mathbf{d}_{\mathrm{grid},ij,k}||$ 为欧几里得距离。该几何表述为刻画波束传播奠定了基础，其中每个网格点处的接收辐照度由高斯光束的发散和角度对准情况决定。

B. 高斯光束传播与接收辐照度

设 $\theta_{\mathrm{diff},ij,k}$ 表示波束轴 $\mathbf{d}_{\mathrm{beam},k}$ 与指向网格点$\mathbf{d}_{\mathrm{grid},ij,k}$的方向向量之间的角偏差，定义为

$$\theta_{\mathrm{diff},ij,k} = \arccos\left(\frac{\mathbf{d}_{\mathrm{beam},k}\cdot\mathbf{d}_{\mathrm{grid},ij,k}}{\|\mathbf{d}_{\mathrm{beam},k}\|\,\|\mathbf{d}_{\mathrm{grid},ij,k}\|}\right). \tag{3}$$

接收平面处的光束强度遵循高斯光束传播规律，光束半径随距离变化为$w(z)$。因此，发射器 k 在网格点 $(i,\ j)$ 处产生的接收辐照度为

$$I_{\mathrm{grid},ij,k} = \frac{2P_t}{\pi\, w^2(d_{ij,k})}\exp\left[-2\left(\frac{d_{ij,k}\,\theta_{\mathrm{diff},ij,k}}{w(d_{ij,k})}\right)^2\right], \tag{4}$$

其中，P_t 为发射光功率。这里，$w(d_{ij,\ k})$描述了光束随距离的远场扩展，而指数项则刻画了高斯分布的角度衰减。利用 $w(d_{ij,\ k}) \approx d_{ij,\ k}\,\theta_{\mathrm{divergence}}$ [5]，式（4）中的辐照度可简化为

$$I_{\mathrm{grid},ij,k} \approx \frac{2P_t}{\pi[\theta_{\mathrm{divergence}} d_{ij,k}]^2}\exp\left[-2\left(\frac{\theta_{\mathrm{diff},ij,k}}{\theta_{\mathrm{divergence}}}\right)^2\right]. \tag{5}$$

如式（5）所示，发散角决定空间覆盖范围。增大发散角会使光束和覆盖区域变宽，但也会因功率泄漏而加剧干扰。因此，必须进行优化，以平衡覆盖与干扰。

C. 接收器前端模型与最终接收功率

为了得到最终收集到的光功率，必须应用接收器前端模型。假设接收器水平安装，因此其法向量与入射光线之间的入射角为 Ψ_k 。来自第 k 个发射器的总接收功率计算如下： [5]

$$P_{\mathrm{rec},k} = I_{\mathrm{grid},k}\cdot A_d\cdot\cos(\Psi_k)\cdot G_{\mathrm{conc}}\cdot\Pi\left(\frac{\Psi_k}{\Psi_{\mathrm{FoV}}}\right), \tag{6}$$

其中， A_d 为有效光电探测器面积， G_{conc} 为光学聚光器增益， Ψ_{FoV} 为接收器视场角（FoV）。因子 $\cos(\Psi_k)$ 表示几何损耗，而门函数 $\Pi(\cdot)$ 确保仅收集位于接收器视场角内的光线。

III. 发散角控制的优化框架

A. 问题表述

如前所述，目标是在给定环境状态 s 下，确定使空间覆盖最大化的最优发散角 θ_{opt}。状态 s 主要由接收器高度 h_r 决定，但也可能包含链路配置的其他条件。覆盖率 C 定义为满足最低信号干扰噪声比（SINR）阈值 Γ_{th} 的网格点所占的比例。因此，优化问题可表示为

$$\underset{\theta_{\text{divergence}} \in \mathcal{A}}{\text{maximize}} \quad C(\theta_{\text{divergence}}|s), \tag{7}$$

其中， $\mathcal{A}$ 是受光学硬件[1]约束的发散角可行集。覆盖函数定义为

$$C(\theta_{\text{divergence}}|s) = \frac{1}{N_{\text{grid}}} \sum_{i,j} \mathbb{I}(\text{SINR}_{i,j}(\theta_{\text{divergence}}) \geq \Gamma_{\text{th}}), \tag{8}$$

其中， N_{grid} 表示离散网格位置的数量， $\mathbb{I}(\cdot)$ 表示指示函数。

在每个网格点（i， j），SINR 定义为

$$\begin{aligned} \text{SINR}_{ij}(\theta_{\text{divergence}}) &= \frac{P_{\text{rec},ij,k^*}(\theta_{\text{divergence}})}{\sum_{l \neq k^*} P_{\text{rec},ij,l}(\theta_{\text{divergence}}) + N_0}, \\ k^* &= \arg\max_k P_{\text{rec},ij,k}(\theta_{\text{divergence}}). \end{aligned} \tag{9}$$

其中，$P_{\text{rec},k,i,j}$ 表示发射器 k在位置（i、 j）处的接收功率， k^* 表示主导波束的折射率。分母同时包含残余干扰和恒定噪声功率 N_0。因此，优化问题是在服务区域内确定 θ_{opt}，使 C 最大化。

B. 分析挑战与动态特性

由于结构和环境的复杂性，使用传统方法求解式 (7) 中的优化问题难以实现。覆盖函数 $C(\theta_{\text{divergence}})$ 是非线性且非凸的，而最大化算子 $\max_k\{P_{\text{rec},k}\}$ 会因服务波束索引突然切换而引入不连续性。此外，指示函数 $\mathbb{I}(\cdot)$ 产生分段常数且不可微的目标函数。这些性质不符合基于梯度的优化方法所需的假设，因此这类方法无法应用于此情形 [12]。更具体地，令 $f(\theta_{\text{divergence}})$ =信干噪比$_{i,\ j}(\theta_{\text{divergence}}) - \Gamma_{\text{th}}$。则覆盖函数可写为

$$C(\theta_{\text{divergence}}) = \frac{1}{N_{\text{grid}}} \sum_{i,j} \mathbb{I}(f(\theta_{\text{divergence}}) \geq 0). \tag{10}$$

由于 $\mathbb{I}(\cdot)$ 在零点处不连续，而在其他位置保持平坦，梯度 $\partial C/\partial \theta_{\text{divergence}}$ 在经典意义下并不存在，次梯度近似也无法为优化提供有用的方向。这导致了一个类似于混合整数规划问题的情形，而众所周知，这类问题是 NP难问题 [12]。

除了在数学上难以求解之外，该问题本质上也是动态的。最优发散角 θ_{opt}取决于瞬时环境状态 s。随着接收器高度变化，链路几何结构也会随之改变，先前的最优设置可能不再最优。针对单一运行条件优化的静态发散角，会导致其他条件下的覆盖性能显著下降。同样，在时间变化快速的环境中，依赖预先计算的查找表并不足够。因此，该优化问题既具有非凸的结构特性，也具有非平稳环境动态，这促使我们采用自适应学习型控制方法，而非闭式解或静态解。

IV. 强化学习问题表述

为应对上一节讨论的解析求解难题和环境动态变化这两大挑战，我们采用强化学习框架。该方法将发散角控制任务重新表述为一个序贯决策过程，使发射器能够学习一种自适应策略，将瞬时状态（接收器高度）映射为最优发散角。通过与定制环境进行迭代交互，强化学习智能体不断改进其控制策略，并逐渐收敛到一种能够最大化基于覆盖性能的期望累积奖励的策略。这种基于强化学习的方法通过与环境直接交互进行学习，具备实时适应能力，使系统能够持续优化性能，而无需依赖显式解析模型或先验信道知识。波束选择问题被表述为一个有限状态马尔可夫决策过程，其离散动作空间对应于可用的垂直腔面发射激光器光束。在这些条件下，表格型Q学习提供了一种自然且高效的解决框架。它具有较低的计算复杂度、稳定的收敛特性，并且所学习的动作价值函数可直接解释。

[1]实际上，发散角可以通过集成在发射器侧的光学子系统进行动态调节，例如液晶透镜、变焦光学器件或自适应光束成形元件；这些器件能够实时对发射光束的轮廓进行精细控制 [8], [9], [11]。

A. 马尔可夫决策过程建模

发散角优化任务被建模为一个有限马尔可夫决策过程（MDP），由以下元组定义

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, R, \gamma), \tag{11}$$

其中每个元素表示决策环境的一个关键组成部分，具体如下:

- 状态空间： $\mathcal{S} = \{s_1, s_2, \ldots, s_{N_S}\}$ 表示可观测环境状态的集合。在本研究中，每个状态对应于一个离散接收器高度。这种离散化刻画了典型室内环境中的代表性工作高度，用户可能将设备握持或放置在桌面高度、胸部高度或站立高度。离散状态假设简化了学习过程，同时保留了链路几何的关键空间变化。不过，该框架可以方便地扩展，以纳入连续高度值或其他上下文参数。

- 动作空间： $\mathcal{A} = \{\theta_1, \ldots, \theta_{N_A}\}$ 表示可行的发散角。 $\mathcal{A}$ 的范围与分辨率由阵列中垂直腔面发射激光器的光学设计和驱动限制决定，这些因素限定了实时发散调节的幅度与粒度。

- 转移概率： $P(s'|s,a)$定义了在状态s下执行动作 a 后转移到状态 s' 的概率。我们假设垂直腔面发射激光器发散控制动作 a 不会影响该转移，因为用户移动动态独立于光学控制策略 [13]。这种解耦简化了马尔可夫决策过程的时间结构，并使智能体能够专注于在观测状态 s下最大化瞬时覆盖奖励$R(s,a)$，从而与为每个离散高度确定最优静态发散角的目标相一致。

- 奖励函数： $R(s,\ a)$ 用于量化在状态 s 下执行动作 a 后的系统性能。其定义为

$$R(s,a) = 100 - \mathrm{Hole}(s,a), \tag{12}$$

其中，$\mathrm{Hole}(s,\ a)$ 表示由发散角配置导致的未覆盖网格点（覆盖盲区）所占比例。如果某个网格点对应的信干噪比低于预定义阈值 Γ_{th}，则认为该点未被覆盖。因此，最大化 $R(s,\ a)$ 可直接提升空间覆盖。

- 折扣因子： $\gamma \in [0,1]$ 用于控制未来奖励的权重。鉴于状态演化具有准静态和确定性特征，优化主要由即时覆盖奖励主导，因此采用适度的折扣因子，以确保学习稳定，同时保留对各状态性能的敏感性。

随后，发射端的强化学习智能体旨在学习一种最优策略

$$\pi^* : \mathcal{S} \to \mathcal{A}, \tag{13}$$

以最大化预期的累积折扣奖励

$$J(\pi) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right]. \tag{14}$$

该策略使发射器能够根据每种观测到的接收器条件，自主选择可实现最高预期覆盖的发散角 θ_{opt}。

需要注意的是，尽管状态转移不受智能体动作的影响，该问题仍在强化学习框架下进行建模，以体现时变条件下波束配置所具有的序贯决策特性。在状态转移完全独立于动作这一特殊情形下，该建模可简化为上下文老虎机问题。然而，采用强化学习建模可直接扩展至更现实的场景，例如涉及相关移动、受限移动或波束选择决策之间存在时间耦合的场景，而无需改变核心框架。

B. Q-learning 框架

我们采用无模型 Q-learning 算法来估计最优动作价值函数 $Q^*(s,\ a)$，该函数表示在状态 s 下执行动作 a，并在此后遵循最优策略所能获得的预期长期回报。Q-learning 更新规则表示为[13]

$$Q(s,a) \leftarrow Q(s,a) + \alpha\big[R(s,a) + \gamma \max_{a'} Q(s',a') - Q(s,a)\big], \tag{15}$$

其中，α 表示学习率。在多个回合中反复迭代后，$Q(s,\ a)$会收敛到 $Q^*(s,\ a)$，由此可得最优策略:

$$\pi^*(s) = \arg\max_a Q^*(s,a). \tag{16}$$

此外，还采用 ϵ-贪心探索策略，以平衡探索与利用。探索率随训练回合呈指数衰减

$$\epsilon = \epsilon_{\min} + (\epsilon_{\max} - \epsilon_{\min}) e^{-\lambda n_{\mathrm{ep}}}, \tag{17}$$

其中，n_{ep} 为回合索引，λ 为衰减率。这确保智能体在训练初期进行广泛探索，并在后期专注于微调策略。完整流程总结于算法 1 中。

C. 实现与收敛行为

所提出的学习框架通过自定义环境实现，该环境准确复现了第 II 节所述的几何配置、波束传播特性和接收器前端模型。在训练过程中，每个回合依次遍历离散接收机状态，同时根据每个发散动作获得的覆盖奖励迭代更新 Q 表。如算法 1 所述，智能体反复与自定义环境交互，采用 ϵ-贪心策略，在探索与

利用之间取得平衡。在每一步中，观测到的覆盖奖励均根据公式 (15) 更新 Q值。通过持续交互，Q 表收敛至对最优动作价值函数的稳定估计，从而得到策略 π^*，该策略将每个环境状态映射到相应的最优发散角

$$\theta_{\text{opt}} = \pi^*(s_{\text{current}}). \tag{18}$$

通常在 10^3–10^4 个回合内收敛，具体取决于探索设置。训练完成后，发散度自适应调整的计算效率很高，只需查询一次 Q 表。强化学习智能体在非平稳条件下仍保持稳健，并通过交互持续优化其策略。

算法 1 自适应发散控制的 Q学习

要求：状态集 $\mathcal{S}$、动作集 $\mathcal{A}$；学习率 α；折扣因子 γ；回合 N_{ep}；探索参数 $(\epsilon_{\max}, \epsilon_{\min}, \lambda)$.

初始化 Q 表 $Q(s,a) \leftarrow 0,\ \forall s \in \mathcal{S}, a \in \mathcal{A}$。
初始化环境 $\mathcal{E}$ (VCSELEnv)。
对 $n = 1$ 到 N_{ep} do
  $s \leftarrow \mathcal{E}$.重置() ▷从初始状态开始回合
  $\epsilon \leftarrow \epsilon_{\min} + (\epsilon_{\max} - \epsilon_{\min}) e^{-\lambda n}$ ▷逐回合衰减
  完成 ←false
  当未完成时执行
    ▷ ϵ-贪心动作选择
    如果 Uniform(0,1) $< \epsilon$ 则
      $a \leftarrow$随机($\mathcal{A}$) ▷ 探索
    else
      $a \leftarrow \arg\max_{a' \in \mathcal{A}} Q(s,a')$ ▷利用；打破平局任意地
    结束 if
    $(s', R,$ 完成$) \leftarrow \mathcal{E}$.步长(a) ▷应用发散动作 ，观察奖励
    ▷贝尔曼更新（不进行自举，除非处于终止状态）
    如果已结束，则
      $Q(s,a) \leftarrow Q(s,a) + \alpha\big[R - Q(s,a)\big]$
    else
      $Q(s,a) \leftarrow Q(s,a) + \alpha\big[R + \gamma$最大值$_{a' \in \mathcal{A}} Q(s',a') - Q(s,a)\big]$
    结束如果
    $s \leftarrow s'$
  结束当
结束对于
输出： $\pi^*(s) = \arg\max_{a \in \mathcal{A}} Q(s,a)$。

V. 仿真结果与讨论

我们使用第二节中描述的模型，在仿真室内环境中评估所提出的方法。仿真参数列于表 I。另需说明的是，本文的分析聚焦于受控单接入点室内场景，其中接收机平面固定，用户高度状态为准静态。特意选择这种设置，是为了单独考察

表 I 仿真参数。

参数	值
房间（$L \times W \times H$）	$8 \times 8 \times 3$ 米
网格分辨率 Δ	0.2 米
发射器 N_{TX}	225（15×15）
单个垂直腔面发射激光器的功率 P_t	10 毫瓦
探测器面积 A_d	10^{-4} m^2
视场角 / 折射率 (Ψ_{FoV}, n)	75°, 1.5
强化学习 $(\alpha, \gamma, N_{\text{ep}})$	0.1, 0.9, 2000
强化学习（$\epsilon_{\max}$，$\epsilon_{\min}$，λ）	1, 0.01, 0.005

光束发散角控制和基于学习的决策对覆盖性能的影响。本研究不考虑用户移动性、阻塞和多用户干扰等动态效应，这些问题将留待未来工作研究。

首先，为了进一步深入了解优化光束发散角的重要性，图2展示了固定垂直腔面发射激光器发散角为 $\theta_{\text{divergence}} = 15^\circ$ 时的模拟三维覆盖性能。该图清晰揭示了室内光无线通信环境中静态光束配置的根本局限性。当接收器高度从 0.5 m 增加到 2.0 m 时，整体覆盖质量从约 80% 急剧下降至 48%。这种退化源于采用固定的发散角。在较低高度（$h_r = 0.5$ m）下，该15° 光束过宽，会产生强烈的光束间干扰和分散的覆盖盲区。相反，在较高高度（$h_r = 2.0$ m）下，相同的发散角无法提供足够的空间照明，从而导致接收平面上的功率损耗和覆盖不完整。这些结果表明，光束功率集中度与空间扩散之间的权衡对链路的垂直几何结构高度敏感。最佳发散角会随接收器高度动态变化，但由于发射器阵列的离散特性以及不可微的覆盖函数，这种依赖关系具有非线性、非凸性，且难以通过解析方法求解。为解决这一局限性，我们提出的基于强化学习的控制框架能够自主且自适应地调整发散角。

A. 收敛行为与学习稳定性

所提出的基于强化学习的发散控制框架的收敛行为如图3所示。该图同时展示了平均奖励（代表所达到的覆盖率）和探索率 ϵ 随训练回合的变化情况。智能体一开始采用完全探索策略（$\epsilon = 1$），由于随机探索次优动作，早期奖励出现大幅波动。随着训练的推进，探索率呈指数衰减，从而使智能体能够逐步利用累积经验并强化高奖励动作。

相应地，覆盖奖励的移动平均值呈现出明显的上升趋势，并平稳收敛

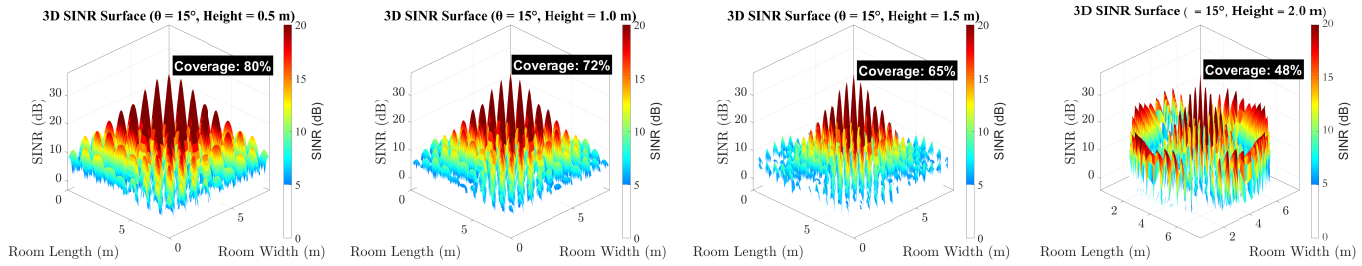

图2. 不同高度下 $\theta_{\text{divergence}} = 15°$ 的覆盖性能。这里， $\Gamma_{\text{th}} = 5$ dB，零空间表示信干噪比 $< \Gamma_{\text{th}}$的区域。

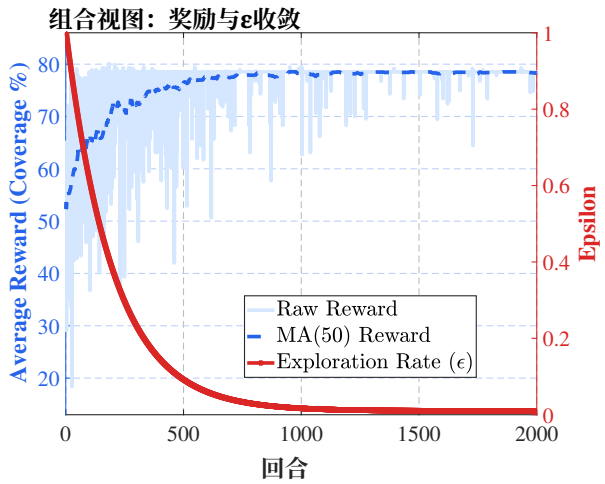

图3. 所提出的基于强化学习的发散控制框架的收敛行为。蓝色曲线显示每回合平均覆盖奖励的变化（采用50回合移动平均（MA）），红色曲线表示探索率（ϵ）。

至约 85% 的稳定平台。这表明智能体已成功学习到环境状态（接收器高度）与最优发散角 $\theta_{\text{divergence}}$ 之间的一致映射。稳态阶段未出现明显振荡，进一步证明了学习得到的策略具有稳定性和鲁棒性。不过需要指出的是，在所考虑的配置中，覆盖率并未达到 100%。这一局限并非源于学习框架本身，而是由所考虑系统模型的物理约束造成的：每个用户位置仅由一束垂直腔面发射激光器光束照射。若要进一步提升覆盖率，尤其是在照明重叠或较弱的区域，则需要采用多光束支持（即为每个用户分配多个发射器） [10]。此类配置能够提供冗余和空间分集，从而减少覆盖空隙，并使系统性能更接近充分利用空间。

B. 结果与学习得到的发散策略

收敛曲线凸显了两个关键观察。首先，强化学习智能体通过 ϵ的自适应衰减有效平衡探索与利用，从而促进快速策略改进，尽管环境具有不可微特性。其次，所得策略达到与穷举优化方法相当的近最优性能，但计算复杂度大幅更低，且不依赖先验系统知识。这些发现验证了所提出的强化学习框架在为动态三维光无线通信系统自主优化光束发散方面的有效性。

为进一步确认学习得到的策略的有效性并评估其物理一致性，表 II 总结了在不同接收器高度下获得的最优发散角及对应覆盖率。从表中可以看出，强化学习智能体成功捕捉到最优发散角与接收器高度之间的非线性关系，从而自适应地平衡光束扩展与干扰。因此，在较低高度（$h_r = 0.5$–1.0 m）下，策略会选择相对较窄的光束角（8°–9°），有效集中光功率并最大限度地减少光束间重叠。随着接收器高度增加，策略学会逐渐增大发散角，在 $h_r = 2.0$ m 处达到 14°，从而实现光束的几何扩展，并在整个接收平面上维持充足的照明。

从定量结果来看，学习得到的策略实现了与穷举优化所得结果高度接近的覆盖水平，在大多数高度下的偏差不超过3%。例如，在 $h_r = 0.5$ 米和 $h_r = 1.0$ 米处，学习得到的 9° 和 8° 发散度分别实现了92.39%和85.72%的覆盖率，与穷举搜索基线的93.33%和88.81%几乎相同。即使在最具挑战性的配置（$h_r = 2.0$ 米）下，强化学习策略仍保持了50.03%的覆盖率，而穷举搜索（ES）为53.8%，这反映的是单光束运行固有的几何局限，而非学习框架的局限。在较高接收器高度下观察到的覆盖概率饱和现象（在2米处约为50%），主要是每用户单光束传输模型所致，而不是学习框架本身的局限。这些结果为未来扩展至多光束或协作传输策略提供了动力，在这些场景中，

表 II 不同接收器高度下学习得到的最优发散角和覆盖性能

h_r [m]	强化学习得到的策略		穷举基线	
	$\theta_{\text{divergence}}$ [°]	覆盖率 [%]	$\theta_{\text{divergence}}$ [°]	覆盖率 [%]
0.5	9	92.39	10	93.33
1.0	8	85.72	10	88.81
1.5	9	85.96	8	89.70
2.0	14	50.03	11	53.80

表 III 穷举搜索与基于强化学习的发散控制的计算复杂度分析

方法	离线	在线
ES	$\mathcal{O}(\lvert\mathcal{A}\rvert N_{\text{grid}} N_{\text{TX}})$	$\mathcal{O}(1)$
强化学习（Q学习）	$\mathcal{O}\left(N_e \cdot \lvert\mathcal{S}\rvert\right)$	$\mathcal{O}(1)$

两种方法都需要 $\mathcal{O}\left(\lvert\mathcal{S}\rvert \times \lvert\mathcal{A}\rvert\right)$ 内存。 N_e：训练回合；N_{grid}：网格点。较大的 N_{grid} 和 N_{TX} 会使穷举搜索的离线成本高得令人难以承受。

多个垂直腔面发射激光器单元协同服务于用户或区域。

尽管所提出的Q学习公式基于离散化的接收器高度运行，并且有限的光束发散动作集合限制了其向完全连续或快速时变环境直接泛化的能力，但它能够实现稳定学习、较低的计算开销以及清晰的物理可解释性。利用函数逼近或深度强化学习扩展至连续状态空间是自然的下一步，但本文有意避免采用这些方法，以保持分析透明性，并避免在小样本场景下出现过拟合。

除了准确性之外，所提出的强化学习方法还具有计算优势，这对于扩展到真实环境并缓解环境漂移至关重要。虽然穷举搜索和强化学习策略的在线推理复杂度在技术上均为 $\mathcal{O}(1)$（单次查表），但这种等价性忽略了初始生成穷举搜索表所需的不切实际的离线成本；该成本会随着 $\mathcal{O}(|\mathcal{A}|N_{\text{grid}}N_{\text{TX}})$ 增长，变得高得难以承受。每当环境发生变化时（例如，新的平面布局或大型家具改变了有效光信道），都必须支付这笔高昂成本。相比之下，强化学习框架将巨大的计算负担转移到一个高效、数据驱动的训练过程中。强化学习的训练复杂度仅随回合数和状态数增长（$\mathcal{O}(N_e|\mathcal{S}|)$），而与昂贵的空间采样（$N_{\text{grid}}$ 和 N_{TX}）无关。这种对物理空间的极低内存需求和训练依赖性表明，对于密集的垂直腔面发射激光器阵列和高速运行而言，强化学习框架是唯一可扩展且易于维护的解决方案。

VI. 结论

我们提出了一种用于三维室内光无线通信系统中自适应光束发散角控制的强化学习框架。强化学习智能体在不依赖解析模型或穷举搜索的情况下，学习接收器高度与光束发散之间的高效映射。所得轻量级Q表策略可提供近最优覆盖和实时适应性。结果表明，数据驱动的发散角控制可增强密集型垂直腔面发射激光器网络的可扩展性与响应能力。未来工作将考虑动态用户移动性、遮挡效应和多用户干扰，以及连续状态学习方法和协作式多光束传输。这些扩展有望进一步提升真实室内部署中的鲁棒性与可扩展性。

致谢

本研究是 Project REASON 的一项成果。该项目由英国政府通过未来开放网络研究挑战（FONRC）资助，并由科学、创新与技术部（DSIT）赞助。作者还感谢工程与物理科学研究委员会（EPSRC）通过项目编号 EP/X04047X/1 - EP/Y037243/1 对本研究的支持。

参考文献

[1] Krishnamoorthy, Aravindh, 等，“光无线通信：构建下一代网络之网络”，IEEE Veh. Technol. Mag., vol. 20, no. 2, pp. 20-39, June 2025. [2] A. Ihsan, M. Asif, H. Safi, I. Tavakkolnia and H. Haas，“混合 WiFi/LiFi 网络中的高效服务差异化与能耗管理”，IEEE Trans. Green Commun. Netw., vol. 10, pp. 1335-1351, 2026. [3] R. Michalzik，《垂直腔面发射激光器：基础、技术与应用》。Springer, 2012. [4] M. D. Soltani 等，“太比特室内激光无线通信：面向 6G 的 LiFi 2.0”，IEEE Wirel. Commun., vol. 30, no. 5, pp. 36–43, Oct. 2023. [5] Z. Zeng, M. D. Soltani, M. Safari and H. Haas，“采用新型光束激活机制的 VCSEL 阵列传输系统”，IEEE Trans. Commun., vol. 70, no. 3, pp. 1886-1900, Mar. 2022.[6]H. Kazemi 等，“采用波束聚类的新型太比特波束网格光无线多用户接入网络”，IEEE Trans. Commun., 2025（提前发表）。[7] Y. Liu 等，“具有均匀覆盖的高容量光无线 VCSEL 阵列发射机”，SPIE Free-Space Laser Commun. XXXV, vol. 12413, p. 124130J, Mar. 2023. [8] M. Zohrabi, R. H. Cormack, and J. T. Gopinath，“利用液体透镜实现宽角度非机械光束转向”，Opt. Exp., vol. 24, no. 21, pp. 23798–23809, Sep. 2016. [9] V. Mai and H. Kim，“利用变焦透镜实现自由空间光通信的非机械光束转向与自适应波束控制”，J. Lightw. Technol., vol. 39, no. 24, pp. 7600–7608, Dec. 2021. [10] H. Safi 等，“面向 6G 应用的室内 OWC 系统中 VCSEL 阵列的三维覆盖性能”，IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS), London, UK, May 2025.
[11] C. Liu 等，“可调液体透镜：新兴技术与未来展望”，Laser Photonics Reviews, vol. 17, no. 11, p. 2300274, 2023. [12] S. Boyd, S. P. Boyd, and L. Vandenberghe，《凸优化》。美国纽约州纽约：Cambridge Univ. Press，2004. [13] J. Clifton and E. Laber，“Q学习：理论与应用”，Annu. Rev. Statist. Appl., vol. 7, pp. 279–301, Mar. 2020.