# 研究背景

随着无人机技术的快速发展，其在多个方面展现了显著优势，包括无人员伤亡、制造成本低、使用方便、用途广泛、机动性强等特点。无人机凭借这些优势，在军事和民用领域展现出巨大的应用潜力。例如，在森林防护和临时战场等场景中，无人机网络因其灵活的移动能力而能够满足多变的通信需求。然而，这些场景中的快速移动特性会导致网络覆盖范围的变化，使得传统蜂窝无线网络难以提供稳定可靠的通信服务。为应对这一挑战，无人机之间通过自主组网，构建协同通信网络来提升无线通信的传输质量。无人机自组网（Ad Hoc Network）可以借助单跳或多跳通信来提供可靠的实时服务，具有更强的可拓展性和生存能力。无人机群组成的通信网络可以发挥多方面的作用，无人机网络中的节点可作为飞行的用户、中继和基站，通过有效协调各无人机可构建复杂网络，多个具有不同功能的无人机可以集成为功能强大的系统，辅助地面设施提供全方位的通信覆盖。与传统无线通信相比，无人机通信网络具有节点高速移动、能量受限、信道可变的特点，快速飞行的无人机使得网络拓扑高度动态和通信链路质量短时间内变动。

无人机网络具有快速部署、低成本和易于控制的优点，多架无人机可协同共享网络资源，从而完成复杂的通信任务。当部分无人机因故障或恶意干扰而导致通信质量下降时，其他无人机能够及时协助受影响的节点继续传输信息，显著提高了网络的传输可靠性，有效解决了单一无人机抗毁性弱和覆盖范围有限的问题。同时，通过无人机群的协作，可以扩大无线网络的覆盖范围。当前，无人机广泛应用于军事领域的空中侦察、战场监视、目标定位、边境巡逻等任务，以及民用领域的航空摄影、地球物理勘探、灾情监测、交通巡逻等场景。

同时无人机网络的开放特性使它容易受到各种物理层的干扰攻击包括欺骗和压制等，其中各类干扰攻击都会对无人机网络通信安全造成严重威胁。作为一种恶意攻击，干扰器通过发射足够功率的无线电信号来中断正常通信，导致在接收器端发生冲突。一旦无人机受到攻击，网络通信质量将下降，甚至无法满足当前需求导致任务中断，当无人机遭受严重干扰时，它们不再能够与其他无人机以及控制站点建立连接，因此提出有效的无人机网络抗干扰方法有重要意义。

强化学习（Reinforcement Learning, RL）是一种半监督的机器学习方法，其通过奖励和惩罚机制使智能体在与环境的交互中不断改进行为策略，以实现特定任务目标。强化学习最初受到动物行为和控制理论的启发，随着Q-learning等算法的提出，强化学习逐渐形成了系统的理论框架。进入21世纪后，深度学习技术的快速发展带动了深度强化学习（Deep Reinforcement Learning, DRL）的突破，解决了基于表学习的有限状态空间的限制，其应用前景大大拓展，特别是在机器人控制、游戏、自动驾驶等复杂任务中取得显著成就。强化学习的特点包括以下几个方面：试错学习：智能体通过与环境的不断交互、试错，逐步学习最佳的行为策略；延迟奖励：强化学习可以处理延迟奖励的问题，即行为的结果和反馈之间可以有较长的时间延迟；策略优化：通过不断调整策略，以使智能体的长期回报最大化。现在强化学习在自动化控制、自然语言处理、推荐系统等多个领域表现出广阔的应用潜力。

将强化学习用于解决无人集群抗干扰问题的研究也随着强化学习的发展而增加。利用强化学习，不仅可以用于优化传统的抗干扰技术方案，也可以设计不同的抗干扰方案，提升抗干扰效果。本课题研究的方向就是利用强化学习算法设计抗干扰方法，从而提升无人机集群的抗干扰性能，满足不同任务场景下无人集群的通信性能指标要求。下面将具体介绍无人集群抗干扰以及基于强化学习的无人集群抗干扰方法的研究进展与现状。

# 国内外研究进展与现状

## 无人集群抗干扰研究进展与现状

### 干扰方式分类

由于无线电的共享介质传播，无线传输很容易受到干扰攻击，干扰器可通过发射不需要的无线电信号来破坏合法传输。

根据实际产生干扰信号的类型，无人集群通信中的主要有两种干扰方式，一种是压制式干扰，另外一种就是欺骗式干扰[1]。压制干扰是指干扰器持续发射的干扰信号功率大于无人机数据链信号功率的一种人为通信干扰,使数据链中的通信节点不能正确接收射频信号,导致通信链路中断[2]。根据干扰信号的形式,压制干扰可分为单频干扰、窄带干扰和宽带干扰三种类型。由于压制干扰功率淹没了有用信号或阻塞了射频前端,其调制信息无关紧要。欺骗干扰通过发送与无人机数据链信号结构相似的欺骗信号,使其无法检测出诱导接收机跟踪捕获的欺骗信号,从而达到降低其抗干扰能力的目的,而且可以采用与数据链信号近似的功率,避免功率过大而被检测到并降低成本。在实际常见于军事领域种，接收机往往面临着高动态、弱信号、强干扰或信号遮挡等复杂场景下。

而根据干扰器生成干扰信号的策略来看，干扰器又分为持续干扰、随机干扰、扫频干扰、自适应干扰、智能干扰和反应式干扰六种[3]。具体介绍如下：

（1）持续干扰：干扰器在信道上连续发送干扰信号，会降低合法接收机的接收信号质量以及使无线信道总是忙阻止发射机对信道的访问。

（2）随机干扰：在每个时隙随机从信道集合中选择一组信道来感知，并基于感知结果在空闲信道上发射干扰信号，干扰决策与感知历史无关。

（3）扫频干扰:是将一个带宽较窄的信号在某段时间内与整个频段上扫描, 同宽带或部分频段噪声干扰类似。在扫频过程中的任意时刻，干扰机的中心频率固定不变，因而有效的干扰频谱仅限于一个很窄的带宽之中，很宽范围内的多条频谱可能在短时间内被不断变化频率的扫频干扰干扰到。

（4）自适应干扰：也是一种认知攻击者，运用多臂赌博机 multiarmed bandit (MAB）算法，干扰器每次选择一组信道感知，并基于感知历史和过去的观测结果来干扰空闲信道。干扰器能利用结果来调整干扰策略，比如之前在传输信道上的干扰是否成功[4]。

（5）智能干扰：能够学习用户的传输策略并调整干扰策略，以最大化干扰效果[5]。

（6）反应式干扰：只有在检测到节点正在发送数据包时，反应式干扰器才开始发送干扰信号。这种类型的干扰器首先感知无线信道，在检测到信道忙即进行合法传输时，它为了破坏接收器处的数据接收而发送干扰信号[6]。反应式干扰器的性能取决于对合法用户状态的感知准确性。当干扰器接收到的信号较弱时可能不被监听到，它就不会发射干扰信号。由于信道空闲时干扰器保持安静，比持续干扰更节能而且不容易被检测到。

### 无人集群抗干扰技术现状与发展

#### 基于跳频和扩频的抗干扰方式

传统无线通信网络中的扩频和跳频等抗干扰技术，可以迁移应用至无人机通

信系统中。利用干扰源的检测和定位技术也可以辅助无人机通信制定轨迹控制等抗干扰策略，实现抗干扰的效果[7]。例如文献[8]提出一种跳频通信和固定信道结合的通信方案，利用跳频通信技术对无人机控制和感知链路的链接进行保障,使用固定信道进行高速率的数据传输以实现频谱共享。基于扩频通信技术的无人机抗干扰方案要求通信系统具有较为充裕的带宽，而在实际的无人机通信系统中，通信资源往往较为有限。此外，基于扩频通信和干扰检测技术的抗干扰方案也较难抵御动态优化自身干扰策略的智能干扰机。文献[9]详细分析了当前常用的使用跳频通信的无人机控制器的安全性能，指出只要跳频速度不超过 USRP等软件定义无线电设备的处理能力，就能有效地解出跳频所用的伪随机序列，可将序列应用于智能干扰中，能精确地于扰正在通信的信道，提高干扰机效率。

#### 基于路径规划和功率控制的抗干扰方式

无人机可以利用其高移动性的特点，利用通过控制自身飞行轨迹实现抗干扰。例如，文献[10]将无人机抗于扰问题建模为零和追击逃逸博弈问题，通过Issacs 方法求解最优无人机移动轨迹，减少无人机受到干扰的时间长度。Wu等人通过块坐标下降和连续凸近似方法，优化无人机的轨迹，在地面敌意干扰机的干扰下提高无人机的通信速率和降低能耗[11]。在无人机感知网络等典型无人机通信场景中，无人机还可通过块坐标下降和连续凸近似方法，联合优化三维轨迹和地面传感器调度策略，在存在强干扰的情况下提升无人机网络的通信速率。在无人机辅助的非正交多址接入通信系统中，利用基于块坐标下降的迭代算法进行轨迹控制优化，在满足无人机网络互相干扰低于阈值的前提下最大化网络吞吐量[12]。

除了轨迹控制之外，无人机也可以结合功率控制等联合控制方法，使无人机在提高抗干扰性能的同时，降低网络的能耗。例如，Xu等人将无人机抗干扰通信建模为斯塔伯格博弈模型，使用基于最优响应的方法优化无人机轨迹和功率控制联合策略，提高网络通信质量并降低能耗[13]。文献[14]设计了一种基于连续凸近似算法的方案，在多无人机系统中优化无人机的轨迹和发射功率，降低网络中无人机之间的交互干扰对通信的影响，并对算法进行并行化以进一步提高计算效率。Zhang等人利用块坐标下降和连续凸近似方法，对无人机的轨迹和发射功率进行联合优化，对抗地面窃听者，提高无人机通信链路的安全性[15]。

基于连续凸近似等算法优化轨迹和功率控制的无人机抗干扰技术，在对无人机的信道、任务和干扰机的干扰策略拥有精确模型时可以获得最优的抗干扰性能。然而在实际的无人机通信网络中，无人机信道模型较为复杂，且干扰机可能使用智能算法动态优化干扰策略，使得无人机难以获取信道模型和干扰模型，可能导致抗干扰性能下降。

#### 基于机器学习的抗干扰方式

无人机抗干扰可以通过引入机器学习技术，解决无人机状态信息获取困难、 传统抗干扰策略优化计算开销高等问题。例如，无人机可使用基于循环神经网络的干扰图谱构建方法，利用无人机周边环境信息和无人机载传感器的时序信息，实现对高风险的干扰区域的预测和回避[16]。文献[17]在无人机自组织网中利用联邦学习技术，使用信号强度和包送达率等无人机感知信息，实现网络参数共享，可在保护无人机感知数据隐私性的前提下，获得检测精度在85%以上的干扰检测模型，提升无人机网络的整体抗干扰性能。

然而，基于机器学习的抗干扰方案需要大量无人机抗于扰通信数据对分类器进行训练，以实现较高的抗于扰性能。当训练数据不足时，这些算法难以在真实的通信环境中取得较好的抗干扰能力。此外，真实无人机通信网络的高度复杂性和异构性，也使得深度学习方案难以获得较好的泛化能力,

## 强化学习研究进展与现状

强化学习（RL）是机器学习的一个子领域，目标是让代理学习如何根据环境状态采取行动，从而最天化预期的长期回报,其中学习问题通常可以建模为马尔可夫决策问题。早期的RL研究主要集中在表格和基于近似的算法。由于缺乏表示能力,传统的 RL 算法只能解决具有低维状态和动作空间的任务。然而更复杂、更接近现实世界的任务通常具有更高维度的状态空间和连续的动作空间,从而限制了RL的应用[18]。深度学习具有强大的表示能力,可以从高维抽象输入中提取多层次的特征.此外,深度神经网络已被证明是一般的函数逼近器,可用于逼近具有高维输入的复杂任务中的值函数和策略。因此,深度强化字习(DRL)近年来受到了广泛关注。自从深度Q网络 （DON)在游戏中成功应用以来,越来越多的深度学习技术和算法与RL相结合,不仅用于解决困难的传统 RL 任务,还激发了新的研究领域（元 DRL、迁移 DRL等）。DRL不仅在理论上取得了一些重要进展,在应用方面也取得了一些重要进展,例如机器人控制、游戏、NLP、驾驶、推荐系统和计算机视觉[19]。

强化学习方法分为基于价值和基于策略两种。基于价值的方法是强化学习方法中的重点在于表示价值函数并寻找最优价值函数。Q-learning[20]是最经典的基于价值的算法。在基于价值的方法中，动作选择的策略是不变的，如固定使用贪婪法作为策略选择方法。即在时间步t的状态下，选择动作的方式是固定的。而基于策略的强化学习方法则不同，在策略梯度法中，个体会学习不同的策略。基于价值的所有方法最终都是通过值函数来学习的，这些值函数可能和动作或状态有关。而基于策略的方法则是用梯度的方法直接学习策略，其在高维状态空间以及连续动作空间情况下取得更好的效果。

这里主要介绍强化学习中最重要的DQN发展研究。DQN在2015年由Mnih等人[21]首次提出。然而,使用神经网络来近似值函数已被证明是不稳定的,并且可能由于来自相关样本的偏差而导致发散[22],为了使样本不相关,Mnih等人创建了一个目标网络t,由β参数化,该网络每N步从估计网络t更新一次。此外,生成的样本存储在经验重放存储器中。然后从经验重放中随机检索样本并输入到训练过程中。DQN变体的第一个也是最简单的形式是[23]中提出的双DQN (DDQN)。DDQN的思想是将“贪婪”动作的选择与动作评估分开。通过这种方式,DDQN希望减少训练过程中对Q值的高估。DQN的另一个缺点是它使用四帧历史作为策略网络的输入。因此,DQN无法有效解决当前状态依赖于天量历史信息的问题,这种问题通常被称为部分可观察的 MDP问题。直接的解决方案是用循环长短期记忆替换策略网络最后一个卷积层之后的全连接层,如[24]中所述。这个 DQN 的变体称为深度循环Q网络(DRQN)在“Double Dunk”和“Frostbite”游戏中的表现比标准DQN高达700%。DRQN的另一个有趣变体是深度注意循环Q网络DARON)[25].在那篇文章中,Sorokin等人在 DRQN中添加了注意力机制,以便网络可以只关注游戏中的重要区域,从而允许更小的网络参数,从而加快训练过程。

## 基于强化学习抗干扰的研究进展与现状

利用强化学习技术可以解决抗干扰通信过程中因无人机移动性和视距信道造成的信道状态波动大，以及干扰模型未知等问题[26]。然而强化学习技术在面对状态动作空间较大的无人机抗干扰通信场景时，存在维度爆炸、环境探索效率较低等问题，导致抗干扰性能下降。针对上述问题，可以结合基于模型的强化学习、深度神经网络和最大熵[27]等技术实现高效的状态动作探索，缩短抗干扰算法的收敛时间。

在无人机和无人机辅助的通信网络中，可以使用强化学习进行轨迹、功率等 控制实现抗干扰通信[28]。例如，文献[29]设计了一种无人机对地面进行感知并回传数据的模型，并提出使用强化学习可以在轨迹控制、功率分配、子信道分配等方面进行优化，提升通信性能。作者设计了一个基于Q学习的轨迹无人机轨迹控制方案，使用无人机自身俭置信息优化飞行轨迹，提高感知和通信的效益。在有多个无人机作为中继的无人机抗十扰通信网络中，文献[30]设计了一种分布式强化学习方案，对每个光人机的中继功率独立地进行优化，降低整体无人机通信网络的能耗和误比特率。无人机还可结合多智能体强化学习技术，持续优化无人机向地面边缘设备卸载的数据量并选择通信协议类型，实现终端用户的低时延体验和无人机的高能效通信[31]。在无人机辅助车联网中，文献[32]提出了一种基于热启动Policyhillclimbing的强化学习方案，通过选择无人机的中继策略对抗智能干扰攻击。文献[33]基于深度强化学习和迁移学习技术，动态选择无人机抗干扰中继功率，对抗蜂窝网络中的智能干扰攻击，并解决通倍环境的高维状态空间和网络动态性等问题。文献[34]出了一种基于领域知识的强化学习无人机抗于扰方案对抗智能干扰攻击，使用领域知识压缩无人机抗干扰的高维状态空间，提高无人机轨迹和功率抗干扰策略的收敛速度。文献[35]把无人机抗干扰通信建模为斯塔伯格分层博奔模型，并设计了-种基于Q学习的信道选择算法，动态优化无人机的信道接入策略，可以提升无人机网络通信总速率。文献[36]所提方案观测自身的信道状态和干扰机的干扰功率，使用Q学习算法获得无人机功率控制策略。考虑到于扰攻击者的主观性，文献[37]提出一种基于前景理论的无人机抗干扰模型，并利用深度强化学习算法优化无人机的功率控制策略实现安全通信速率和信干比的提升。在无人机网络中，还可以通过机器人中继技术，利用强化学习优化机器人的移动性和中继功率，实现高能效、高可靠的无人机通信5。针对无人机视频传输场景，文献[38]提出基于强化学习的抗干扰视频传输方案，联合优化了视频传输参数和无人机的通信信道、功率等通信参数，来满足用户的视频传输体验质量需求并降低能耗。文献[39]提出可以使用基于WoLF—PHC算法联合优化发射功率和无人机轨迹实现无人机抗干扰效益和信干比的提升。

到目前为止在利用强化学习进行抗干扰的研究已经有很多，但是仅仅基于无人机之间的通信抗干扰的场景研究仍然较少，目前的研究难点主要在于模型训练的收敛性问题与模型的泛化性能保障两点。

# 研究意义和挑战

## 研究意义

无人机通信网络部署灵活、成本较低，在军事和民用领域都有巨大发展前景。然而无人机通信往往具有移动性较强、存在视距传输信道、通信距离较远等特点，且无人机自身计算、存储和通信资源较为有限，因此无人机通信的可靠性面临挑战。并且由于无线通信信道的开放特性，无人机通信容易受到智能敌意干扰攻击，可能导致业务中断、电量耗尽，甚至无人机坠毁等严重后果。因此研究存在智能敌意干扰机时如何保证无人机的通信安全具有重要意义。

传统无线网络中的抗干扰技术，如跳频和扩频通信等，由于无人机通信中的带宽资源和能耗受限，直接应用有一定局限。无人机还可以利用自身高移动性通过轨迹控制实现抗干扰通信[40]，但现有的无人机轨迹控制方案往往基于特定的无人机信道模型、干扰机位置和干扰策略等信息。此外还可以使用凸优化等方法获得轨迹控制策略，然而无人机的信道模型较为复杂，对现实无人机信道进行精确建模较为困难，且干抗机可能采用智能算法对干找策略持纹进行优化，使得干找策略更以预测，进而导致抗干扰性能下降。

功率控制技术可以在能量受限的无人机通信系统中，提高无人机通信的抗干扰性能，并节约网络中通信设备的能耗[41]。但是当通信时选择的发射功率较低时，可能会因为智能干扰机的强干扰导致通信质量急剧下降。因此，如何选择合适的发射功率，在通信能耗和通信质量之间进行权衡，是使用功率控制抗干扰技术的一个重要挑战。

综合以上研究难点，强化学习技术可以直接从部分可观察的无人机通信环境中获取抗干扰通信经验，结合运用多种抗干扰手段，组合优化抗干扰局策略，无需依赖于干扰机位置和干扰策略等难以在实际无人机通信环境中获取的信息[42]，直接进行端到端的抗干扰实现方式，因此具有很大的研究意义。但是由于无人机通信具有较大的状态动作空间，高维度的动作空间将导致训练难度的大幅度增加，利用该方法进行抗干扰任务时也面临着许多挑战，在设计基于强化学习的无人机抗干扰方案时，如何高效地探索抗干扰状态和策略是一个难点。

1. 张攀.无人机数据链抗干扰技术的研究与仿真[D].西安电子科技大学,2017.
2. T.Liu,J.Huang,J.Guo,and Y.Shan,“Survey on Anti-jamming Technology of UAV Communication”in 6GN for Future Wireless Networks,A.Li,Y.Shi,and L.Xi,Eds.,Cham:Springer Nature Switzerland,2023,pp.111–121.
3. 张新宇.无人机网络抗干扰方法研究[D].北京邮电大学,2019.
4. Wang,Q.,Ren, K.,Ning,P.,et al Jamming-resistant multiradio multichannel opportunistic spectrum access in cognitive radio networks [J].IEEE Transactions on Vehicular Technology,2016, 65(10): 8331-8344
5. Yang, D., Xue, G., Zhang, J., et al. Coping with a smart jammer in wireless networks: A Stackelberg game approach [J]. IEEE Transactionson Wireless Communications, 2013, 12(8): 4038-4047
6. Rawat, D B., Song, M. Securing space communication systems against reactive cognitive jammer [A].2015 IEEE Wireless Communicationsand Networking Conference (WCNC) [C], Washington: IEEE Computer Society, 2015: 1428-1433.
7. 林紫涵.基于强化学习的无人机通信抗干扰技术研究[D].厦门大学,2022.
8. Baek H, Lim J. Design of future UAV-relay tactical data link for reliable UAV control and situational awareness[J]. IEEE Communications Magazine, 2018, 56(10): 144--150.
9. Shin H, Choi K, Park Y, et al. Security analysis of FHSS-type drone controller[C]//International Workshop on Information Security Applications. Jeju Island, Korea: Springer, 2015: 240-253
10. Bhattacharya S, Basar T. Game-theoretic analysis of an aerial jamming attack on a UAV communication network[C]//American Control Conference. Baltimore, Maryland: IEEE, 2010: 818--823.
11. Wu Y, Guan X, Yang W, et al. UAV swarm communication under malicious jamming: joint trajectory and clustering design[J]. IEEE Wireless Communications Letters, 2021,10(10):2264-2268.
12. Zhao N, Pang X, Li Z, et al. Joint trajectory and precoding optimization for UAV-assisted Network[J].2019 673:3723-3735.
13. Shen C, Chang T-H, Gong J, et al. Multi-UAV interference coordination via joint trajectory and power control[J]. IEEE Transactions onSignal Processing, 2020, 68; 843-858.
14. Xu Y, Ren G, Chen J, et al. Joint power and trajectory optimization in UAV anti-jamming communication networks[C]WIEEE International Conference on Communications. Shanghai China: IEEE, 2019: 1--5.
15. ZhangG, Wu Q, Cui M, et al. Securing UAV communications via joint trajectory and power controllJ IEEE Transactions on Wireless Coumunications, 2019。18(2):1376-1389.
16. Challita U, Ferdowsi A, Chen M, et al. Machine learming for wireless connectivity and security of cellular-connected UAVs[J]. IEEE Wireless Communications, 2019, 26(1): 28-35.
17. Mowla N 1, Tran N H, Doh I, et ai. Federated leaming-based cognitive detection of jamming attack in flving ad-hoc network[J]. IEEE Access. 2020.8:4338-4350.
18. R. Munos and A. Moore,“Variable resolution discretization in optimalcontrol" Marh Learn 49 no 2 pn291-323.2002
19. A. T. Azar et al., “Drone Deep Reinforcement Learning: A Review,” Electronics, vol. 10, no. 9, Art. no. 9, Jan. 2021.
20. C. J. Watkins and P. Dayan,“Q-learning,” Mach. Learn,vol. 8 nos.3-4, pp, 279-292, 1992
21. V.Mnih et al.,Human-level control through deep reinforcement learning" Nature, vol. 518, no. 7540, pp. 529-533,2015
22. J. N. Tsitsiklis and B. Van Roy, "Analysis of temporal-difference learning with function approximation"in Proc. Adv. Neural Inf Process Syst.1997, pp. 1075-1081.
23. H. V. Hasselt, “Double Q-learning," in Proc. Adv. Neural Inf. Process Syst.,2010、pp. 2613-2621.
24. M.J. Hausknecht and P. Stone,“Deep recurrent Q-learning for partially observable MDPs," in Proc. AAAI Fall Symp. Series, Sep. 2015 pp. 29-37.
25. I. Sorokin,A.Seleznev,M. Pavlov,A. Fedorov, and A. Ignateva“Deep attention recurrent Q-network,"2015.
26. Meng F, Chen P, Wu L, et al. Power allocation in multi-user cellular networks: deep reinforcement leaming approaches[J]. IEEE Transactions on Wireless Communications, 2020, 19(10):6255-6267.
27. Haarnoja T, Zhou A, Abbeel P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement leaming with a stochastic actor[Cy Intemational Conference on Machine Learming. PMLR, 2018： 1861-1870.
28. Peng J, Zhang Z, Wu Q, et al. Anti-jamming communications in UAV swarms: a reinforcement learning approach[J]. IEEE Access, 2019，7： 180532-180543.
29. Hu J, Zhang 1-1, Song L, et al. Reinforcement learning for a cellular interne of UAVs: protocol design, trajectory control, and resource management[J]. IEEE Wireless Communications, 2020, 27(1): 116- 123.
30. Wang W, Lv Z, Lu X, et al. Distributed reinforcement learning based framework for energy- efficient UAV relay against jamming[J]. Intelligent and Converged Networks, 2021.
31. Sacco A, Esposito F, Marchetto G, et al. Sustainable task offloading in UAV networks via multi-agent reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2021, 70(5): 5003-5015.
32. Xiao L, Lu X, Xu D, et al. UAV relay in VANETs against smart jamming with reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2018, 67(5): 4087-4097.
33. Lu X, Xiao L, Dai C, et al. UAV-aided cellular communications with deep reinforcement learning against jamming[J]. IEEE Wireless Communications, 2020, 27(4): 48-53.
34. Li Z, Lu Y, Li X, et al. UAV networks against multiple maneuvering smart jamming with knowledge-based reinforcement learning[J]. IEEE Internet of Things Journal, 2021, 8(15): 12289-12310.
35. 范超琼, 赵成林,无人机网络中基于分层博弈的干扰对抗频谱接入优化[J].通信学报,2020,41(6): 26-33. DOI:10.11959/j.issn.1000-436x.2020114.
36. Lv S, Xiao L, Hu Q, et al. Anti-jamming power control game in unmanned aerial vehicle networks[C]//IEEE Global Communications Conference. Singapore: IEEE, 2017: 1-6.
37. Xiao L, Xie C, Min M, et al. User-centric view of unmanned aerial vehicle transmission against smart attacks[J]. IEEE Transactions on Vehicular Technology, 2018, 67(4): 3420-3430.
38. Xiao L, Ding Y, Huang J, et al. UAV anti-jamming video transmissions with QoE guarantee: a reinforcement learning-based approach[J]. IEEE Transactions on Communications, 2021, 69(9): 5933-5947.
39. 张孟杰,赵睿,王培臣,等.基于强化学习的无人机辅助物联网抗敌意干扰算法[J]. 信号处, 2021, 37(01): 11-18. DOI:10.16798/j.issn. 1003-0530.2021.01.002.
40. Zhao N, Pang X, Li Z, et al. Joint trajectory and precoding optimization for UAV-assisted NOMA networks[J]. IEEE Transactions on Communications, 2019, 67(5): 3723-3735.
41. Xu Y, Ren G, Chen J, et al. Joint power and trajectory optimization in UAV anti-jamming communication networks[C]//IEEE International Conference on Communications. Shanghai, China: IEEE, 2019: 1-5.
42. Lu X, Xiao L, Dai C, et al. UAV-aided cellular communications with deep reinforcement learning against jamming[J]. IEEE Wireless Communications, 2020, 27(4): 48-53.

# 课题研究目标

本课题的研究目标是利用深度强化学习网络的自适应性优化决策能力，融合传统的跳频和扩频抗干扰技术、轨迹规划和功率控制技术等实现多种抗干扰手段的协作。通过设计与训练模型一方面实现传统抗干扰手段的自适应优化参数决策，另一方面实现对不同电磁环境干扰手段的自适应选择。

本课题的研究内容主要由以下几点：

1）基于强化学习的组合抗干扰能力模型：本课题通过优化或融合多种现有发展的抗干扰手段，实现强对抗场景下的抗干扰能力的提升，因此根据不同抗干扰手段应对不同干扰类型的效果差异，通过强化学习模型评估该抗干扰方法所带来的增益效果，从而在不同干扰器，例如持续干扰、随机干扰、扫频干扰、自适应干扰、智能干扰和反应式干扰等产生的干扰场景下，采取对应效果的抗干扰方式或者结合多种抗干扰方式实现最佳决策。因此以上研究目标要求所设计的强化学习模型能够正确评估不同抗干扰模型在不同干扰场景下的能力，为此主要研究对深度神经网络部分引入注意力机制，使得接收机在抗干扰时能够关注相对重要的部分，来提高抗干扰性能。通过构建基于混合注意力机制的卷积神经网络和基于高效通道注意力机制的卷积神经网络提高组合抗干扰能力模型的预测精度与模型训练的收敛速度。

2）基于集中式全局信息训练与决策的无人集群智能组网架构

深度强化学习模型的训练与部署是实现基于强化学习抗干扰的关键内容，其中不同的训练和部署模式取决于模型输入信息的获取。本研究方向利用全局信息，包括所有无人机节点收集的动态空间谱信息、移动信息和功率信息等作为模型训练数据，能够实现信息的融合利用，从而充分提取当前全局电磁场景信息和任务信息下的特征，实现优化综合运用多种抗干扰技术决策的智能组网，通过设计该智能组网架构实现对应干扰方式的决策参数和组合参数输出。

3）基于多智能体的分布式强化学习抗干扰决策模型

由于无人机之间通信容量受限，因此抗干扰模型的部署应当尽可能降低获取信息带来的额外通信开销，所以节点抗干扰算法应当只依据本地的局部信息作为建模决策的依据。本研究内容通过利用集中式训练和决策智能组网架构实现分布式模型的部署，并且重点是保障和训练模型的一致性问题。因为各节点不利用全局信息，而是只通过本节点获取的本地信息作为模型的输入，缺少全局决策的优化，节点间的局部决策优化无法实现全局最优决策，因此本课题通过引入多智能体强化学习思想，通过建模各节点行为，以增加局部节点间相互决策的影响部分，实现多智能体联合优化抗干扰决策输出，降低由于局部信息不充分带来的决策无法达到全局最优的问题。

本课题拟解读的关键问题对应而来主要有以下几点：

1)集中式学习分布式执行二者之间模型的预测一致性保障。

2)引入多智能体强化学习算法以提高对环境的建模能力优化分布式决策

3)利用强化学习建模评估不同抗干扰技术对不同干扰器的能力。

4)利用强化学习模型优化传统抗干扰技术参数空间。

5)融合多种抗干扰技术，实现不同抗干扰技术的最优化组合

6)通过限制参数空间保障模型训练效果以及模型的收敛性。

# 研究方案

本课题采取计算机仿真的方式进行研究，通过设计仿真场景和开发所需的仿真功能模块实现并验证算法效果。仿真工具包含两部分：

1）网络仿真采用EXATA网络仿真软件。EXata 是一个由C++语言开发的强大的网络仿真平台，用于建模和分析无线和有线网络的性能。它支持各种网络协议和通信技术，包括移动 ad hoc 网络、传感器网络和多媒体通信。EXata 采用模块化设计，允许用户灵活定制仿真环境，以便进行不同场景和应用的测试。其特点包括高精度的物理层建模、支持大规模网络仿真以及与其他工具的集成能力，使其在学术研究、工程设计和网络优化等领域得到了广泛应用。通过 EXata，研究人员和工程师可以深入了解网络行为，评估新技术的性能，并优化网络架构。

2）强化学习建模与训练过程采用Pytorch。PyTorch 是一个由 Meta（原 Facebook）主导开发的开源深度学习框架，以其灵活性和易用性广受欢迎。它采用动态计算图（Dynamic Computation Graph），允许用户在运行时定义和修改模型结构，使调试和开发更加便捷。PyTorch 提供了强大的张量计算库和自动微分工具，适合处理复杂的深度学习任务，并在计算机视觉、自然语言处理等领域有着广泛应用。PyTorch 拥有丰富的模型层和优化器库，同时支持CPU和GPU加速。它还与其他深度学习库如 torchvision、torchtext 等配合使用，提供了大量预训练模型和工具包。2019 年发布的 PyTorch 1.0 版本集成了高性能的 C++ 后端Libtorch，从而在性能上与 TensorFlow 等框架持平，使 PyTorch 在学术界和工业界获得了广泛采用。此外，PyTorch 还包含分布式训练、混合精度训练等功能，满足从快速原型设计到大规模生产部署的多种需求。本仿真研究使用Libtorch作为强化学习的算法部分，用于在EXATA中完成对网络仿真数据的在线处理。

本课题的研究方案设计分为以下部分：

1）场景建模：通过exata仿真软件建模相应的场景，分为无人机集群部分以及干扰机两部分，无人机集群设计相应的任务信息和网络业务进行通信，包括机间链路和无人机与基站之间的通信链路。MAC层采用TDMA协议，路由协议采用OLSR协议，物理层采用802.11进行传输。干扰机采用压制式干扰，干扰信号的产生方式使用基于Q学习的智能式干扰机，干扰机设计对应的移动该模型。通过无人机网络和干扰机的建模设计相应场景，用于后续数据采集以及模型训练和效果验证。

2）多种抗干扰算法实现：本课题研究的抗干扰方案综合运用多种抗干扰算法，因此需要从不同协议栈实现对应的抗干扰算法，拟实现的抗干扰包括传统的抗干扰算法（如扩频和跳频技术）和轨迹规划与功率控制算法等，需要设计对应的算法实现，并且设计对应的参数接口，用于后续的强化学习训练和部署。

3）强化学习模型设计：设计基于价值的强化学习模型，利用libtorch库进行网络模型的设计并融合到exata中实现在线训练和部署，确定输入的数据维度与指标，明确输出参数的对应。搭建的模型基于DQN进行对应的注意力机制设计以及其他深度神经网络领域相关的最新进展，以获得最优的训练与部署效果。

4）模型训练与调整：通过仿真采集数据进行强化学习模型的集中式训练，并且根据效果进行模型超参数的调整，包括学习率、优化器、回报函数参数等，获得最佳的训练效果并验证模型收敛性。

5）分布式算法验证：将算法实际部署在对应场景中分析抗干扰性能的全方位评估，包括能量管理、性能提升等方面。最后需要验证模型的泛化能力，通过部署不同的干扰场景测试对应的抗干扰性能。

本课题的可行性分析为：本算法利用深度强化学习集中式训练模型并且分布式执行，模型训练阶段输入的数据为各节点接收机收到的信息，包括到达率、功率以及链路质量等信息，通过设计回报函数指导对应的抗干扰参数优化以及对应的抗干扰方式选择，根据抗干扰技术原理以及强化学习特点，该方法能够保证实现，但效果还需要验证。

# 创新点

本课题有以下创新点：

1）基于集中式全局信息训练与决策的无人集群智能组网架构

该智能组网架构利用全局信息进行集中式训练和决策，通过输入全局信息进在整体层面进行决策输出，能够从全局层面优化抗干扰决策。该方案通过实现基于强化学习的组合抗干扰能力建模，和以往抗干扰技术应对某特定的干扰类型不同，能够根据不同干扰场景，综合运用多种抗干扰手段，决策每种抗干扰手段的最优化参数以及多种抗干扰手段的最优化组合参数，以适应复杂时变的强对抗干扰场景，从而提高综合抗干扰能力。

2）基于多智能体的分布式强化学习抗干扰决策模型：

本课题通过引入多智能体强化学习，和以往利用全局信息进行决策的算法不不同，能够在只利用节点本地信息的情况下最大程度实现全局抗干扰决策的最优化输出。本课题适应无人机之间通信容量受限的场景，各节点不利用全局信息，而是只通过本节点获取的本地信息作为模型的输入，由于缺少全局决策的优化，节点间的局部决策优化无法实现全局最优决策，通过引入多智能体强化学习架构，以增加局部节点间相互决策的影响部分，实现多智能体联合优化抗干扰决策输出，降低由于局部信息不充分带来的决策无法达到全局最优的问题。

# 计划进度和预期成果

本课题的计划进度如下表：

|  |  |  |
| --- | --- | --- |
| 时间 | 进度 | 具体内容 |
| 2024.11 | 场景建模 | 建模相应的场景，包括无人机集群和干扰机建模 |
| 2024.12-2025.01 | 多种抗干扰算法实现 | 拟实现包括传统的抗干扰算法（如扩频和跳频技术）和轨迹规划与功率控制算法等 |
| 2025.01-2025.03 | 强化学习模型设计 | 设计基于价值的强化学习模型，利用libtorch库进行网络模型的设计并融合到exata中实现在线训练和部署 |
| 2025.03-2025.04 | 模型训练与调整 | 根据效果进行模型结构参数和超参数的调整 |
| 2025.04-2025.05 | 多智能体算法验证 | 算法实际部署在对应场景中分析抗干扰性能的全方位评估，验证模型的泛化能力 |

本课题的预期成果如下：本利用深度强化学习集中式训练模型并且分布式执行，模型训练阶段输入的数据为各节点接收机收到的信息，包括到达率、功率以及链路质量等信息，通过设计回报函数指导对应的抗干扰参数优化以及对应的抗干扰方式选择，能够处理在压制式智能干扰场景下实现针对单一或传统抗干扰手段抗干扰能力的提升。

# 现有工作积累

本课题现有的有关工作内容

1. 基于策略梯度下降（DDPG）模型的OLSR自适应控制包发包调整算法
2. 基于DQN的OLSR自适应控制包发包调整算法