# 强化学习在无线通信信道估计中的应用：当前研究、成果及创新前沿

## 1. 引言：强化学习在信道估计中的兴起

### 1.1 无线系统中信道估计的演进格局

在无线通信系统中，信道估计扮演着至关重要的角色，其精度直接影响数据传输速率、误码率和信号覆盖范围等关键性能指标 1。传统的信道估计算法，如最小二乘 (LS) 和线性最小均方误差 (LMMSE) 算法，在处理复杂的通信环境时，例如存在显著多径效应、信道状态快速变化或强干扰的场景，往往难以精确捕捉信道特性，导致估计精度下降 1。随着下一代无线网络（NextG, 5G, 6G）的演进，其应用场景日益多样化，例如大规模多输入多输出 (Massive MIMO)、毫米波 (mmWave) 通信、智能反射面 (IRS) 以及车联网 (V2X) 等，这些都对信道估计的准确性和自适应性提出了更为严苛的要求 3。

这种技术需求的转变，不仅仅是对现有算法的增量改进，更是对传统方法在日益复杂和动态的下一代网络环境下根本局限性的一种回应。传统方法通常依赖于高斯、瑞利等统计信道模型，这些模型在高度动态、非平稳或充满干扰的真实场景中可能不再适用 1。虽然监督式深度学习能够学习这些复杂映射，但其前提是需要针对每一种可能的信道条件提供大量的标注数据，这在实际中几乎是不可能实现的。

### 1.2 强化学习：自适应信道估计的范式革新

强化学习 (RL) 为自适应信道估计提供了一种全新的范式。它是一种数据驱动的、模型无关（或基于模型）的方法，其中智能体通过与环境的交互来学习制定最优决策 7。这种特性使其特别适用于动态变化的、未知的信道环境，在这些环境中，传统的数学模型可能不准确或难以获取 7。强化学习能够在动态信道环境中持续优化决策，从而提高信道估计的鲁棒性 1。

强化学习在信道估计领域的应用，标志着通信系统向更高级别的自主化和智能化发展，系统能够自我优化物理层功能。智能体通过强化学习进行信道估计决策，例如选择导频符号 12、配置智能反射面 (IRS) 1 或联合优化收发机 10，这代表了比预编程算法或离线训练的监督模型更高程度的自动化。这与AI原生6G网络的广阔前景相契合 4。

### 1.3 强化学习相对于传统及常规机器学习方法的优势

与传统的信道估计算法以及常规的机器学习方法（如监督学习）相比，强化学习展现出多方面的优势：

* **自适应性 (Adaptability):** 强化学习的核心优势在于其能够通过持续学习和调整策略来适应动态变化的环境 1。与监督学习需要大量标记数据集（通常难以获取所有信道条件下的标记数据）不同，强化学习从经验和奖励中学习。
* **模型无关能力 (Model-Free Capability):** 许多强化学习算法不需要信道的精确数学模型，这在信道模型复杂或未知，特别是在端到端 (E2E) 系统或存在非线性特性的场景中尤为重要 7。虽然这种“模型无关”特性为处理未建模动态提供了鲁棒性，但也可能导致比基于模型的方法或具有准确标签的监督学习更高的样本复杂度和确保收敛到全局最优解的挑战。奖励函数的设计变得至关重要，并且确保其引导智能体实现期望的信道估计目标并非易事 7。
* **长期性能优化 (Optimization of Long-Term Performance):** 强化学习的目标是最大化累积奖励，从而制定出长期最优的策略，而非仅仅是眼前的贪婪选择 15。
* **处理非线性问题 (Handling Non-Linearities):** 深度强化学习 (DRL) 将强化学习与深度神经网络相结合，能够通过自动特征提取有效处理复杂信道环境中的非线性和动态波动 1。

## 2. 面向信道估计的核心强化学习方法论

强化学习为信道估计提供了多样化的算法工具箱。这些方法大致可分为基于值函数的方法和基于策略梯度的方法，后者通常与Actor-Critic架构相结合。近年来，深度强化学习（DRL）通过引入深度神经网络作为函数逼近器，极大地扩展了这些方法处理高维复杂问题的能力，这对于现代通信系统中复杂的信道状态表示和连续控制参数空间至关重要 8。

### 2.1 基于值函数的方法：基础与应用

基于值函数的方法旨在学习一个值函数，该函数评估在特定状态下采取某个动作或处于某个状态的“价值”。

* **Q学习 (Q-Learning) 和深度Q网络 (DQN): 原理及信道估计策略**
  + **Q学习:** 是一种模型无关的、离策略的强化学习算法，它学习一个状态-动作值函数（Q值），表示在给定状态下采取某个动作所能获得的预期累积奖励 7。它通常使用一个Q表来存储这些值。在信道估计中，Q学习已被应用于多种场景。例如，在多输入单输出非正交多址接入 (MISO-NOMA) 系统中，基站处的Q学习智能体可以预测下行信道系数，以最大化系统和速率并减少估计损失 15。在此类应用中，状态可以包括当前用户的信道条件，动作是对信道系数的预测或调整，奖励则基于系统和速率 15。另一个应用是在MIMO OFDM系统中进行逐次去噪，Q学习通过识别信道曲率来判断不可靠的信道估计，并应用基于几何的去噪动作来更新信道估计。其状态是量化的信道估计值，动作是选择子载波进行去噪，奖励则反映了均方误差 (MSE) 的降低 17。
  + **深度Q网络 (DQN):** DQN是Q学习的扩展，它使用深度神经网络来近似Q值函数，从而能够处理大规模的连续状态空间 7。DQN通常采用经验回放和目标网络等技术来增强学习的稳定性。在NOMA系统的信道系数估计中，DQN被用于预测信道参数，目标是最大化下行链路的和速率 18。DQN能够克服传统Q学习在Q表存储方面的限制 18。此外，双重DQN (DDQN) 等改进算法还可以缓解DQN中的过高估计偏差 19。
* 性能洞察与固有局限性  
  Q学习及其基于深度学习的变体在特定信道估计任务中取得了显著的性能提升。例如，在误码率 (BER)、中断概率和系统和速率方面，它们均优于传统的MMSE等方法 15。DQN能够有效处理高维状态空间 19。  
  然而，这些方法也存在固有的局限性。Q学习在面对复杂问题时，其Q表的大小可能变得异常庞大，导致存储和计算困难 15。DQN虽然解决了状态空间问题，但可能存在过高估计偏差 7，并且对超参数的选择较为敏感 19。此外，这两类方法通常适用于离散动作空间，这对于信道估计中许多连续参数（如精确的相位偏移）的优化是一个限制 7。计算复杂度和收敛时间也是需要关注的问题 15。

### 2.2 策略梯度与Actor-Critic方法：高级策略

策略梯度方法直接学习一个策略函数，该函数将状态映射到动作。Actor-Critic方法是策略梯度方法的一种流行实现，它包含一个“演员”（Actor）网络负责学习策略，以及一个“评论家”（Critic）网络负责评估策略的好坏。

* **深度确定性策略梯度 (DDPG):** DDPG是一种离策略、模型无关的Actor-Critic算法，专为连续动作空间设计。Actor网络学习确定性策略（将状态映射到特定动作），Critic网络学习Q值函数（评估Actor选择的动作）1。DDPG在信道估计领域的应用包括：利用CNN-GRU模型提取的特征作为状态输入，优化IRS的反射参数，DDPG智能体持续学习并调整IRS配置，以改善信号质量和覆盖范围 1。在端到端通信系统中，DDPG被用于在未知信道条件下联合训练发射机和接收机，无需先验信道状态信息 (CSI)，而是利用接收端的损失作为奖励信号 10。
* **近端策略优化 (PPO):** PPO是一种在策略 (on-policy) Actor-Critic算法，以其学习的稳定性和样本效率而闻名。它使用一种裁剪的替代目标函数来限制策略更新的幅度，从而避免过大的策略变化 7。PPO的应用案例包括：通过考虑信噪比 (SNR) 和先前的RIS相移来优化RIS相移、无人机高度和通信调度 7；在IEEE 802.11ah物联网网络中优化受限接入窗口 (RAW) 参数以提高吞吐量 10；以及在集成传感与通信 (ISAC) 系统中优化信道采样模式 28。
* **双延迟DDPG (TD3):** TD3是DDPG的扩展，旨在解决DDPG的过高估计偏差问题并提高其稳定性。它通过采用裁剪的双Q学习、延迟策略更新和目标策略平滑等技术来实现这些改进 7。TD3的应用包括：考虑信道对局部信息的响应来优化RIS辅助网络中的RIS相移矩阵，以最大化总的可实现有限块长度速率 7；以及在集成的卫星-无人机-RIS网络中联合优化卫星波束赋形、无人机定位和功率分配 30。文献 21 中还提出了一种增强的TD3变体EE-DDPG。
* 性能洞察与固有局限性  
  诸如DDPG和PPO之类的Actor-Critic方法非常适用于连续动作空间，这对于信道估计中的许多参数（例如波束赋形向量、相移）至关重要 1。DDPG在无需信道知识的端到端系统中可以取得显著性能 10。PPO提供稳定高效的学习过程 16。TD3则改进了DDPG的稳定性和过高估计问题 7。  
  然而，这些高级策略也面临挑战。DDPG可能存在过高估计偏差，并且由于包含多个神经网络，其计算复杂度较高 7。PPO作为一种在策略算法，可能样本效率较低。TD3在DDPG的基础上增加了额外的复杂性 7。此外，超参数调整对于这些算法仍然是一个挑战 19。

强化学习算法的选择与信道估计问题的动作空间特性（离散与连续）密切相关。例如，在数据辅助信道估计中选择检测到的数据符号这类问题 12，天然适合Q学习/DQN的离散动作空间。相反，优化连续参数如IRS相移 1、波束赋形向量或功率水平 21，则需要采用DDPG、PPO或TD3等为连续动作空间设计的策略梯度或Actor-Critic方法 7。

DRL算法的演进，例如从DQN到DDQN，从DDPG到TD3，反映了研究者在解决强化学习基本挑战（如过高估计偏差、学习稳定性和样本效率）方面持续的努力，这些对于实现可靠和实用的信道估计至关重要。过高估计偏差（由DDQN 19 和TD3 7 解决）可能导致信道估计中的次优策略，从而降低性能。不稳定的学习过程会使得训练结果不可靠。这些改进不仅仅是学术上的，对于在真实通信系统中部署强化学习（其中一致的性能至关重要）也是必不可少的。PPO等算法的开发 7 也强调了在性能、稳定性和易实现性之间找到平衡。

此外，一个反复出现、至关重要但往往处理微妙的方面是“奖励函数设计”，它极大地影响强化学习在信道估计中的成功。文献中提到了各种奖励，如最小化信道估计误差 12、最大化和速率 15、改善信号质量 1 或使用接收端损失 10。强化学习智能体的有效性直接取决于这个奖励信号与信道估计的真实目标的一致程度。设计不当的奖励可能导致智能体学习到意外的或次优的行为，即使强化学习算法本身很强大。这是研究人员必须仔细调整的关键“旋钮”。

下表1总结了用于信道估计的主要强化学习算法及其关键特性。

**表1：用于信道估计的关键强化学习算法总结**

| **算法** | **类型** | **关键特性 (例如，离散/连续动作空间，在策略/离策略，模型无关)** | **信道估计的典型MDP组件 (状态；动作；奖励)** | **在信道估计中的主要应用领域 (例如，数据辅助CE，IRS优化，E2E学习，去噪)** | **主要参考文献** |
| --- | --- | --- | --- | --- | --- |
| Q-学习 | 基于值 | 离散动作空间，离策略，模型无关 | **状态:** 接收信号，导频数据，历史信道估计，量化信道估计； **动作:** 选择导频/数据符号，调整信道系数值，选择去噪子载波； **奖励:** MSE/NMSE降低，BER改善，和速率最大化 | 数据辅助CE，信道预测，MIMO OFDM去噪 | 7 |
| DQN | 基于值 | 离散动作空间 (通常)，离策略，模型无关，使用DNN近似Q函数 | **状态:** 高维信道特征，接收信号； **动作:** 选择导频/数据符号，预测信道参数； **奖励:** 和速率最大化，SINR增强 | NOMA系统信道估计，大规模状态空间问题 | 7 |
| DDPG | Actor-Critic | 连续动作空间，离策略，模型无关 | **状态:** CSI特征 (例如来自CNN-GRU)，接收端损失信息； **动作:** 调整IRS相位，设置波束赋形向量，调整发射机参数； **奖励:** 信号质量改善，覆盖范围提升，BER/BLER降低 | IRS优化，端到端(E2E)学习，连续参数控制 | 1 |
| PPO | Actor-Critic | 离散或连续动作空间，在策略，模型无关，裁剪替代目标函数 | **状态:** SNR，先前RIS相移，网络状态； **动作:** 调整IRS相移/UAV高度/调度，设置RAW参数，选择信道采样模式； **奖励:** 信息年龄最小化，吞吐量提升，SINR增强 | RIS优化，物联网网络参数优化，ISAC系统 | 7 |
| TD3 | Actor-Critic | 连续动作空间，离策略，模型无关，解决DDPG过高估计问题 | **状态:** 本地CSI，系统状态； **动作:** 优化RIS相移矩阵，调整波束赋形/功率/速率分割； **奖励:** 有限块长度速率最大化，能量效率最大化 | RIS优化，集成网络资源管理 | 7 |

## 3. 强化学习在行动：跨多种通信场景的信道估计

强化学习的自适应和学习能力使其能够应对各种通信场景中独特的信道估计挑战。从智能反射面到高度动态的移动环境，再到新兴的毫米波和物联网应用，强化学习正被积极探索以提升信道估计的性能和鲁棒性。这些应用场景的特性，如信道动态性、硬件限制和系统目标，直接决定了马尔可夫决策过程（MDP）的构建方式（包括状态、动作和奖励的定义）以及相应强化学习算法的选择。

### 3.1 智能反射面 (IRS) / 可重构智能表面 (RIS)：强化学习驱动的优化

* **问题背景:** IRS/RIS技术通过大量低成本的无源反射单元来智能地调控无线传播环境，从而增强接收信号质量或抑制干扰。然而，IRS/RIS的引入使得信道估计变得更加复杂，通常涉及到用户-IRS-基站的级联信道，其维度高且难以直接估计和跟踪 1。传统方法难以适应IRS的动态配置需求 2。
* **强化学习应用:** 深度强化学习，特别是DDPG和PPO等算法，被广泛用于学习IRS/RIS反射单元的最优相移配置 1。强化学习智能体接收当前的状态信息（例如，提取的信道特征、先前的相移配置、接收信号的信噪比等），并输出动作（即对相移进行调整）。奖励函数通常基于接收信号质量、信干噪比 (SINR) 或系统和速率等指标进行设计。
* **典型案例:** 文献 1 提出了一种结合卷积神经网络 (CNN) 和门控循环单元 (GRU) 提取的特征作为状态输入，利用DDPG算法优化IRS配置的智能信道估计模型。该模型在动态信道环境中持续学习和调整IRS配置策略，显著提高了信道估计的准确性和鲁棒性，优于传统的LS和LMMSE方法。文献 7 的综述也指出，DDPG可用于基于CSI优化RIS相移，同时还调研了DQN、TD3、PPO等算法在RIS中的应用 7。
* **主要成果:** 通过强化学习优化IRS/RIS，可以显著增强SINR、提高信道容量，并在动态环境中实现自适应优化，从而提升整体通信系统性能 1。

### 3.2 应对动态性：强化学习在时变和双色散信道中的应用

* **问题背景:** 无线信道本质上是时变的，尤其是在高速移动场景下，信道特性会随时间快速变化。更为复杂的是双色散信道，其特性在时间和频率维度上都发生变化。这些快速变化给传统的信道估计算法带来了巨大挑战，因为这些算法通常假设信道在一段时间内是准静态的 3。
* **强化学习应用:** 强化学习被用于在时变信道中选择已检测的数据符号作为额外的导频，以辅助信道估计，从而更好地适应信道的变化 12。通过将符号选择问题建模为MDP，强化学习算法可以计算出最优的符号选择策略。状态元素细化和滑动窗口等机制有助于跟踪信道的动态变化 12。
* **典型案例:** 文献 12 提出了一种用于时变MIMO系统的RL辅助信道估计器，该估计器通过有效捕捉信道变化，其性能优于传统估计器。该工作使用一阶高斯-马尔可夫过程对信道进行建模，并根据信道变化调整数据符号的权重。文献 14 也针对时变MIMO系统强调了类似方法，并指出由于计算复杂性和信道动态性，推导最优解非常困难。
* **主要成果:** 强化学习方法能够更有效地跟踪信道的时间变化，在动态场景下其性能优于传统估计算法 12。

### 3.3 增强多天线系统：强化学习在MIMO和Massive MIMO信道估计中的应用

* **问题背景:** MIMO系统通过使用多根天线来提高频谱效率和链路可靠性，但这要求准确获取多天线间的CSI，其维度较高。Massive MIMO系统将天线数量扩展到非常大的规模，进一步加剧了导频开销和估计复杂度的挑战 5。此外，Massive MIMO中常采用的混合模数 (HAD) 架构也增加了信道估计的难度，因为接收端只能观测到原始信号的少数线性组合 32。
* **强化学习应用:** 强化学习被用于数据辅助的信道估计，通过有选择地利用已检测的符号作为额外的导频来减少估计误差 13。为了应对计算需求，研究者提出了低复杂度的强化学习算法，例如通过将数据块划分为子块来降低处理复杂度 13。Q学习算法也被用于MIMO OFDM系统中信道估计的逐次去噪 17。在Massive MIMO中，可以将注意力机制与深度学习相结合，使模型能够关注于相关的信道特征，从而提高估计精度 32。
* **典型案例:** 文献 13 提出了一种用于MIMO系统的低复杂度RL信道估计器，该估计器利用检测到的符号，并将问题定义为MDP以最小化MSE，同时引入了折扣因子。文献 17 则在MIMO OFDM系统中基于信道曲率使用Q学习进行逐次去噪。
* **主要成果:** 强化学习的应用显著降低了MSE/NMSE和BER，在导频数量有限的情况下提高了性能，并在时变信道中表现出良好的鲁棒性 13。

### 3.4 超越显式模型：强化学习在端到端 (E2E) 通信系统中的应用

* **问题背景:** 传统的端到端学习系统通常假设已知先验信道信息或存在可微分的信道模型以便进行反向传播，这在许多实际的未知或不可微信道场景中是不切实际的 10。强化学习，特别是DDPG，为这类“黑箱”信道估计问题提供了解决方案。
* **强化学习应用:** 基于DDPG的端到端系统允许在没有先验信道知识的情况下联合训练发射机和接收机的神经网络 10。在这种架构中，发射机充当DDPG智能体，而接收端的损失信息被用作奖励信号。信道估计被神经网络隐式地处理，系统通过学习在未知信道上进行有效通信。
* **典型案例:** 文献 10 和 10 描述了一种DDPG-E2E系统，该系统利用来自接收机的损失反馈来优化发射机，在复杂无线信道（如莱斯、瑞利和3GPP信道）下，尤其是在处理大块长度数据时，其块差错率 (BLER) 性能优于需要信道模型的解决方案。文献 11 对这种DDPG-E2E方法进行了技术解释和批判性分析。
* **主要成果:** 实现了在没有显式信道模型的情况下对发射机和接收机的联合优化，并在复杂无线信道上获得了更好的BLER性能 10。

### 3.5 强化学习在新兴前沿领域的探索：毫米波、物联网、V2X和水下信道估计

* **毫米波 (mmWave):** 毫米波通信面临高路径损耗和对阻塞敏感等挑战。强化学习可用于联合优化毫米波MIMO系统中低分辨率ADC的波束赋形和ADC阈值 34。在RIS辅助的毫米波系统中，可以利用统计性CSI (S-CSI) 进行优化 35。
* **物联网 (IoT) / 反向散射通信 (Backscatter):** 低功耗是物联网应用的关键要求。强化学习算法（如DQN, DDQN, DDPG, PPO）可以自适应地调整传输参数（如调制方式、发射功率），并通过适应SNR和干扰变化来增强反向散射系统中的信道估计 19。例如，PPO已被用于信道跳频或信道黑名单机制，以避免干扰，提高通信可靠性 26。
* **车联网 (V2X):** V2X通信环境具有高度动态性，信道变化快速。深度强化学习（包括单智能体RL (SARL) 和多智能体RL (MARL)）被用于频谱共享和资源分配，这些决策隐式地依赖于对信道状态的理解 9。DRL还可以基于传感状态信息优化波束赋形，从而减少对导频进行信道估计的依赖 39。
* **水下通信 (UWC):** 水下信道环境复杂、时变，并且通常遭受严重的多径效应。强化学习（例如Q学习、DQN）被探索用于基于（可能过时的）CSI进行自适应调制 41，以及优化传输参数 7。虽然直接将RL用于水下信道估计的研究在提供的文献中不突出，但深度学习方法（如神经网络）已被应用 43，RL可以作为这些方法的补充。

在这些新兴领域，虽然强化学习有时是隐式地处理信道估计（例如在端到端系统或资源分配任务中），但其潜力巨大。一个值得注意的趋势是，与MIMO或IRS等领域相比，在某些新兴领域（如V2X或UWC）中，专门用于*直接*信道参数估计的强化学习框架可能尚不成熟。这表明，在这些领域开发更明确的、针对信道估计任务本身的强化学习方法，可能是一个富有前景的创新方向。

强化学习在各种通信场景中的成功应用预示着未来通信系统将高度自适应，即使在知之甚少或快速变化的环境中也能有效运行，从而减少对人工系统调整和大量预部署建模的需求。这对于未来网络（6G及以后）至关重要，因为这些网络预计将远比当前更为复杂和异构 4。

下表2总结了强化学习在不同通信场景中针对信道估计问题的应用。

**表2：强化学习在不同通信场景中用于信道估计的应用**

| **通信场景** | **解决的信道估计问题 (例如，级联信道估计，动态信道跟踪，高维CSI获取，E2E中的隐式CE，波束赋形/ADC优化，干扰避免，恶劣环境下的鲁棒CE)** | **采用的RL方法 (例如，DDPG, PPO, Q学习, DQN, Actor-Critic)** | **主要性能结果/改进 (例如，NMSE/BER降低，和速率/SINR增加，更快的自适应，无CSI操作)** | **该场景下的挑战 (例如，维度，动态性，导频开销，功率限制)** | **主要参考文献** |
| --- | --- | --- | --- | --- | --- |
| IRS/RIS | 级联信道估计，动态相移优化 | DDPG, PPO, DQN, TD3 | NMSE降低，SINR/和速率增加，动态自适应 | 高维度，级联信道，快速变化 | 1 |
| 时变/双色散信道 | 动态信道跟踪，利用数据符号辅助估计 | RL (通用), Q学习 | 有效捕捉信道变化，优于传统估计器 | 快速时变性，计算复杂度 | 3 |
| MIMO/Massive MIMO | 高维CSI获取，导频开销降低，HAD架构下的估计，信道去噪 | RL (低复杂度), Q学习, DL结合注意力机制 | MSE/NMSE/BER降低，有限导频下性能提升，时变信道鲁棒性 | 极高维度，导频污染，计算复杂度 | 13 |
| 端到端 (E2E) 系统 | 无先验CSI的隐式信道估计和联合收发机优化 | DDPG | 无需显式信道模型实现联合优化，复杂信道下BLER改善 | 黑箱信道，梯度传播，大块长度处理 | 10 |
| 毫米波 (mmWave) | 波束赋形与低分辨率ADC联合优化，高路径损耗下的鲁棒估计 | RL (通用), Policy Gradient | 接近穷举搜索的性能，对噪声CSI鲁棒 | 高路径损耗，阻塞敏感，硬件限制 | 34 |
| 物联网 (IoT)/反向散射 | 低功耗下的自适应传输，干扰避免，信道参数自适应 | DQN, DDQN, DDPG, PPO | 适应SNR/干扰变化，提高频谱效率 | 极低功耗限制，简单设备计算能力 | 19 |
| 车联网 (V2X) | 高动态环境下的信道状态理解与资源分配，基于传感的波束赋形 | DRL (SARL, MARL), Actor-Critic | 频谱共享效率提升，减少导频依赖 | 极高动态性，快速信道变化，低延迟要求 | 9 |
| 水下通信 (UWC) | 恶劣多径与时变信道下的自适应调制与参数优化 | Q学习, DQN | 基于过时CSI的自适应调制性能改善 | 极端信道条件，严重多径，带宽限制 | 7 |

## 4. 强化学习信道估计的显著成果与性能基准

将强化学习应用于信道估计领域，已经取得了一系列令人鼓舞的成果，尤其是在性能提升和环境适应性方面，相较于传统方法和其他机器学习技术展现出明显优势。这些成果通常通过关键性能指标（KPIs）来量化，并与现有基准进行比较。

### 4.1 相对于传统估计器 (LS, LMMSE) 和其他机器学习技术的量化改进

研究表明，基于强化学习的框架，特别是深度强化学习，在多种通信场景下，其信道估计误差（例如归一化均方误差NMSE，均方误差MSE）显著低于传统的LS和LMMSE方法 1。例如，一个采用DDPG的IRS辅助信道估计模型，其信道估计误差远小于LS和LMMSE方法 1。在MISO-NOMA系统中，基于Q学习的信道预测不仅在BER方面有所改进，其和速率也优于MMSE甚至基于深度学习的LSTM方法 15。此外，针对MIMO OFDM系统的RL辅助去噪方法，其性能接近理想的LMMSE估计（需要完美的信道统计信息），并显著优于实用的LS估计 17。

这种性能上的超越，尤其是在动态或复杂场景（如IRS、时变信道）中，暗示了强化学习学习非线性映射和适应变化统计数据的能力是其关键的区分因素。LS/LMMSE通常基于线性模型和二阶统计量 1，难以有效捕捉复杂信道现象（如多径、非线性失真、快速衰落）。而DRL凭借神经网络作为函数逼近器，能够直接从数据中学习这些复杂关系 1，从而在如 1 等研究中展现出卓越性能。

### 4.2 关键性能指标 (KPIs) 与已实现的增益

* **误码率 (BER) / 块差错率 (BLER):** 强化学习方法已被证明能够改善BER 13。例如，DDPG-E2E系统在多种信道条件下实现了更优的BLER性能 10。
* **均方误差 (MSE) / 归一化均方误差 (NMSE):** MSE/NMSE的显著降低是强化学习信道估计中普遍报告的成果 1。
* **信干噪比 (SINR) / 和速率 (Sum-Rate):** 在RIS和NOMA等系统中，SINR和和速率的提升通常是强化学习智能体的优化目标，并且也取得了实际的性能增益 7。

### 4.3 在复杂和动态环境中的自适应性与鲁棒性得到验证

强化学习使模型能够在动态信道环境中持续优化决策，从而增强了信道估计的鲁棒性 1。RL辅助的估计器能够有效捕捉时变MIMO系统中的信道变化 12。DDPG-E2E系统在莱斯、瑞利和3GPP等复杂信道模型下表现出良好的鲁棒性 10。一个重要的特性是，强化学习方法可以在没有先验信道知识的情况下运行，从而适应未知的通信环境 7。

文献中频繁提及的“自适应性”和“鲁棒性”等成果，是强化学习学习范式的直接体现——即通过与环境交互并基于反馈（奖励）进行优化。与固定的算法或在静态数据集上训练的监督模型不同，强化学习智能体根据其在当前（且可能变化的）环境中行动的结果来持续改进其策略 1。这种固有的反馈回路使其能够适应初始训练集中未出现的意外信道条件或动态，从而获得了观察到的鲁棒性 1。

强化学习在信道估计领域的这些成就，正为构建更自主和自我优化的无线网络铺平道路，减少了对人工干预和为每种可能场景预定义模型的需求。强化学习在难以获取精确信道模型的场景（例如E2E系统 10、IRS 1）或信道高度动态的场景（12）中取得高性能的能力，指向了能够自我配置和优化的系统。这符合未来智能网络的目标，即能够以最少的人工监督来管理复杂资源并适应不可预测的条件。

然而，需要注意的是，虽然强化学习通常优于传统方法，但报告的增益通常特定于问题公式（MDP、奖励函数）和所使用的DRL架构。将这些增益推广到所有可能的信道估计任务或实现“即插即用”的强化学习解决方案仍然是一个开放的挑战。成功的案例（例如 1）都涉及到针对特定问题（例如IRS相位控制、符号选择、去噪）精心设计的状态、动作和奖励。一个在IRS控制方面表现出色的DDPG智能体 1，若不对其学习框架进行重大重新设计，可能无法直接在例如V2X信道跟踪中获得最佳性能。这突出表明，虽然强化学习*范式*本身功能强大，但其*应用*仍需要特定领域的专业知识。

## 5. 克服障碍：强化学习在信道估计应用中的挑战与局限

尽管强化学习在信道估计领域展现出巨大潜力，但在将其广泛部署于实际通信系统之前，仍有一些关键的挑战和固有的局限性需要克服。这些挑战涉及计算开销、可扩展性、数据效率、鲁棒性以及安全性等多个方面。

### 5.1 计算需求：复杂性与延迟问题

强化学习算法，特别是包含深度神经网络的深度强化学习 (DRL)，在训练和推理阶段都可能具有很高的计算强度 7。计算最优策略可能带来相当大的复杂度和延迟 12。例如，DDPG算法由于包含多个神经网络，其计算开销较大 7，而传统Q学习的Q表在状态空间较大时会变得难以管理 15。这些计算需求构成了实际部署的障碍，尤其是在实时系统或资源受限的终端设备上 11。

### 5.2 大规模部署中的可扩展性与实用性

将强化学习解决方案扩展到具有大量用户、天线（如Massive MIMO）或IRS单元的系统中仍然是一个严峻的挑战 7。随着系统规模的扩大，状态空间和动作空间可能呈指数级增长，使得许多强化学习算法的学习过程变得难以处理 13。计算复杂度、可扩展性和数据效率这几个挑战往往是相互关联的。一个高度复杂的DRL模型可能提供更好的性能，但同时也需要更多的数据和计算能力，从而阻碍其可扩展性。因此，研究人员必须在性能与实际可行性之间进行权衡，特别是在实际系统中。开发“轻量级”DRL的努力 46 正是直接针对这一系列挑战。

### 5.3 MDP构建与奖励工程的关键作用

定义合适的状态、动作，尤其是奖励函数，对于强化学习的成功至关重要，但这并非易事 7。奖励函数必须准确地反映信道估计的目标。不恰当的MDP设计可能导致学习到次优策略或收敛缓慢。DRL模型的“黑箱”特性虽然有利于处理未知环境，但也增加了奖励工程和确保泛化能力的难度。由于缺乏对DRL智能体决策过程的清晰分析模型，设计完美的奖励函数变得更加困难 7。同时，也更难预测智能体在全新的、分布之外的场景中的行为，从而影响其泛化能力 11。这与基于模型的优化形成了对比，后者的行为更具可预测性。

### 5.4 数据效率、样本复杂度与泛化能力

强化学习算法可能存在样本效率低的问题，即需要与环境进行大量的交互才能学习到有效的策略 11。确保模型能够泛化到训练阶段未曾遇到的信道条件或与训练设置不同的环境中至关重要 3。此外，DRL模型的离线训练通常需要大规模数据集 46，这在某些场景下难以获取。

### 5.5 安全性考量：对抗性操纵的脆弱性

基于深度学习的信道估计模型，包括那些使用强化学习的模型，容易受到对抗性机器学习攻击的影响 4。攻击者可能精心制作扰动，以错误的信道估计结果欺骗接收机，从而构成严重的安全风险 5。随着机器学习/强化学习模型日益成为通信系统的组成部分，安全性正成为一个日益受到重视的挑战。如果作为物理层基本过程的信道估计可以通过对强化学习智能体的攻击而被恶意操纵，那么整个通信链路都可能受到损害。这已超出了性能优化的范畴，进入了系统可信度和韧性的领域，这也是下一代网络的关键主题 4。

克服这些挑战对于将基于强化学习的信道估计从学术研究推向广泛的实际部署至关重要。当前的局限性界定了该领域成熟所需的主要研究方向。在这些领域（例如，高效算法、用于提高数据效率的迁移/元学习、用于安全性的鲁棒强化学习）的解决方案，将使强化学习成为通信工具箱中的标准工具。

## 6. 未来之路：强化学习在信道估计领域的创新前沿与研究方向

尽管强化学习在信道估计领域已取得显著进展，但仍有广阔的创新空间和诸多值得探索的研究方向。未来的研究趋势主要集中在提升强化学习的实用性、自适应性、效率、可信度以及与其他通信模块的集成度，以应对下一代无线通信系统日益增长的复杂性和多样性需求。

### 6.1 提升自适应性与效率的高级学习范式

* **元强化学习 (Meta-RL): 实现对新信道环境的快速适应**
  + **概念:** 元强化学习，或称“学会学习”，旨在训练智能体使其能够利用少量数据快速适应新的任务，例如不同的信道环境、用户移动模式或新的服务质量 (QoS) 需求 48。这对于处理非平稳信道至关重要。
  + **应用:** 可以训练模型找到最优的初始参数，使其能够针对新的信道条件快速收敛 48。模型无关元学习 (MAML) 是一种突出的算法 48。例如，文献 48 将MAML与DQN结合用于无人机通信中的飞行轨迹和调度策略优化，以最小化年龄信息 (AoI) 和传输功率的组合。文献 49 介绍了GMA，一种基于上下文的元强化学习方法，结合专家混合 (MoE) 来改进表示学习，用于MAC协议的快速适应。
  + **潜力:** 减少新环境下的训练开销，提高泛化能力 48。元强化学习还可以调整强化学习解决方案的超参数以实现快速适应 48。
  + **相关研究:** 48 (MAML与DQN结合用于无人机AoI/功率优化)49 (GMA，基于上下文的元RL与MoE结合用于MAC协议自适应)54 (元学习用于下行信道预测，用少量数据适应新环境)52 (MAML用于时变OFDM信道估计，少量样本即可快速适应)50 (离线元RL)。
* **迁移学习 (Transfer Learning, TL): 跨通信任务与场景的知识迁移**
  + **概念:** 利用从源任务/环境中获得的知识来改进目标任务/环境中的学习效果，从而减少对目标领域大量数据的依赖 8。
  + **应用:** 可以在来自各种环境的数据上预训练模型，然后使用来自新环境的有限数据进行微调 54。例如，文献 54 提出了直接迁移算法，在FDD大规模MIMO系统中，利用先前所有环境的数据训练模型，然后在新环境中使用少量标记数据进行微调，以从上行CSI预测下行CSI。迁移学习还可以用于初始化异构联邦学习 (HFL) 中的本地模型参数 55。
  + **潜力:** 加速学习过程，提高数据稀缺场景下的性能，更有效地适应新的系统配置或信道统计特性。
  + **相关研究:** 54 (直接迁移和元学习用于FDD大规模MIMO信道预测)55 (TL用于增强RIS辅助的无蜂窝MIMO系统中信道估计的HFL初始化)。
* **联邦与分布式强化学习: 迈向隐私保护和可扩展的信道估计**
  + **概念:** 在多个设备（用户、基站）之间协同训练强化学习模型，而无需共享原始本地数据，从而增强隐私保护 4。分布式强化学习允许智能体基于本地信息学习策略，同时为全局目标做出贡献。
  + **应用:** 联邦学习可用于RIS辅助的无蜂窝MIMO系统中的信道估计，其中用户形成联盟，并使用分布式DRL（例如基于Qmix的方法）进行决策 55。
  + **潜力:** 提高可扩展性，减少集中式训练的通信开销，保护用户数据隐私，并支持设备端学习。
  + **相关研究:** 55 (HFL结合联盟形成和DRL进行信道估计)7 (提及FL/SL作为RIS的未来研究方向)4 (基于分布式学习的信道估计模型用于后5G系统)。 对高度动态和多样化的6G场景（例如，变化的服务、用户密度、信道条件）的快速适应需求，是推动元强化学习和迁移学习研究的强大动力。传统的强化学习甚至深度强化学习在针对某一特定场景训练后，当条件发生显著变化时，其性能可能会大幅下降。从头开始重新训练通常过于耗时或数据密集 49。元强化学习 48 和迁移学习 54 提供了“学会适应”或“迁移知识”的机制，这对于未来网络所设想的灵活性至关重要。

### 6.2 设计轻量级与计算高效的DRL算法

* **问题:** DRL的高计算成本阻碍了其在资源受限设备上的部署 11。
* **方法:**
  + 开发更简单的神经网络架构，例如轻量级CNN与Transformer的结合 46。
  + 探索无限宽度神经网络/神经正切核 (CNTK) 等可能不需要大量数据集或广泛训练的方法 46。这些方法通过利用无限宽网络的训练动态可以用封闭形式方程表示的特性，仅使用导频位置的已知值来估算缺失的信道响应。
  + 进行算法级优化以降低Q学习或Actor-Critic方法的复杂度，例如文献 13 中提出的子块划分策略。
  + 应用剪枝、量化或知识蒸馏等模型压缩技术于DRL模型。
* **潜力:** 实现设备端的强化学习信道估计，降低延迟，使强化学习在实时应用中更具实用性。
* **相关研究:** 46 (CNTK用于信道估计，减少数据和计算需求)13 (通过子块划分实现低复杂度RL)16 (PPO用于物联网RAW优化，暗示对效率的关注)。文献 20 提及DQN可用于处理大规模状态空间，但仍需关注计算效率。 对轻量级DRL 46 和联邦/分布式RL 4 的推动是相辅相成的。轻量级模型更容易部署在边缘设备上，这是许多联邦学习场景的先决条件。联邦学习涉及在本地设备上训练模型 55。如果这些设备资源受限（在物联网、V2X中很常见），则DRL模型必须计算成本低廉。因此，轻量级DRL的进步直接促进了更有效和广泛的联邦学习在信道估计中的应用。

### 6.3 构建鲁棒的RL智能体以应对非平稳、高干扰和对抗性环境

* **问题:** 真实的无线信道通常是非平稳的，会受到不可预测的干扰，并且可能面临对抗性攻击 1。
* **方法:**
  + 采用本身就为动态环境设计的强化学习算法，例如文献 1 中用于IRS在复杂条件下进行调整的DDPG，以及文献 12 中用于时变信道的RL方法。
  + 开发具有明确鲁棒性目标或约束的强化学习智能体。
  + 将强化学习与干扰抑制技术相结合，例如文献 58 中通过DRL优化的速率分割多址接入 (RSMA) 来减轻干扰。
  + 研究安全强化学习方法，以防御针对信道估计器的对抗性攻击 4。这可能涉及对抗性训练或鲁棒的奖励函数设计。
* **潜力:** 即使在具有挑战性和敌对性的无线环境中，也能确保可靠的通信性能。
* **相关研究:** 1 (DDPG用于动态IRS调整)58 (DRL用于RSMA预编码以减轻干扰)4 (安全担忧及对鲁棒模型的需求)。

### 6.4 迈向可解释人工智能 (XAI)：增强RL驱动的信道估计器的可解释性

* **问题:** DRL模型，特别是深度神经网络，通常被视为“黑箱”，使其决策过程难以理解 11。这种缺乏可解释性会阻碍信任和调试。
* **方法:**
  + 开发专为通信系统中强化学习智能体定制的XAI技术。
  + 可视化学习到的策略或值函数，例如文献 32 中用于Massive MIMO的注意力图。
  + 设计本质上更具可解释性的强化学习架构。
  + 利用学习到的世界模型来理解智能体的预测，如文献 59 中基于智能体内部状态或世界模型预测未来动作。
* **潜力:** 增强对基于强化学习的系统的信任度，便于调试和改进，并提供对学习到的信道特性的洞察。
* **相关研究:** 32 (可视化注意力图以增强可解释性)11 (缺乏可解释性被认为是DDPG-E2E的局限性)59 (基于内部状态或世界模型预测智能体未来行为以增进理解)。

### 6.5 整体优化：用于联合信道估计和跨层设计的强化学习

* **概念:** 超越孤立地优化信道估计，强化学习可用于联合优化信道估计与其他通信任务，如资源分配、波束赋形、功率控制，甚至更高层协议。
* **应用:**
  + 联合训练导频信号和信道估计器 32。
  + 联合优化波束赋形、功率分配和RIS配置 7。
  + DDPG-E2E联合优化发射机和接收机，隐式处理信道估计 10。
  + 基于信息价值 (VoI) 优化V2X中的通信时机、内容和方式，集成控制与通信决策 36。文献 29 提出联合优化物联网设备卸载、STAR-RIS的幅度和相移系数以及功率控制。文献 30 则研究了卫星波束赋形、无人机定位、功率分配和速率分割的联合优化。
* **潜力:** 实现通过优化单个组件无法达成的系统级性能增益，催生更集成和智能的通信系统。 利用强化学习进行联合优化和端到端学习的趋势 10，预示着一种潜在的范式转变，即通信系统中传统的模块化设计（分离的信道估计、均衡、解码等）可能会被更整体的、人工智能驱动的设计所取代，这些设计能够隐式地、集成地学习执行这些任务。DDPG-E2E 10 已经通过在未知信道上学习通信而无需显式估计模块来证明了这一点。扩展这一理念，强化学习可以动态地协调各种物理层和MAC层功能，从而实现真正的“智能”收发机。这与当前的工程实践有显著不同。
* **相关研究:** 10 (DDPG-E2E)32 (导频与估计器联合训练)34 (波束赋形与ADC阈值联合优化)36 (CAV的通信与控制联合优化)。

然而，追求更强大、更复杂的DRL模型（例如用于联合优化或处理非常复杂状态的模型）与对轻量级、可解释模型的需求之间存在潜在的张力。虽然复杂的DRL架构可能在特定的复杂任务中实现更高的峰值性能（例如，30 使用TRPO、TD3、A3C进行多参数优化），但它们可能更难部署 11、调试和信任 11。研究界将需要找到平衡点或开发能够同时提供能力和实用性/可解释性的混合方法。

下表3概述了强化学习在信道估计领域的创新前沿和未来研究方向。

**表3：强化学习在信道估计领域的创新前沿与未来研究方向**

| **创新领域** | **待探索的特定RL技术/概念 (例如，MAML, Reptile, 离线元RL；参数迁移，实例迁移；带RL的联邦平均，多智能体RL；CNTK, 剪枝/量化DRL；对抗性RL, 鲁棒MDP；基于注意力的XAI, 代理模型；多目标RL, 分层RL)** | **对信道估计的潜在影响 (例如，快速适应未知信道，减少数据需求，隐私与可扩展性，设备端部署，抗干扰/攻击的韧性，可信AI，系统级性能增益)** | **需解决的关键挑战 (例如，元任务定义，负迁移，FL中的通信开销，轻量级模型性能保持，定义合适的威胁模型，解释性与性能的平衡，联合优化的复杂性)** | **相关参考文献** |
| --- | --- | --- | --- | --- |
| 元强化学习 (Meta-RL) | MAML, Reptile, 离线元RL, 基于上下文的元RL与MoE | 快速适应未知信道，减少新环境的训练开销，提高泛化能力 | 元任务定义，少量样本学习的稳定性，计算复杂度 | 48 |
| 迁移学习 (TL) | 参数迁移，实例迁移，特征表示迁移，与DRL结合 (DTL) | 减少对目标域大量数据的依赖，加速学习，提高数据稀缺场景性能 | 负迁移风险，源任务与目标任务的关联性度量，知识迁移的有效性 | 8 |
| 联邦/分布式RL | 带RL的联邦平均，多智能体RL (MARL)，Qmix，异构FL (HFL) | 隐私保护，可扩展性，减少通信开销，支持设备端学习 | 通信开销，异构性管理，收敛性保证，激励机制设计 | 4 |
| 轻量级DRL | 简化NN架构 (轻量级CNN/Transformer)，无限宽度NN (CNTK)，算法级优化 (例如子块划分)，模型压缩 (剪枝，量化，知识蒸馏) | 设备端部署，降低延迟，提高实时应用可行性 | 在降低复杂度的同时保持高性能，模型压缩与RL的结合 | 13 |
| 鲁棒RL (应对恶劣环境) | 针对动态环境的RL算法 (DDPG)，鲁棒性目标/约束，与干扰抑制技术结合 (RSMA)，安全RL (对抗性训练，鲁棒奖励设计) | 在非平稳、高干扰、对抗性设置下保证可靠通信 | 定义合适的威胁模型，鲁棒性与性能的权衡，对抗样本的生成与防御 | 1 |
| 可解释RL (XAI for CE) | 针对RL的XAI技术，策略/值函数可视化 (例如注意力图)，固有可解释架构，学习世界模型 | 增强系统信任度，便于调试与改进，提供对学习到的信道特性的洞察 | 解释性与性能的平衡，复杂RL模型的解释难度，缺乏标准化评估指标 | 11 |
| 联合CE与跨层优化 | 联合导频与估计器训练，联合波束赋形/功率/RIS配置，E2E学习，多目标RL，分层RL | 系统级性能增益，更集成和智能的通信系统 | 联合优化问题的复杂性，多目标冲突与权衡，跨层信息交互 | 7 |

## 7. 结论：强化学习对信道估计的变革性影响

强化学习的引入，正在深刻地改变无线通信领域中信道估计的理论与实践。通过赋予通信系统前所未有的自学习和自适应能力，强化学习不仅克服了传统估计算法在面对日益复杂的无线环境时的诸多局限，更为未来智能通信系统的构建奠定了坚实基础。

回顾强化学习在信道估计中的贡献，其核心价值在于能够有效应对动态变化、模型未知或模型复杂的信道场景。无论是通过Q学习及其变体在MIMO和NOMA系统中优化符号选择和资源分配 15，还是利用DDPG、PPO等先进的Actor-Critic方法在IRS/RIS配置 1 和端到端通信 10 中实现智能决策，强化学习都展现出提升估计精度、增强系统鲁棒性以及优化整体通信性能的巨大潜力。其在时变信道跟踪 12、Massive MIMO高效估计 13 以及毫米波、物联网、车联网等新兴领域的初步探索 9，均预示着一个由数据驱动的、更加智能化的物理层正在形成。

展望未来，如元强化学习、迁移学习、联邦学习等高级学习范式的应用，将进一步提升强化学习在信道估计中的适应性和效率，使其能够快速响应新环境、利用先验知识并保护用户隐私 48。轻量级DRL算法的研发将推动其在资源受限设备上的部署，而鲁棒强化学习和可解释人工智能的研究则致力于构建更值得信赖和易于理解的智能体 5。尤为重要的是，强化学习驱动的联合信道估计与跨层优化，例如端到端学习和通信与控制的协同，预示着通信系统设计理念可能从传统的模块化、分离式优化，转向更加整体化、一体化的人工智能原生设计 10。

尽管计算复杂度、样本效率、奖励设计以及安全性等方面仍存在挑战，但持续的研究和创新正在不断突破这些瓶颈。强化学习的终极目标是助力构建能够自主感知、理解、适应并优化其运行状态的无线网络。随着这些创新方向的不断深入，强化学习有望在6G及更远未来的通信系统中扮演核心角色，赋能真正意义上的智能、自主和高韧性的无线连接，从而彻底改变我们设计和运营通信网络的方式。

#### 引用的著作

1. Research on channel estimation based on joint perception and deep enhancement learning in complex communication scenarios - PeerJ, 访问时间为 五月 14, 2025， <https://peerj.com/articles/cs-2852.pdf>
2. Research on channel estimation based on joint perception and deep enhancement learning in complex communication scenarios - PeerJ, 访问时间为 五月 14, 2025， <https://peerj.com/articles/cs-2852/>
3. [2206.02165] A Survey on Deep Learning based Channel Estimation in Doubly Dispersive Environments - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/abs/2206.02165>
4. CMC | Free Full-Text | Secure Channel Estimation Using Norm Estimation Model for 5G Next Generation Wireless Networks, 访问时间为 五月 14, 2025， <https://www.techscience.com/cmc/v82n1/59225/html>
5. Undermining Deep Learning Based Channel Estimation via Adversarial Wireless Signal Fabrication, 访问时间为 五月 14, 2025， <https://par.nsf.gov/servlets/purl/10339961>
6. Review on Channel Estimation for Reconfigurable Intelligent Surface Assisted Wireless Communication System - MDPI, 访问时间为 五月 14, 2025， <https://www.mdpi.com/2227-7390/11/14/3235>
7. A Survey on Reinforcement Learning for Reconfigurable Intelligent ..., 访问时间为 五月 14, 2025， <https://www.mdpi.com/1424-8220/23/5/2554>
8. deep reinforcement learning: an overview - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/pdf/1701.07274>
9. Spectrum Sharing using Deep Reinforcement Learning in Vehicular Networks - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/pdf/2410.12521>
10. DDPG-E2E: A Novel Policy Gradient Approach for End-to-End Communication Systems, 访问时间为 五月 14, 2025， <https://arxiv.org/html/2404.06257v2>
11. DDPG-E2E: A Novel Policy Gradient Approach for End-to-End Communication Systems, 访问时间为 五月 14, 2025， <https://www.aimodels.fyi/papers/arxiv/ddpg-e2e-novel-policy-gradient-approach-end>
12. Reinforcement Learning-Aided Channel Estimator in Time-Varying MIMO Systems - PMC, 访问时间为 五月 14, 2025， <https://pmc.ncbi.nlm.nih.gov/articles/PMC10304914/>
13. A Low-Complexity Algorithm for a Reinforcement Learning-Based Channel Estimator for MIMO Systems - PMC, 访问时间为 五月 14, 2025， <https://pmc.ncbi.nlm.nih.gov/articles/PMC9229451/>
14. Reinforcement Learning-Aided Channel Estimator in Time-Varying MIMO Systems - MDPI, 访问时间为 五月 14, 2025， <https://www.mdpi.com/1424-8220/23/12/5689>
15. A Study on the Impact of Integrating Reinforcement Learning for ..., 访问时间为 五月 14, 2025， <https://pmc.ncbi.nlm.nih.gov/articles/PMC9921540/>
16. Deep Reinforcement Learning for Optimizing Restricted Access Window in IEEE 802.11ah MAC Layer - MDPI, 访问时间为 五月 14, 2025， <https://www.mdpi.com/1424-8220/24/10/3031>
17. arxiv.org, 访问时间为 五月 14, 2025， <https://arxiv.org/abs/2101.10300>
18. Simplified Deep Reinforcement Learning Approach for Channel Prediction in Power Domain NOMA System - MDPI, 访问时间为 五月 14, 2025， <https://www.mdpi.com/1424-8220/23/21/9010>
19. arxiv.org, 访问时间为 五月 14, 2025， <https://arxiv.org/pdf/2309.12507>
20. Deep Reinforcement Learning for Dynamic Multichannel Access in Wireless Networks, 访问时间为 五月 14, 2025， <https://www.researchgate.net/publication/323302666_Deep_Reinforcement_Learning_for_Dynamic_Multichannel_Access_in_Wireless_Networks>
21. Energy-Efficient Irregular RIS-aided UAV-Assisted Optimization: A Deep Reinforcement Learning Approach - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/html/2504.15031v1>
22. DDPG-E2E: A Novel Policy Gradient Approach for End-to-End Communication Systems, 访问时间为 五月 14, 2025， <https://www.researchgate.net/publication/385231114_DDPG-E2E_A_Novel_Policy_Gradient_Approach_for_End-to-End_Communication_Systems>
23. Nguyen HUYNH | Lecturer | PhD | University of Liverpool, Liverpool | UoL | Department of Electrical Engineering and Electronics | Research profile - ResearchGate, 访问时间为 五月 14, 2025， <https://www.researchgate.net/profile/Nguyen-Huynh-4>
24. 访问时间为 一月 1, 1970， httpshttps://arxiv.org/html/2404.06257v2
25. 访问时间为 一月 1, 1970， <https://arxiv.org/pdf/2404.06257>
26. A PPO-based Channel Hopping Sequence Framework for Time Slot Channel Hopping - ResearchGate, 访问时间为 五月 14, 2025， <https://www.researchgate.net/publication/384851434_A_PPO-based_Channel_Hopping_Sequence_Framework_for_Time_Slot_Channel_Hopping>
27. PPO Simply Explained for Beginners with Detailed Walkthrough (Proximal Policy Optimization) - YouTube, 访问时间为 五月 14, 2025， <https://www.youtube.com/watch?v=5VHLd9eCZ-w>
28. Using Deep Reinforcement Learning to Enhance Channel Sampling Patterns in Integrated Sensing and Communication - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/html/2412.03157v1>
29. Proximal Policy Optimization for Energy-Efficient MEC Systems with STAR-RIS Assistance, 访问时间为 五月 14, 2025， <https://icoin.org/media?key=site/icoin2024/abs/A-3-3.pdf>
30. Energy Efficient RSMA-Based LEO Satellite Communications Assisted by UAV-Mounted BD-Active RIS: A DRL Approach - arXiv, 访问时间为 五月 14, 2025， <https://www.arxiv.org/pdf/2505.04148>
31. Deep Residual Learning for Channel Estimation in Intelligent Reflecting Surface-Assisted Multi-User Communications - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/pdf/2009.01423>
32. An Attention-Aided Deep Learning Framework for Massive MIMO Channel Estimation - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/pdf/2108.09430>
33. Deep Learning-Based Channel Estimation - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/pdf/1810.05893>
34. www.arxiv.org, 访问时间为 五月 14, 2025， <https://www.arxiv.org/pdf/2504.18957>
35. Channel Estimation in RIS-Enabled mmWave Wireless Systems: A Variational Inference Approach - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/pdf/2308.13616>
36. Learning Value of Information towards Joint Communication and Control in 6G V2X - arXiv, 访问时间为 五月 14, 2025， <http://www.arxiv.org/pdf/2505.06978>
37. Learning Value of Information towards Joint Communication and Control in 6G V2X - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/pdf/2505.06978>
38. Reinforcement Learning-Based Resource Allocation Scheme of NR-V2X Sidelink for Joint Communication and Sensing - MDPI, 访问时间为 五月 14, 2025， <https://www.mdpi.com/1424-8220/25/2/302>
39. Energy-Efficient and Intelligent ISAC in V2X Networks with Spiking Neural Networks-Driven DRL - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/html/2501.01038v1>
40. Energy-Efficient and Intelligent ISAC in V2X Networks with Spiking Neural Networks-Driven DRL - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/pdf/2501.01038>
41. Deep Reinforcement Learning-Based Adaptive Modulation for Underwater Acoustic Communication with Outdated Channel State Information - MDPI, 访问时间为 五月 14, 2025， <https://www.mdpi.com/2072-4292/14/16/3947>
42. 访问时间为 一月 1, 1970， <https://www.mdpi.com/2072-4292/14/16/3947/pdf>
43. Robust Channel Estimation for Optical Wireless Communications Using Neural Network, 访问时间为 五月 14, 2025， <https://arxiv.org/html/2504.02134v1>
44. Robust Channel Estimation for Optical Wireless Communications Using Neural Network - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/pdf/2504.02134>
45. Robust Underwater Acoustic Channel Estimation Method Based on Bias-Free Convolutional Neural Network - ResearchGate, 访问时间为 五月 14, 2025， <https://www.researchgate.net/publication/377280019_Robust_Underwater_Acoustic_Channel_Estimation_Method_Based_on_Bias-Free_Convolutional_Neural_Network>
46. Channel Estimation by Infinite Width Convolutional Networks M. Mallik and G. Villemaud are with INSA Lyon, Inria, CITI, UR3720, France. (email:mohammed.mallik@insa-lyon.fr, guillaume.villemaud@insa-lyon.fr).This work was supported by a French government grant managed by the - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/html/2504.08660v1>
47. channel estimation by infinite width convolutional networks - arXiv, 访问时间为 五月 14, 2025， [https://arxiv.org/pdf/2504.08660?](https://arxiv.org/pdf/2504.08660)
48. Age and Power Minimization via Meta-Deep Reinforcement Learning in UAV Networks - arXiv, 访问时间为 五月 14, 2025， [https://arxiv.org/pdf/2501.14603?](https://arxiv.org/pdf/2501.14603)
49. Meta-Reinforcement Learning With Mixture of Experts for Generalizable Multi Access in Heterogeneous Wireless Networks - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/html/2412.03850v1>
50. Offline Meta-Reinforcement Learning with Evolving Gradient Agreement - Unpaywall, 访问时间为 五月 14, 2025， <https://unpaywall.org/10.1109%2FIROS58592.2024.10802167>
51. Reinforcement Learning in Process Industries: Review and Perspective, 访问时间为 五月 14, 2025， <https://www.ieee-jas.net/article/doi/10.1109/JAS.2024.124227>
52. Meta-learning based time-varying channel estimation method, 访问时间为 五月 14, 2025， <https://www.sys-ele.com/EN/10.12305/j.issn.1001-506X.2023.06.32>
53. From Theory to Practice: Implementing Meta-Learning in 6G Wireless Infrastructure, 访问时间为 五月 14, 2025， <https://www.researchgate.net/publication/386523661_From_Theory_to_Practice_Implementing_Meta-Learning_in_6G_Wireless_Infrastructure>
54. arxiv.org, 访问时间为 五月 14, 2025， <http://arxiv.org/pdf/1912.12265>
55. Coalition Formation for Heterogeneous Federated Learning Enabled Channel Estimation in RIS-assisted Cell-free MIMO - arXiv, 访问时间为 五月 14, 2025， <https://arxiv.org/html/2502.05538v1>
56. Meta-learning Approaches for Smart Antenna Systems in 5G Networks Using Reinforcement Learning and Artificial Intelligence - ResearchGate, 访问时间为 五月 14, 2025， <https://www.researchgate.net/publication/389639231_Meta-learning_Approaches_for_Smart_Antenna_Systems_in_5G_Networks_Using_Reinforcement_Learning_and_Artificial_Intelligence>
57. UAV-Aided Cellular Communications with Deep Reinforcement Learning Against Jamming | Request PDF - ResearchGate, 访问时间为 五月 14, 2025， <https://www.researchgate.net/publication/343730113_UAV-Aided_Cellular_Communications_with_Deep_Reinforcement_Learning_Against_Jamming>
58. rate splitting for interference channels with deep reinforcement learning - Middle East Technical University, 访问时间为 五月 14, 2025， <https://avesis.metu.edu.tr/dosya?id=7d2388a3-edfa-4214-a2cc-363324c4432d>
59. Predicting Future Actions of Reinforcement Learning Agents - NIPS papers, 访问时间为 五月 14, 2025， <https://proceedings.neurips.cc/paper_files/paper/2024/file/5aea56eefab60e06f35016478e21aae6-Paper-Conference.pdf>