arXiv:2603.23252v1 [cs.NI] 24 Mar 2026

非地面网络开放式无线接入网中分拆式无线智能控制器编排的人工智能生命周期感知可行性框架

DANIELE TARCHI，*IEEE* 高级会员

摘要—**将人工智能（AI）融入非地面网络（NTN），受到卫星尺寸、重量和功耗（SWaP）与馈电链路容量共同限制的制约，进而直接影响开放式无线接入网（O-RAN）的闭环控制和模型生命周期管理。本文研究通过分拆式无线智能控制器架构，将 O-RAN 控制层级分布于地面、低地球轨道和地球静止轨道各段的可行性。我们比较了三种部署场景：（i）采用遥测流传输的地面中心式控制；（ii）采用星载推理和存储转发学习的地面–低地球轨道分拆式无线智能控制器；以及（iii）由星间链路支持的地球静止轨道–低地球轨道多层控制。针对每种场景，我们推导了生命周期能耗和生命周期时延的闭式表达式，并将训练数据传输、模型分发和近实时推理纳入考量。针对馈电链路条件、模型复杂度和轨道间歇性开展数值敏感性分析后，得出了与运营商相关的可行性区域，明确了星载推理和非地面学习闭环在物理层面优于地面卸载的情形。**

索引词—**服务管理与编排（SMO）、开放式无线接入网、无线接入网智能控制器（无线智能控制器）、非地面网络、人工智能/机器学习生命周期管理、边缘智能**

一、引言

非地面网络（NTN）融入 6G 生态系统，有望弥合数字鸿沟，为服务不足的地区和偏远地区提供无处不在的覆盖 [1]。然而，运行一个跨越高容量地球静止轨道（GEO）卫星和高速移动的低地球轨道（LEO）星座的多层非地面网络，会带来时变拓扑、间歇性连接和显著的资源不对称性。因此，越来越需要通过机器学习（ML）实现自动化，以支持可扩展的网络与服务管理功能，例如资源控制、异常检测和自适应无线优化 [2]。因此，近期研究探索了将计算能力部署到星载节点上 [3], [4]。然而，机器学习在太空中的实际部署仍然有限

D. Tarchi 任职于意大利佛罗伦萨 50139 佛罗伦萨大学信息工程系，同时隶属 CNIT—佛罗伦萨大学研究组，地址为意大利佛罗伦萨 50139。本研究由欧盟委员会通过“5G-STARDUST”项目共同资助。该项目获得欧盟“地平线欧洲”研究与创新计划下智能网络与服务联合事业（SNS JU）的资助，资助协议编号为 101096573。文中观点仅代表作者本人，不一定代表该项目。欧盟委员会不对其中任何信息的使用承担责任。本研究还获得瑞士国家教育、研究与创新秘书处（SERI）的资助。本研究已提交 IEEE，拟申请发表。版权可能在未另行通知的情况下转让；此后，本版本可能不再可供访问。

这是因为存在一个根本性悖论：现代深度学习（DL）模型的计算开销很大，而卫星平台在尺寸、重量和功耗（SWaP）方面受到严重限制。具体而言，在不影响任务运行的前提下，训练和频繁更新模型可能与低地球轨道载荷计算的能量预算不相容。

通过虚拟网络功能（VNFs）、软件定义网络（SDN）和网络切片实现的网络软化，已成为管理此类异构基础设施的关键推动因素 [5]。在此背景下，开放无线接入网（O-RAN）通过标准化接口（例如 E2、A1、O1接口）和无线接入网智能控制器（RIC） [6]，为虚拟化、解耦式无线接入网实现提供了参考架构。开放式无线接入网最初是为回传链路充足且有电网供电的地面部署而设计的，但其控制环层级对于非地面网络尤其具有吸引力，因为它自然地将长时间尺度功能（非实时RIC）与时延敏感功能（近实时RIC）区分开来 [7]。在非地面网络环境中，这种区分使智能能够跨地空边界分布，但也引出了一个核心的服务管理问题：人工智能生命周期的哪些部分应在机载执行，哪些应卸载？通过馈电链路进行管理的成本又是多少？

尽管具有这种潜力，非地面网络、开放式无线接入网与人工智能的融合仍受到当前最先进水平中两项根本局限的制约。首先，许多研究将空间开放无线接入网视为把地面架构直接映射到轨道上 [8],[9], ，并且往往假设连接理想或星载计算能力充足。这导致生命周期建模支离破碎：以往研究通常只优化人工智能流水线的某一个阶段（例如推理时延或分布式训练收敛），却没有量化完整生命周期的端到端能量与时间成本，其中包括训练数据传输、模型分发和重复推理。尤其是，通过O1接口传输海量遥测信息与通过A1接口分发紧凑型模型制品所带来的管理平面负担，鲜少在统一框架中进行评估。其次，现有方案缺乏可行性边界。它们常常倡导固定的“全边缘”或“全云端”部署，却未指出这些选择在哪些物理条件下切实可行。因此，目前缺乏相关指导，无法将馈电链路条件、轨道间歇性（等待时间）和模型复杂度等关键非地面网络参数，映射到可行的无线接入网智能控制器部署位置集合。

本文通过研究面向受尺寸、重量和功耗限制的非地面网络运行而设计的分拆式无线智能控制器部署的可行性，弥补了这些不足。我们考虑一种解耦的人工智能生命周期，其中训练和模型更新集中在高

容量节点（地面或地球静止轨道）上，而对时延敏感的推理则下沉至低地球轨道边缘。本文并不旨在提出在线编排算法，而是提供支持决策的分析基础工具和与运营商相关的可行性区域，用于量化在现实的连接条件和资源约束下，何时星载智能在物理上优于地面卸载。

本研究旨在为非地面网络开放式无线接入网中的人工智能生命周期管理提供支持决策的可行性框架，通过推导闭式模型和运行边界，确定如何在地面、低地球轨道和地球静止轨道各层部署训练、模型更新和近实时推理。我们的分析量化了尺寸、重量和功耗约束、馈电链路状况与轨道间歇性如何共同影响生命周期能耗和时延，并划定了各分拆式无线智能控制器部署方案可行的物理条件。我们旨在建立分析基础要素和可行性区域，供未来服务管理与编排层优化及闭环保障方案用作约束和性能指标。我们提供可直接供服务管理与编排使用的决策支持型可行性边界，以便在尺寸、重量和功耗及馈电链路约束下规划 xApp/rApp 部署和人工智能生命周期操作。本文的具体贡献如下:

- **面向人工智能赋能的开放式无线接入网管理的生命周期成本建模：**我们推导了非地面网络开放式无线接入网中人工智能工作流的生命周期能耗（$\mathcal{J}_{cycle}$）和生命周期时延（$\mathcal{T}_{cycle}$）的闭式表达式，并明确计入了（i）训练数据传输、（ii）模型分发以及（iii）重复的近实时推理。该公式将更新开销（静态成本）与推理运行成本（动态成本）区分开来，从而能够定量分析通信与计算权衡，并可直接用作服务管理与编排层规划的性能指标。

- **地面/低地球轨道/地球同步轨道 Split-RIC 部署的比较可行性分析：**我们对开放式无线接入网智能的三种代表性部署方式进行了形式化建模和评估：地面中心式（*S1*）、地面–低轨 *Split-RIC*（*S2*）（采用星载推理和间歇式学习），以及依托星间链路的地球同步轨道–低地球轨道多层架构（*S3*）。该比较量化了地球静止轨道节点作为训练枢纽缓解低地球轨道间歇性的作用，并从运营角度阐明了持续的非地面学习循环在何种情况下有益。
- **与运营商相关的可行性区域和设计指南：**通过针对遥测数据量、馈电链路条件、模型复杂度和更新紧迫性（轨道等待时间）开展广泛的数值敏感性分析，我们推导出多维可行性区域，界定了在何种情况下，星载推理和/或非地面训练在物理层面上优于地面卸载。这些区域为 6G 非地面网络运营商提供了可行的阈值和设计指南，以便在符合实际的尺寸、重量和功耗以及连接约束条件下，规划人工智能功能部署和生命周期管理。

本文其余部分安排如下。第二节回顾开放式无线接入网的基础知识、非地面网络的轨道约束，

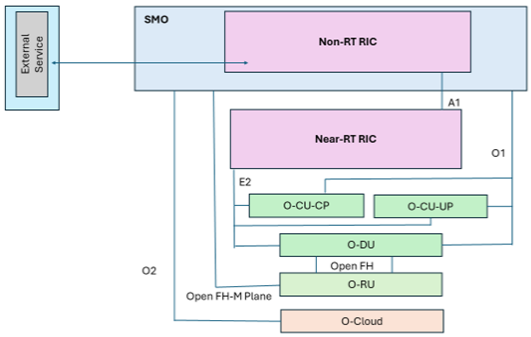


图 1. 开放式无线接入网功能架构

并从管理与生命周期的角度介绍卫星边缘智能的相关研究。第三节通过刻画地面/低地球轨道/地球同步轨道的资源配置，并形式化定义人工智能工作负载及其生命周期成本构成，建立系统模型。第四节介绍三种对比部署方案（地面中心式、分拆式无线智能控制器和多层）及其管理平面数据流。第五节推导分析可行性框架，并给出生命周期能耗和生命周期时延的闭式表达式。第六节报告数值敏感性分析，并划定星载智能与地面卸载的运行可行区域。最后，第七节总结全文并讨论未来研究方向。

II. 背景与相关工作

本研究探讨了三个各不相同但相互关联的领域交汇处固有的工程挑战：开放式无线接入网软件架构、轨道力学的物理约束（特别是低地球轨道和地球静止轨道），以及人工智能生命周期的计算成本。为构建所提出的分拆式无线智能控制器架构所需的技术基础，本节首先剖析与分布式智能相关的开放式无线接入网核心组件，随后阐述非地面网络的独特约束，最后评估将标准人工智能工作流整合到太空环境中的可行性。本分析指出了最先进的卫星边缘计算领域中亟待解决的技术差距，而这正是本研究旨在解决的问题。

A. 分布式智能的开放式无线接入网基础知识 *ence*

O-RAN 架构通过促进无线接入网解耦为开放、模块化且可互操作的组件，从根本上颠覆了传统无线接入网模型；这些组件过去曾紧密耦合为单体式黑盒。对于非地面网络和所提出的分拆式无线智能控制器架构，这种模块化并非仅仅为了便利，而是管理分布式卫星星座的必要条件。尽管开放式无线接入网规范内容广泛 [6], ，我们聚焦于对所提出的架构至关重要的实体：控制器（RIC）及其相关开放接口（A1接口、E2接口和O1接口）。图1展示了 O-RAN 功能架构的主要组成部分。

1）智能控制器（*RIC*）和应用：O-RAN 引入了一种由模块化应用驱动的分层控制环路机制：

- **非实时RIC（非实时 RIC）：** 该控制器位于服务管理与编排（SMO）框架内，运行时间尺度为 $>$ 1s。它承载 rApp（即非实时应用），这些模块化微服务负责提供高层指导、长期分析，并利用历史数据训练大型机器学习模型。在我们的架构中，rApp 部署在地面（或地球静止轨道），以利用无限的计算资源生成策略。

- **近实时RIC（近实时 RIC）：** 该控制器的运行时间尺度为 10ms 至 1 s。它承载 xApp（即近实时应用），这些轻量级、低时延微服务旨在执行特定的控制逻辑（例如波束跳变、干扰抑制）和机器学习推理。在我们所提出的架构中，这些 xApp 直接部署在低地球轨道卫星边缘，以确保实时响应能力。

2）关键接口：解耦通过标准化开放接口实现：

- **A1 接口：** 连接非实时 RIC 与近实时 RIC。对本文而言，最重要的是，该接口可扩展以传输机器学习模型权重 [10]。在太空中，这对应于馈电链路，用于将模型更新从地球传送至轨道。
- **E2 接口：** 连接近实时 RIC 与 E2节点（O-CU 和 O-DU）。该接口便于收集实时遥测，并直接在卫星载荷上执行控制动作。
- **O1 接口：** 用于故障、配置、计费、性能与安全（FCAPS）管理的流程。该接口使地面站能够收集原始数据轨迹。正如我们将要讨论的那样，过度使用该接口收集数据会造成瓶颈。

B. 轨道类型与约束

在太空中，尤其是在非地面网络环境中部署开放式无线接入网软件组件，会受到严苛的物理条件和信道约束的影响，其复杂性与地面部署截然不同。我们必须区分两种主要轨道类型，因为它们对人工智能和控制环路部署构成了截然不同但同样严峻的约束。

1) 低地球轨道（*LEO*）：低地球轨道卫星运行于靠近地球的轨道（$\approx$ 300千米至 2000 千米），并以高速飞行（$\approx$ 7.6千米/秒）。低地球轨道平台（例如立方体卫星）面临严格的尺寸、重量和功耗约束。典型的载荷功率预算低于 100 W，因此不适合使用高功耗图形处理器进行星载推理。此外，低地球轨道卫星与地面站之间的可见窗口很短，这使得传输大量数据颇具挑战。虽然低地球轨道馈电链路往返时延到地面网关可能相对较小（通常约为$\approx$ 20毫秒至 40 毫秒，包括传播时延和协议时延），但对于时延敏感的近实时控制环路（例如切换或波束跳变）而言，这一时延仍然过高，而此类控制环路通常要求当控制环路通过地面段闭合时，响应时间低于 10 毫秒。

2）地球静止轨道（地球静止轨道）：地球静止轨道卫星位于约 35 786 km 的高度，基本保持静止。信号单向传输大约需要 250 ms，往返时间超过 500 ms。尽管地球静止轨道平台通常具有千瓦级功率预算，足以支持更强大的星载计算，但较大的时延使得从地面进行实时控制变得不可行。地面操作员发出指令时，信道条件早已发生变化。因此，人工智能必须部署在星上才能发挥作用。

3）馈电链路约束：卫星与地面站之间的无线链路，即馈电链路，是一个瓶颈。它承载汇总后的用户业务流量、遥测、跟踪与指令（TT&C）数据，以及开放式无线接入网控制平面流量。传输原始 I/Q 采样以进行基于地面的机器学习训练，会消耗稀缺的带宽和大量发射功率（P_{tx}），直接影响卫星能量预算。

C. 人工智能生命周期

在卫星网络中实施机器学习并非一次性事件，而是一个持续的生命周期，包含数据收集、模型训练、模型部署和实时推理。与这些阶段通常共存于同一数据中心的地面网络不同，非地面网络环境要求根据各项任务特定的计算和通信特征，对其进行审慎的物理分布。

1) 训练阶段：深度神经网络（DNN）的训练是生命周期中资源消耗最大的阶段。它涉及迭代优化算法（例如随机梯度下降）以及通过反向传播计算梯度。训练需要大量矩阵乘法，而且至关重要的是，需要使用高精度浮点格式（FP32 或 FP64）以确保模型正确收敛。相比之下，抗辐射卫星载荷通常针对定点运算或低位宽整数（如 INT8）进行设计，以尽量降低功耗。在航天器上直接进行密集的浮点计算会产生过多热量，并迅速耗尽电池。因此，在我们所提出的架构中，训练过程完全集中在高容量节点上，主要位于地面段，也可能位于地球静止轨道平台。与高度受限的低地球轨道边缘相比，这种方法能够利用这些环境中显著更高的功率预算和高性能计算（HPC）资源。

2）推理阶段：推理是神经网络的前向传播，用于根据实时输入数据生成控制决策。虽然推理的计算开销低于训练，但由于执行频率高（例如，每个传输时间间隔（TTI）都要执行一次）且时延截止期限严格，因此对实际运行要求很高。在传统的弯管式架构中，卫星充当透明中继，将原始遥测数据（例如 I/Q 采样）传输到地面进行处理。虽然这样可以免去卫星的计算负担，却会使稀缺的馈电链路带宽达到饱和，并造成显著的往返时延，使实时控制难以实现。将推理引擎移至

卫星边缘后，我们便可在本地处理原始数据。这样无需传输高带宽遥测数据，且星载能耗几乎可以忽略不计。

3) 基本权衡：空间开放无线接入网的架构挑战归结为一个涉及两种相互耦合权衡的多维优化问题：

- **能量（计算与通信的权衡）：** 我们必须在将原始遥测数据卸载到地面所需的通信能耗（发射功率较高 P_{tx}）与在本地处理这些数据所需的计算能耗（星载电池消耗）之间取得平衡。
- **时延（传播与处理的权衡）：** 我们必须在地面环路固有的传播时延（往返时间）与受限星载加速器的处理时延之间取得平衡。地面服务器速度更快，但距离较远；卫星离用户更近，但计算速度较慢。

我们的可行性分析旨在确定可行区域，使星载推理在满足近实时 RIC 严格时序要求（$\tau < 10$毫秒）的同时，最大限度地降低总能耗。

D. 相关工作与差距分析

非地面网络、开放式无线接入网架构与人工智能的融合是一个新兴但发展迅速的领域。我们将最先进水平的研究归纳为三个主要方向。

1）非地面网络中的开放式无线接入网集成：第三代合作伙伴计划（第17/18版）和 O-RAN联盟早期的标准化工作，重点是定义非地面网络所需的协议栈适配。多项研究考察了开放式无线接入网功能（O-CU、O-DU、O-RU）在卫星有效载荷中的最佳部署位置。 [8]和 [9] 等研究提出了自适应功能拆分，以应对低地球轨道和地球静止轨道链路不断变化的时延。然而，这些研究主要关注协议同步和吞吐量优化，将卫星视为静态计算节点，而非动态的能量受限平台。 [11] 和 [12]的研究探讨了分层架构，其中低地球轨道卫星充当地球静止轨道控制器的中继。虽然这些研究定义了拓扑连接，但往往忽视了 O-RAN接口（A1接口、E2接口、O1接口）经由无线馈电链路引入的控制平面开销。

2）卫星通信中的人工智能与机器学习：机器学习在优化卫星运行方面的应用已有充分记载，但往往与开放式无线接入网框架相脱节。已有大量文献探讨利用深度强化学习（DRL）进行波束跳变、功率分配和干扰管理 [13], [14]。然而，这些研究大多假设模型以“黑盒”方式部署，而忽略了运行推理引擎的能耗成本 [15]。已有研究提出时间序列预测模型（例如长短期记忆网络（LSTM）、Transformer），用于预测低地球轨道星座中的业务负载 [16]。这些研究通常假设在地面进行集中式训练，却未量化通过遥测接口持续收集高分辨率训练数据所需的带宽成本。

3）分布式边缘智能与生命周期管理：在更广泛的地面6G领域，拆分人工智能训练与推理的概念正受到越来越多的关注。研究者已探索跨边缘节点分布模型训练的技术，以保护隐私并减少带宽占用 [17]。然而，将这些地面协议应用于太空场景面临独特挑战，尤其是低地球轨道卫星的间歇性连接，以及星载反向传播带来的巨大能耗，这些问题在当前文献中仍未得到充分研究 [18]。尽管量化和剪枝是嵌入式人工智能中的标准做法，但它们在开放式无线接入网A1接口卫星链路中的具体应用仍有待研究，因为每一比特的控制开销都要与用户数据争用资源。

4) 差距分析： 尽管取得了这些进展，当前的技术水平仍存在两项根本性局限，阻碍了开放式无线接入网在空间中的实际部署：

- **碎片化的生命周期建模：** 现有研究通常只优化人工智能流水线的单个阶段，要么仅关注推理准确率，要么仅关注训练收敛。当前严重缺乏定量模型来比较完整生命周期的端到端能耗，尤其是在现实馈电链路预算下，对比通过 O1接口传输海量原始遥测数据与通过 A1接口传输压缩模型权重之间的权衡。

- **静态架构假设：** 以往工作往往主张采用固定的部署策略（例如“全边缘”或“全云端”），却没有界定其可行性的物理边界。它们未能确定具体的交叉点，即星上处理在何种情况下会在物理层面优于地面卸载。

我们的贡献： 本研究通过为分体式 RIC 架构建立首个严谨的分析可行性框架，弥合了这些差距。不同于以往的定性综述，我们开展了全面的敏感性分析，以绘制具体的可行性区域：在这些区域内，根据链路速率、数据量和轨道等条件，星上人工智能在能耗和时间方面均优于传统的地面控制。

III. 系统模型与资源形式化

我们将多层非地面网络建模为时变图（TEG） $\mathcal{G}(t) = (\mathcal{V},\ \mathcal{E}(t))$。在这一动态拓扑中，节点集合 $\mathcal{V}$ 分为两个物理上截然不同的域：资源充足的地面段（$\mathcal{G}_{st}$）和资源受限的空间段（$\mathcal{S}$）。不同于无线智能控制器可集中部署在供电可靠的城域边缘云中的地面网络，空间开放无线接入网 架构必须在物理上解耦控制层级。我们提出一种分拆式无线智能控制器框架，将学习生命周期分布式部署：繁重的训练阶段位于地球上，最终也可部署在地球静止轨道卫星上；而对时延要求严苛的推理阶段则下沉至低地球轨道卫星边缘。为评估分布式部署人工智能工作负载的可行性，我们首先对物理域（即节点）、连接层（即链路）以及人工智能任务的特性进行形式化定义。

表 I 节点资源规格

参数	地面（$\mathcal{G}_{st}$）	低地球轨道（$\mathcal{S}_{LEO}$）	地球静止轨道（$\mathcal{S}_{GEO}$）
硬件类型	云端 GPU	抗辐射 FPGA	空间数字信号处理器
容量 (F_n)	>1 PFLOPS	$\approx$ 10每秒万亿次浮点运算	$\approx$ 200每秒万亿次浮点运算
功率（P_n^{max}）	无限	20 W	1000 瓦
效率（ϵ_n）	200 皮焦耳/浮点运算	20 皮焦耳/浮点运算	100 皮焦耳/浮点运算

A. 物理域与节点能力

该网络由三个不同的物理域组成。我们通过资源配置 $\mathcal{R}_n = \langle F_n, P_n^{max}, \epsilon_n \rangle$来刻画每个节点 n，分别表示计算能力、功率预算和硬件能效。表 I 总结了所采用的数值，这些数值来源于地面及航天级处理器的最新硬件规格。

1) 地面段（$\mathcal{G}_{st}$）：地面段 $\mathcal{G}_{st}$ 由测控地面站及相关地面云基础设施组成。它作为全球遥测数据的汇集端，能够存储数 PB 的历史数据集 $\mathcal{D}$ 而不受存储限制，同时也是非地面网络的物理锚点，提供回传连接并接入地面数据中心。与天基节点不同，地面段的特点是拥有近乎无限的电力供应和高性能计算集群。我们将地面段节点 $n_{gnd} \in \mathcal{G}_{st}$ 建模为相对于空间段具有无限计算能力（$F_{gnd} \to \infty$）。尽管性能强大，地面 GPU（例如 NVIDIA A100/H100）优先考虑吞吐量而非能效，因此每次运算的能耗成本（$\epsilon_{gnd} \approx$200 pJ/浮点运算）高于专用边缘加速器。

2) 第一级空间段：*LEO*边缘（$\mathcal{S}_{LEO}$）：低地球轨道卫星（$\approx$ 600km）作为面向用户设备的资源受限边缘接口。它们的特点是轨道动力学变化剧烈，且受到严格的尺寸、重量和功耗约束。有效载荷通常依赖节能硬件加速器，例如抗辐射现场可编程门阵列或片上系统（SoC）加速器（如 Xilinx Versal AI Core、NVIDIA Jetson）。可用于开放无线接入网处理的功率预算，P_{budget}^{LEO}，严格受太阳能电池阵列发电量和电池深度限制（计算有效载荷通常为 $\approx$ 20W）。不过，这些设备通过低精度整数运算（INT8）实现了较高能效（$\epsilon_{LEO} \approx$ 20pJ/浮点运算）。由于靠近地球，低地球轨道节点到其服务用户的传播时延最低（单跳时延约为几毫秒至 $\approx$ 10ms，具体取决于几何位置）。因此，当推理和执行在星上本地进行（即无需经过馈电链路）时，低地球轨道节点能够支持近实时控制环路。

3）第二级空间段：地球静止轨道枢纽（$\mathcal{S}_{GEO}$）：地球静止轨道卫星（$\approx$ 36 000km）充当高容量区域枢纽。地球静止轨道平台采用更大的卫星平台，可搭载机架式服务器刀片或强大的数字信号处理器集群。虽然其能力不及地面，但地球静止轨道功率预算显著更高（$P_{GEO}^{max} \approx 1$ kW），可支持复杂处理器持续运行。较高的传播时延（$>$ 250ms）使地球静止轨道节点不适合亚毫秒级无线电环路控制。不过，其强大的计算能力（$F_{GEO} \gg F_{LEO}$）以及对低地球轨道卫星轨道面的持续可视性，使其适合承载聚合任务、分布式训练或对时延不敏感的网络编排。

B. 接口映射与链路动态特性

物理段之间的连接由标准 O-RAN 逻辑接口控制。然而，在非地面网络环境中，这些逻辑接口必须映射到具有不同信道特性和资源约束的物理无线链路上。

1) 馈电链路（物理 *A1/O1* 接口）：地面至卫星馈电链路充当垂直 O-RAN 接口的物理传输层。该链路承载 A1 接口（从非实时 RIC 到近实时 RIC 的策略/模型更新）和 O1 接口（从卫星到服务管理与编排的管理/遥测）。与基于光纤的地面前传不同，该链路是一种共享介质，承载用户面流量、遥测、跟踪与指令数据以及开放无线接入网控制面消息。带宽B_{feed} 受到严格限制，可实现的上行数据速率$R_{feed}(t)$受香农极限约束，取决于随时间变化的斜距 $d_{s,\ g}(t)$ 以及卫星的发射功率 P_{tx}。从太空向地面传输数据会消耗大量功率 P_{tx}^{feed}。

2）星间链路（物理 *Xn/E2*）：星间链路（ISL）通常采用光学激光技术（OISL），为卫星之间提供水平连接。这些链路承载 Xn 接口（卫星之间的切换协调），并在分布式场景中承载 E2接口（例如，一颗卫星上的近实时 RIC 控制另一颗卫星上的 O-RU）。与射频馈电链路相比，星间链路（ISL）具有更高的带宽（R_{ISL}）和更低的单位比特传输能耗（P_{tx}^{ISL}）。低地球轨道与地球静止轨道之间的连接持续可用，无需再“等待”地面站过境。

C. 人工智能工作负载特征分析

为了将智能映射到物理节点上，我们对人工智能过程进行数学表征。不同于仅考虑单个计算块的传统任务卸载模型，我们将人工智能生命周期建模为一对相互关联但各有区别的工作负载：推理（频繁执行）和训练（定期执行）。

令 $\mathcal{K} = \{1, \ldots, K\}$ 表示网络所需的人工智能任务集合。每个任务 $k \in \mathcal{K}$ 由一个复合元组 $\Phi_k = \langle \mathcal{W}_k^{inf}, \mathcal{W}_k^{train}, \sigma_k \rangle$ 定义，分别对应推理配置、训练配置和模型制品。

推理配置是一个元组 $\mathcal{W}_k^{inf} = \langle \delta_k^{in}, \delta_k^{out}, \omega_k^{inf}, \tau_k^{max} \rangle$，，，，用于定义近实时执行要求，其中 δ_k^{in}是推理决策所需的原始输入遥测数据大小（例如，一个 I/Q 快照），δ_k^{out}是推理输出的大小（例如，波束索引控制命令），ω_k^{inf} 是量化模型前向传播的计算复杂度（以浮点运算次数计），τ_k^{max}是严格的时延截止期限（例如，近实时 RIC 为 10 ms）。通常 $\delta_k^{out} \ll \delta_k^{in}$，但必须将其传输到执行器（O-DU/O-RU）。

训练配置是一个元组 $\mathcal{W}_k^{train} = \langle \Delta_k^{train}, \Omega_k^{train} \rangle$，用于定义模型更新要求，其更新周期为 T_{update}，其中 Δ_k^{train}是历史训练数据集 $\mathcal{D}$ 的大小，Ω_k^{train}是反向传播过程的计算复杂度（训练轮数 × 、批量大小 × 及梯度计算）。请注意，$\Delta_k^{train} \gg \delta_k^{in}$，因为它汇集了数千个样本。

模型制品为 σ_k，对应于必须通过 A1 接口从训练节点传输到推理节点的压缩模型权重大小。

我们将这些任务映射到可用物理节点集合 $\mathcal{V} = \{\mathcal{G}_{st}, \mathcal{S}_{LEO}, \mathcal{S}_{GEO}\}$。每个节点 $n \in \mathcal{V}$ 的特征由其资源配置 $\mathcal{R}_n = \langle F_n,\ P_n^{max},\ \epsilon_n \rangle$ 决定，其中F_n 是节点的最大计算吞吐量（$F_{Gnd} \gg F_{GEO} > F_{LEO}$），$P_n^{max}$是分配给人工智能处理的最大功率预算，$\epsilon_n$ 是硬件加速器的能效。专用低地球轨道级现场可编程门阵列通常具有更高的能效（更低的焦耳/浮点运算，即 $\epsilon_{LEO} < \epsilon_{Gnd}$），而地面 GPU 优先考虑原始吞吐量（$F_{Gnd}$），代价是功耗更高。

D. 广义成本模型

为评估不同架构拆分在各种场景下的表现，我们定义了一个统一的成本模型。总系统成本是两个正交指标的函数：在指定节点执行任务并传输必要数据产物（输入遥测数据、训练数据集或模型权重）所需的能耗 ($\mathcal{J}$)，以及闭合控制回路或完成学习周期所需的端到端时延 ($\mathcal{T}$)。设 n_{proc} 为分配用于处理（推理或训练）的节点，n_{src} 为数据来源节点。

1) 能耗模型（$\mathcal{J}$）：总能耗成本为：

$$\mathcal{J}_k(n_{src}, n_{proc}) = E_{comp}(n_{proc}, \Omega) + E_{comm}(n_{src} \to n_{proc}, D) \quad (1)$$

其中，E_{comp} 为计算成本，E_{comm} 为通信成本。计算成本对应于节点 n_{proc}处理复杂度为 Ω 的工作负载所需的能量：

$$E_{comp}(n_{proc}, \Omega) = \Omega \cdot \epsilon_{n_{proc}} \quad (2)$$

其中，$\epsilon_{n_{proc}}$ 为硬件效率（焦耳/浮点运算）。通信成本是通过链路 l 将数据量 D 从源节点 n_{src} 传输到目标节点 n_{proc} 所需的能量：

$$E_{comm}(n_{src} \to n_{proc}, D) = \begin{cases} P_{tx}^{link} \cdot \frac{D}{R_{link}} & \text{if } n_{src} \neq n_{proc} \\ 0 & \text{if } n_{src} = n_{proc} \end{cases} \quad (3)$$

2) 时延模型（$\mathcal{T}$）：总时延定义为传输时延、传播时延、处理时延和传输窗口等待时延之和：

$$\mathcal{T}_k(n_{src}, n_{proc}) = T_{comm}(n_{src} \to n_{proc}, D) + T_{comp}(n_{proc}, \Omega) + T_{wait} \quad (4)$$

表 II 建模假设汇总

领域	假设	依据
Data	原始遥测数据	我们假设 O1 数据（δ^{in}）未经压缩以建模最坏情况下的带宽负载频谱监测（例如 I/Q 采样）。
计算	硬件异构性	低地球轨道节点采用节能型 INT8 量化，而地面/GEO 节点采用 ϵ_{LEO}，采用高精度 FP32 (ϵ_{GND})，这体现了标准的边缘与云端 AI 实践。
功率	热节流	我们假设低地球轨道功率预算 (P_{comp}) 是一个考虑了占空比的平均数值-为防止热过加热。
拓扑	星间链路可用性	假设地球静止轨道—低地球轨道光链路在更新期间始终可用窗口，将地球同步轨道卫星视为“数据中继”骨干网。
训练	单样本学习	我们将训练阶段建模为一个离散事件（T_{train}）；迭代式联邦学习多个轮次合并为单一能量成本，以便清晰说明。

其中，T_{comm} 是包含序列化延迟和物理传播延迟的通信时延：

$$T_{comm} = \begin{cases} \frac{D}{R_{link}} + \text{RTT}_{link}/2 & \text{if } n_{src} \neq n_{proc} \\ 0 & \text{if } n_{src} = n_{proc} \end{cases} \quad (5)$$

T_{comp} 表示在目标硬件上执行工作负载所需的计算时延：

$$T_{comp} = \frac{\Omega}{F_{n_{proc}}} \quad (6)$$

T_{wait} 表示在可用传输窗口（即与低地球轨道卫星之间的传输窗口）出现前应等待的时长。

这种通用表述使我们能够量化不同架构选择的成本。例如，卸载到地面会导致较高的 T_{comm}（由于往返时延）以及可能较高的 T_{wait}；而采用星上处理则会将 T_{comm}降至零，但会增加星载能耗负担。

E. 模型假设与适用范围

为确保分析上的可处理性，同时保留非地面网络环境的物理真实性，我们采用以下建模假设。这些选择总结于表 II。关键约束的依据：

- **遥测压缩：** 虽然无损压缩可以降低 δ^{in}，但许多 6 G xApp（例如神经波束成形）需要原始复值信号，无法接受压缩伪影。因此，我们对传输成本的上限进行建模。
- **量化级别：** 我们基于采用 INT8 现场可编程门阵列加速器的方案，设定每 $\epsilon_{LEO} = 20$浮点运算为/pJ，因为在电池电量受限的立方体卫星上运行完整 FP32 推理通常并不可行。地面站不受此类限制，因此采用标准 FP32精度，以实现最高模型准确率。
- **星间链路稳定性：** 与容易受到高多普勒效应和频繁切换影响的低地球轨道星间链路不同，低地球轨道—地球静止轨道链路在较长时间内（数十分钟）具有几何稳定性，因此，下一节所述的场景 3 中假设信道持续可用是合理的。

IV. 对比部署架构

基于上述系统模型，我们确定了开放式无线接入网控制层级的三种不同架构配置。这些场景如图 2 所示，代表了人工智能工作负载（$\mathcal{W}_{inf}$，$\mathcal{W}_{train}$）到物理节点（$\mathcal{G}_{st}$，$\mathcal{S}_{LEO}$，$\mathcal{S}_{GEO}$）的特定映射，从而形成具有不同拓扑特征的成本分布。

A. 场景 *1*：基线地面中心控制（*S1*）

此场景代表传统的“弯管式”架构（符合 3GPP Rel-17 非地面网络与地面网络互通），卫星严格作为透明物理层中继器运行。整个 O-RAN 智能栈集中部署在地面。地面段同时承载非实时 RIC（用于训练）和近实时 RIC（用于推理），将卫星节点视为不具备本地决策能力的远程无线电单元（O-RU）。运行数据流分为三个阶段：(i) 低地球轨道卫星捕获射频信号并将其数字化，然后通过馈电链路（映射至 O1 接口）直接向地面站传输原始、未压缩的 I/Q 遥测数据（δ^{in}）；(ii) 地面近实时 RIC 使用高性能图形处理器处理这一海量数据流，以计算所需的控制动作（例如波束加权）；(iii) 控制命令通过馈电链路（映射至 E2接口）传回卫星，以供执行。

虽然这种架构最大限度地降低了星载尺寸、重量和功耗成本（$E_{comp} \approx 0$），但却给馈电链路带来了巨大负担。系统严格受往返传播时延（RTT$_{feed}$ ≈ 20毫秒）限制，因此无法满足亚毫秒级控制环路的要求。

B. 场景 *2*：拟议的分拆式无线智能控制器（*S2*）

这是我们为实现能效运行提出的主要方案，通过引入“功能拆分”，将学习生命周期解耦，使其与节点能力相匹配。我们提出将 RIC 进行物理分离：近实时 RIC 直接嵌入低地球轨道卫星的有效载荷中（利用高效的现场可编程门阵列执行任务），而非实时 RIC 则保留在地面段（利用云集群进行训练）。运行数据流分为三个阶段：（i）原始遥测数据在星上本地处理。xApp 执行推理（ω^{inf}），无需与馈电链路交互即可立即生成控制动作；（ii）历史训练样本 Δ^{train} 缓存在星上，仅在有效可见窗口期间卸载至地面（存储转发）；（iii）在地面完成训练后，更新后的模型权重 σ 经压缩后通过 A1 接口上传至 LEO 节点。在此配置中，近实时 RIC 与卫星上的 O-DU/O-RU 功能共置于星载处理器上。因此，E2 接口通过内部总线通信实现（例如，在 FPGA 片上系统上使用 AXI 互连），与外部 A1 和 O1 链路相比，其时延可忽略不计。因此，我们将 E2 交互建模为相对于轨道传播时延而言瞬时完成。

该架构通过传输紧凑模型（σ）而非原始数据（δ^{in}），降低馈电链路能耗。然而，“学习循环”并非持续进行，而是受到访问地面站所需轨道等待时间（T_{wait}）的限制。

C. 场景*3*：多层*GEO-LEO*控制（*S3*）

这一高级场景利用空间网络的层级特性，解决场景2中的连接间歇性问题。控制平面完全位于非地面环境。近实时 RIC部署在低地球轨道边缘，而非实时 RIC则托管在高容量GEO枢纽上。通过星间链路（ISL）维持连接。运行数据流分为三个阶段：(i) 与场景2类似，在低地球轨道节点上本地执行实时控制；(ii) 低地球轨道卫星无需等待地面过境，而是通过星间链路持续向GEO节点卸载训练数据 Δ^{train}（通常使用高带宽光链路）；(iii) GEO节点执行训练，并通过星间链路上的A1接口将更新推送回低地球轨道。

无论卫星位于海洋上空还是偏远地区，这都能实现持续学习（$T_{wait} \approx 0$）。该方案以射频馈电链路较高的能耗成本，换取光学星间链路通常较低的单位比特能耗成本，因此非常适合地面网络稀疏的场景。需要注意的是，尽管GEO中继会引入约120毫秒的传播时延，但该链路仅用于非实时 RIC功能（A1策略更新和O1遥测聚合）。这些管理平面操作的时间尺度为数秒至数分钟。对时延敏感的推理循环仍在低地球轨道节点本地运行，因此较高的GEO时延不会影响无线接入网所需的亚毫秒级实时控制。

V. 分析可行性框架

为评估所提出架构的可行性，我们针对第四节定义的三种场景，推导出总生命周期能耗($\mathcal{J}$)和生命周期时延($\mathcal{T}$)的闭式表达式。这些表达式将第三节中的广义成本模型映射到各部署策略的具体拓扑约束。

A. 生命周期定义

我们基于一个完整的从学习到推理的周期来定义可行性指标。单个周期由以下三个连续阶段组成：

1) **训练阶段：** 将大小为$\Delta^{train} = |\mathcal{D}|$（比特）的历史数据集从边缘卸载到训练节点。2) **更新阶段：** 将大小为 σ（比特）的已训练模型权重传输回推理节点。3) **推理阶段：** 在模型的生命周期内执行 N_{inf} 次推理操作。每次操作需要 δ^{in}个输入比特和 ω^{inf} 次浮点运算。

参数 N_{inf} 表示模型寿命——即在概念漂移要求启动新的训练周期之前，可以进行多少次推理决策。

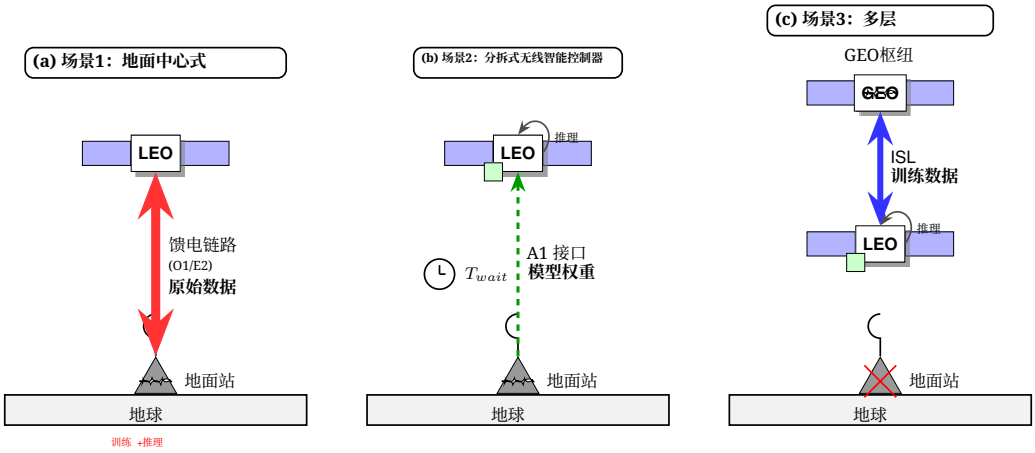


图 2. 开放式无线接入网部署架构对比。(a) S1：传统弯管式方案，所有智能功能均部署在地面。(b) S2：所提出的分拆式无线智能控制器方案，在边缘进行推理，在地面进行训练。(c) S3：分层方案，在边缘进行推理，并通过星间链路在地球静止轨道枢纽上进行训练。

B. 场景*1*：基线地面中心式（*S1*）

在这一基线方案场景中，卫星充当透明中继。训练和推理均在地面网段（$\mathcal{G}_{st}$）中进行。

1) 生命周期能耗（$\mathcal{J}_{S1}$）：卫星仅消耗能量用于通信：传输训练数据集，并在此后为每个推理事件传输原始遥测数据：

$$\mathcal{J}_{S1} = \underbrace{P_{tx}^{feed} \cdot \frac{\Delta^{train}}{R_{ul}^{feed}}}_{\text{Training Offload}} + \underbrace{N_{inf} \cdot \left(P_{tx}^{feed} \cdot \frac{\delta^{in}}{R_{ul}^{feed}} \right)}_{\text{Real-Time Inference Streaming}} \tag{7}$$

$\mathcal{J}_{S1}$ 随 N_{inf} 和 δ^{in}线性增长。对于数据密集型 xApp（例如频谱分析），第二项占主导地位，会迅速耗尽低轨卫星电池。

2）生命周期时延（$\mathcal{T}_{S1}$）：周期时延包括卸载训练数据所需的轨道等待时间，以及每次推理操作的往返时延（RTT）：

$$\mathcal{T}_{S1} = T_{wait}^{gnd} + \frac{\Delta^{train}}{R_{ul}^{feed}} + T_{train}^{gnd} + N_{inf} \cdot \left(\text{RTT}_{feed} + \frac{\delta^{in}}{R_{ul}^{feed}} \right) \tag{8}$$

推理循环受到馈电链路传播时延（≈ 20毫秒）的影响，因此对于要求时延低于 $\tau < 10$毫秒的控制环路而言，这种架构并不可行。

C. 场景 *2*：拟议的拆分式 *RIC*（*S2*）

在分体式 RIC 架构中，训练仍在地面进行，但推理移至 LEOEdge（$\mathcal{S}_{LEO}$）。

1）生命周期能耗（$\mathcal{J}_{S2}$）：卫星只需一次性支付训练数据的传输成本和模型权重的接收成本。值得注意的是，推理能耗现在由星载硬件效率（ϵ_{LEO}）而非无线链路决定：

$$\mathcal{J}_{S2} = \underbrace{P_{tx}^{feed} \cdot \frac{\Delta^{train}}{R_{ul}^{feed}}}_{\text{Training Offload}} + \underbrace{P_{rx}^{feed} \cdot \frac{\sigma}{R_{dl}^{feed}}}_{\text{Model Rx}} + \underbrace{N_{inf} \cdot (\omega^{inf} \cdot \epsilon_{LEO})}_{\text{On-Board Inference}} \tag{9}$$

由于 $\omega^{inf} \cdot \epsilon_{LEO} \ll P_{tx}^{feed} \cdot \frac{\delta^{in}}{R_{ul}}$，随着 N_{inf} 增长，这种架构的能效会不断提高。这体现了计算与通信增益。

2）生命周期时延（$\mathcal{T}_{S2}$）：尽管推理几乎是瞬时完成的（$T_{comp} \approx 0$），但间歇性的地面链路会增加学习时延。卫星必须等待链路过境才能上传数据，并且可能还要等待下一次过境才能下载模型：

$$\mathcal{T}_{S2} = \underbrace{2 \cdot T_{wait}^{gnd}}_{\text{Double Wait Penalty}} + \frac{\Delta^{train}}{R_{ul}^{feed}} + T_{train}^{gnd} + N_{inf} \cdot T_{comp}^{LEO} \tag{10}$$

D. 场景*3*：多层*GEO-LEO*（*S3*）

在此场景中，控制回路完全位于非地面。低地球轨道卫星负责推理，而地球同步轨道卫星则充当训练中心。

1）生命周期能耗（$\mathcal{J}_{S3}$）：与训练能耗被视为“免费”（由电网供电）的地面中心场景不同，多层场景会在低地球轨道和地球同步轨道节点上产生能耗成本。系统总能耗是低地球轨道推理回路与地球同步轨道训练回路的能耗之和：

$$\mathcal{J}_{S3} = \mathcal{J}_{S3}^{LEO} + \mathcal{J}_{S3}^{GEO} \tag{11}$$

其中：

- **低地球轨道成本（推理 + 卸载）：** LEO节点需要为本地推理以及通过星间链路将训练数据集卸载至地球静止轨道而消耗能量：

$$\mathcal{J}_{S3}^{LEO} = N_{inf}(\omega^{inf} \cdot \epsilon_{LEO}) + P_{tx}^{ISL} \frac{\Delta^{train}}{R_{ISL}} + P_{rx}^{ISL} \frac{\sigma}{R_{ISL}} \tag{12}$$

- **GEO成本（训练 + 模型分发）：** 地球静止轨道节点必须消耗能量来接收海量数据集、执行繁重的反向传播，并将模型传输回来：

$$\mathcal{J}_{S3}^{GEO} = \underbrace{\Omega^{train} \cdot \epsilon_{GEO}}_{\text{Training Compute}} + \underbrace{P_{rx}^{ISL} \frac{\Delta^{train}}{R_{ISL}}}_{\text{Dataset Rx}} + \underbrace{P_{tx}^{ISL} \frac{\sigma}{R_{ISL}}}_{\text{Model Tx}} \tag{13}$$

尽管光学星间链路效率很高，项 $\Omega^{train} \cdot \epsilon_{GEO}$ 仍可能相当大。由于 GEO数字信号处理器 的每浮点运算效率通常低于 LEO边缘加速器（见表 I），S3 通常比 S2 更耗能。其主要优势在于时延，而非能量。

2) 生命周期时延（$\mathcal{T}_{S3}$）：S3 的主要优势是消除了轨道等待时间。时延由星间链路传播时延和 GEO处理速度决定：

$$\mathcal{T}_{S3} = \text{RTT}_{ISL} + \frac{\Delta^{train}}{R_{ISL}} + \frac{\Omega^{train}}{F_{GEO}} + N_{inf} \cdot T_{comp}^{LEO} \tag{14}$$

E. 运行可行性边界

为确定全局最优策略，我们将比较推广至三种架构，以识别各自的具体优势区域。我们分别定义能量（$\mathcal{J}$）和时延（$\mathcal{T}$）的可行性边界。

1) 全局能量可行性：能量最小化问题旨在求 $\min(\mathcal{J}_{S1}, \mathcal{J}_{S2}, \mathcal{J}_{S3})$。

a) 条件 *1*：*“边缘优势”*（*S2* 对比 *S1*）：当星载推理的成本低于传输原始遥测的成本时，场景 2（分拆式无线智能控制器）在能量方面优于场景 1（地面中心式）。忽略更新开销（假设 N_{inf} 较大），可得基本不等式：

$$\underbrace{\omega^{inf} \cdot \epsilon_{LEO}}_{\text{Compute Energy/Op}} < \underbrace{P_{tx}^{feed} \cdot \frac{\delta^{in}}{R_{ul}^{feed}}}_{\text{Transmission Energy/Op}} \tag{15}$$

对于复杂信号处理任务（δ^{in} 较高），该条件成立；但对于简单的标量报告（δ^{in} 较低），该条件不成立，此时 S1 仍是最优选择。

b) 条件 *2*：*“链路效率”*（*S3* 对比 *S2*）：我们比较通过 ISL 将训练数据卸载至地球静止轨道与通过射频将其卸载至地面的成本。只有当光学星间链路的效率足以抵消 GEO 计算所增加的成本时，场景 3 在能量方面才优于场景 2：

$$\underbrace{P_{tx}^{ISL} \frac{\Delta^{train}}{R_{ISL}}}_{\text{ISL Offload}} + \mathcal{J}_{S3}^{GEO} < \underbrace{P_{tx}^{feed} \frac{\Delta^{train}}{R_{ul}^{feed}}}_{\text{RF Offload}} \tag{16}$$

通常，$\mathcal{J}_{S3}^{GEO}$较高。因此，S3 很少是能量最小化方案；它是一种为降低时延而选择的高端架构，而代价则以焦耳计。

2）全局时延可行性：时延最小化问题旨在求 $\min(\mathcal{T}_{S1}, \mathcal{T}_{S2}, \mathcal{T}_{S3})$。

表III 仿真参数

参数	符号	数值
轨道与链路参数		
轨道高度	h_{LEO}	600 千米
馈电链路频率	f_c	28 GHz（Ka 波段）
馈电链路带宽	B_{feed}	400 MHz
馈电发射功率	P_{tx}^{feed}	15 W
馈电接收功率	P_{tx}^{feed}	5 W
星间链路发射/接收功率（光学）	$P_{tx}^{ISL}, P_{rx}^{ISL}$	2 W
星间链路数据速率	R_{ISL}	10 Gbit/s
轨道等待时间	T_{wait}	0 分钟至 60 分钟
人工智能工作负载配置		
输入数据大小（可变）	δ^{in}	10 kB 至 50 000kB
推理复杂度（可变）	ω^{inf}	0.1 每秒十亿次浮点运算至 500每秒十亿次浮点运算
训练数据集大小	Δ^{train}	10 Gbit
模型权重大小	σ	50 Mbit
模型寿命	N_{inf}	10^5 推理事件

a）条件*3*：*“连续性增益”*（*S3* 与 *S1/S2* 对比）：S1 和 S2 都需要等待轨道等待时间（T_{wait}^{gnd}）才能完成学习循环。S3 则消除了这段等待。因此，当等待时间超过 GEO 枢纽的处理开销时，S3 成为全局时延最小化方案：

$$T_{wait}^{gnd} > \underbrace{\left(\text{RTT}_{ISL} + \frac{\Omega^{train}}{F_{GEO}}\right)}_{\text{GEO Overhead}} - \underbrace{\left(\text{RTT}_{feed} + \frac{\Omega^{train}}{F_{gnd}}\right)}_{\text{Ground Overhead}} \tag{17}$$

由于T_{wait}^{gnd}处于分钟量级（例如，低地球轨道为 10 分钟至 45 分钟），而右侧处于秒级，因此对于任何需要快速模型再训练的时间关键型应用，S3 都具有严格优势。

VI. 性能评估与数值结果

本节对第五节推导的解析框架进行数值评估。我们开展多维敏感性分析，以确定各架构（S1、S2、S3）在物理层面具有优势的运行区域。所提出的分拆式无线智能控制器架构（即 S2 和 S3）的性能，以称为 S1 的最先进（SoA）基线方案为参照进行评估。该基线方案对应于传统系统中使用的“弯管式”或透明架构：卫星仅充当中继节点，所有控制智能均集中在地面核心中。我们通过改变四个关键参数来确定所提出架构的运行边界：数据量、信道质量、模型复杂度和环境稳定性。

A. 仿真设置

表 III 总结了仿真参数，这些参数对应于典型的 6G低地球轨道星座场景（例如，Starlink/OneWeb 轨道高度）：馈电链路工作在 Ka 波段，多层场景则使用光学星间链路 (OISL)。硬件配置与第 III 节中的定义一致。

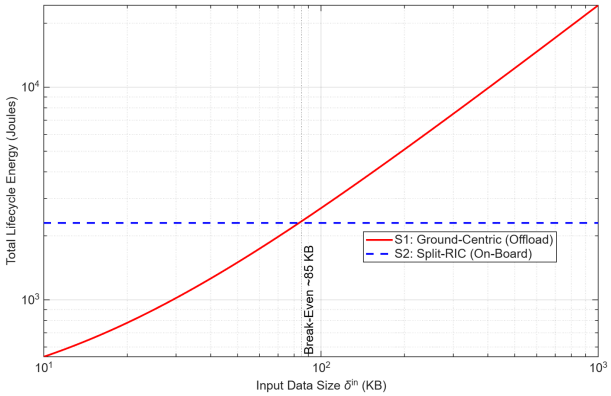


图 3. 能量可行区域：数据量的影响

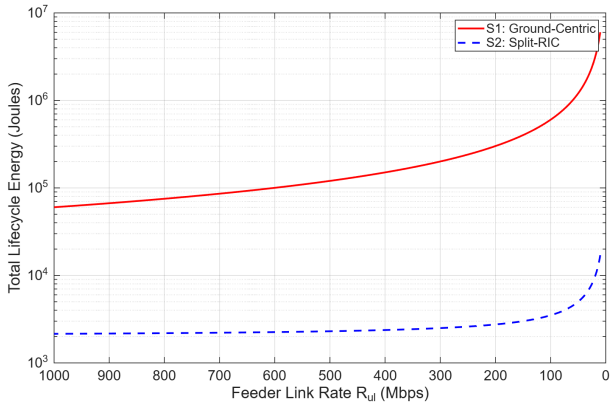


图 4. 信道质量的影响

B. 能量敏感性分析

1）通信约束：我们首先分析馈电链路的特性如何决定架构选择。图 3 比较了基线方案 (S1) 与分拆式无线智能控制器 (S2) 的生命周期总能耗随每次推理事件的遥测数据量变化的情况，并假设模型复杂度固定为 ($\omega^{inf} = 1$每秒十亿次浮点运算)。对于标量指标（例如缓冲区状态报告或简单的关键绩效指标计数器），S1（地面中心式）在能效方面更优。当传输成本可以忽略不计时，启用星载加速器所带来的开销并不划算。对于向量数据（例如波束成形矩阵或原始同相/正交频谱快照），S1 的能耗会急剧线性增长。这是因为卸载大量原始数据需要较高的发射功率 ($P_{tx}^{feed} = 15$ W)。两者的交点约为 $\delta^{in} \approx 85$kB。对于任何需要原始频谱数据（通常为 > 10 MB）的 xApp，分体式 RIC 架构 (S2) 可将总能耗降低 90% 以上，验证了“计算与通信之间的权衡”所带来的收益。

图 4 探讨了一个关键的实际场景：雨衰或低仰角导致馈电链路速率下降（R_{ul}）。我们假设频谱图 xApp 应用的数据量为 5MB。S1 的能耗成本与数据速率（$E \propto 1/R_{ul}$）成反比。当信道劣化（例如，速率从 1 Gbit/s 降至 50 Mbit/s）时，传输时间延长，导致能耗急剧增加。相比之下，分体式 RIC 架构（S2）受信道劣化的影响较小，证实了它是高纬度地区或受天气影响地区的稳健选择。

2) 计算约束：接下来，我们分析星载硬件的限制。图 5 考察星上人工智能的上限。我们将数据大小固定为 $\delta^{in} = 5$MB，并改变神经网络的计算复杂度。S1 的能耗成本保持不变。由于处理在地面进行（相对于电池而言，地面功率可视为无限或免费），增加模型深度不会影响卫星的能量预算。S2 的能耗成本随浮点运算次数线性增长。在约 $\omega^{inf} \approx 250$GFLOPS 时，星载计算成本超过传输成本。这就界定了物理可行性边界：如果某个特定 xApp 需要复杂度超过这一水平的大型 Transformer 模型，则无论数据传输代价如何，都必须将其卸载到地面（S1）或强大的地球静止轨道枢纽（S3）。

图 6 分析了每次推理的摊销能耗。如果

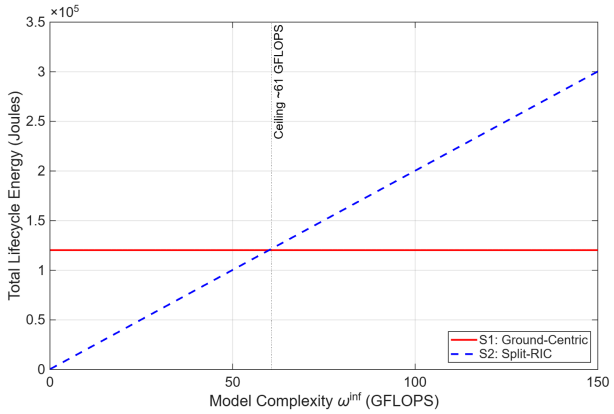


图 5. 能量可行区域：复杂度的影响

环境不稳定，模型需要每隔几分钟重新训练一次（低 N_{inf}），那么接收模型更新（通过 A1 接口）所产生的开销会使 S2 效率低下。S2 仅适用于环境稳定的情况，在这种环境下，模型在 > 1000 次推理事件中保持有效。

C. 时延可行性分析

虽然S2节省了能量，但也带来了时延方面的挑战。我们分析了这些架构的时间关键性。图7绘制了总生命周期时延（$\mathcal{T}$）与

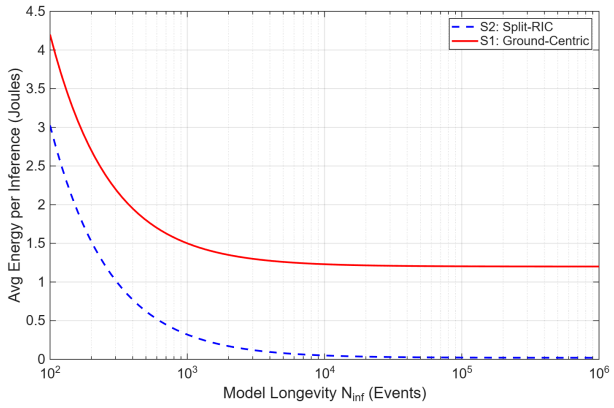


图 6. 概念漂移的影响

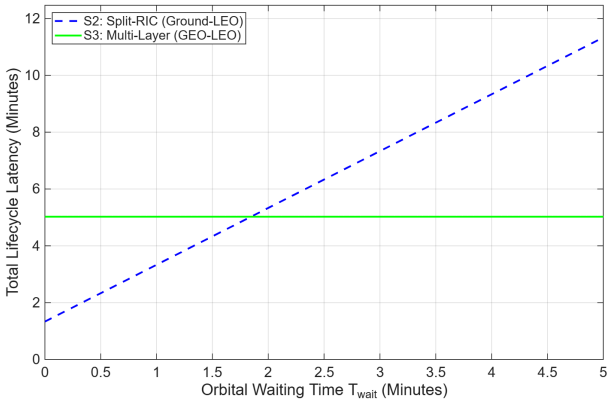


图7. 时延可行性：间歇性的影响

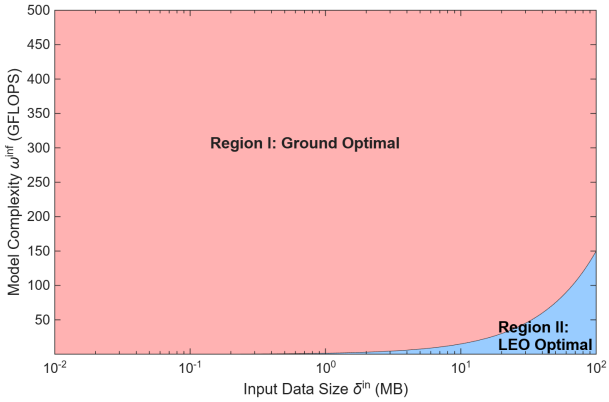


图8. 能量可行性图

平均轨道等待时间（T_{wait}）之间的关系，后者取决于地面站网络的密度。地面站-LEO分离架构的时延主要由存储转发时延（$2 \times T_{wait}$）决定。在稀疏部署中，当 $T_{wait} \approx 45$min时，模型更新周期对于应对动态威胁（例如干扰自适应）而言过于缓慢。无论低地球轨道卫星的位置如何，多层 GEO-LEO 架构都能保持恒定且较低的时延（$\mathcal{T}_{S3} \approx 2$s）。因此，S3是实现时间关键型自适应的必要架构。S2仅适用于静态优化 任务（例如每日流量模式学习），这类任务可以接受更新延迟。

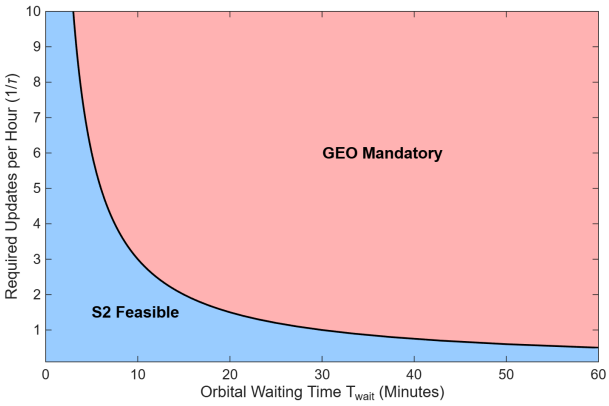


图9. 时延可行性图

D. 多维可行性图

最后，我们将这些权衡综合为两张统一的决策图，以指导网络编排。能量图（图8）是一张等高线图，描绘了地面最优（区域I）与低地球轨道最优（区域II）之间的边界。网络运营商可以将 xApp 映射到该平面上：流量预测（低数据量、高计算量）位于区域I，而波束管理（高数据量、低计算量）位于区域II。时延图（图9）凸显了场景3的必要性。随着所需更新频率提高（Y轴）且地面站密度降低（X轴，等待时间更长），系统会进入暗区，在该区域中，S2 违反时延期限。在此区域，多层地球静止轨道-低地球轨道架构（S3）不仅是一种选择，更是物理上的必然要求。

我们可以将这些结果归纳为一套统一的指导规则。

1) **区域I（地面最优）：** 数据量低。涉及简单报告的任务应使用传统O1接口（S1）。2) **区域II（LEO最优）：** 数据量高 +复杂度中等。涉及原始信号处理（例如波束管理）的任务应在LEO边缘（S2）执行，以最大限度提高能效。3) **区域III（GEO最优）：** 敏捷性要求高。需要立即进行模型再训练的任务必须利用GEO星间链路骨干（S3），以更高的能耗成本换取速度。

E. 模型验证与参数论证

为确保推导出的可行性区域符合物理现实，第VI-A节至第VI-D节所用的仿真参数均以商用现货（COTS）航天硬件和3GPP NTN标准的基准测试数据为依据。

1) 计算效率（ϵ_{LEO}，ϵ_{GEO}）：星上处理效率 $\epsilon_{LEO} = 20$ pJ/浮点运算，符合面向新航天任务的现代片上系统（SoC）加速器的能耗特征，例如 Xilinx Versal AI Core 或 NVIDIA Jetson AGX Orin（以低功耗模式运行）。虽然地面 GPU 的单位比特能耗较低，但抗辐射FPGA通常会带来功率损耗。我们的数值符合 7 纳米节点工艺下 INT8推理的效率，同时满足 6U立方星典型的热节流限制。相比之下，地球静止轨道较高的能耗（$\epsilon_{GEO} = 100$pJ/浮点运算）反映了传统透明转发总线采用较老的高可靠性 DSP（如抗辐射 PowerPC或 LEON 处理器），这类处理器优先考虑抗辐射能力，而非原始浮点运算次数/W。

2) 通信链路预算：馈电链路参数（$R_{ul} = 500$Mbit/s，$P_{tx} = 15$W）依据符合 ETSI 和 3GPP TR 38.821 指南的标准 Ka 波段高通量卫星（HTS）终端链路预算计算得出，该终端的孔径为 30 cm。光学星间链路速率（$R_{ISL} = 10$Gbit/s）相较于 100 Gbit/s

新型激光终端的能力（例如 SDA 传输层标准）较为保守，以确保我们针对场景3（多层）的结果不会高估轨道网状网络的优势。

3) 人工智能工作负载分析：模型复杂度（$\omega^{inf} \in [0.1, 500]$ 每秒十亿次浮点运算）和模型大小（$\sigma \approx$50 Mbit）的扫描范围涵盖了 6G xApp 的多样性。下限代表轻量级时间序列预测模型（例如用于流量预测的长短期记忆网络），而上限（≈250每秒十亿次浮点运算）对应复杂的视觉 Transformer （ ViT ）或大规模 MIMO 波束成形智能体。图 5 中确定的硬件上限实际上充当了验证边界；它证实，在当前低地球轨道能源预算下，部署繁重的地面模型（例如 GPT 级模块）在物理上是不可能的，因此必须采用本文提出的卸载策略。

VII. 结论

本文分析了将开放式无线接入网控制层级分布在地面、低地球轨道和地球静止轨道分段中的可行性，以克服非地面网络的尺寸、重量和功耗与带宽之间的约束。我们的多维敏感性分析表明，不存在一种普遍最优的架构。我们证明，对于数据密集型 xApp（输入$>$ 85kB），只要模型复杂度低于 $\approx$ 250每秒十亿次浮点运算的硬件上限，所提出的分体式 RIC 架构（S2）即可将生命周期能量降低 90 % 以上。然而，对于更新紧迫性超过轨道重访率的时间关键型应用，多层GEO-LEO架构（S3）是弥合连接空缺的必要解决方案。我们得出结论，6G 非地面网络需要一种灵活的分体式 *RIC* 策略，根据瞬时物理约束在地面、边缘和地球静止轨道层之间动态迁移功能。未来的工作将把这一可行性框架扩展为实时编排协议，利用优化技术，在时变流量和能量收集条件下自主确定功能部署。

参考文献

[1] G. Araniti、A. Iera、S. Pizzi 和 F. Rinaldi，《迈向6G非地面网络》，*IEEE Netw.*，第36卷，第1期，第113–120页，2022年。[2] D. Naseh、S. S. Shinde 和 D. Tarchi，《面向6G非地面网络场景中车联网应用的网络切片分布式学习即服务》，*Journal of Sensor and Actuator Networks*，第13卷，第1期，2024年。[3] S. S. Shinde 和 D. Tarchi，《用于多层多服务非地面车载边缘计算的分层强化学习》，*IEEE Trans. Mach. Learn. Commun. Netw.*，第2卷，第1045–1061页，2024年。[4] J. Wigard、E. Juan、J. Stanczak、M. Lauridsen、A. Marcone、S. Hoppe、A. Ahmadzadeh、A. Masri 和 D.-H. Tran，《通过非地面网络实现泛在6G服务》，*IEEE Wireless Commun.*，第30卷，第6期，第12–18页，2023年。[5] S. Ammar、C. Pong Lau 和 B. Shihada，《关于6G地面与非地面融合网络虚拟化技术的深入综述》，*IEEE Open J. Commun. Soc.*，第5卷，第3690–3734页，2024年。[6] M. Polese、L. Bonati、S. D'Oro、S. Basagni 和 T. Melodia，《理解 O-RAN：架构、接口、算法、安全与研究挑战》，*IEEE Commun. Surveys Tuts.*，第25卷，第2期，第1376–1411页，2023年。[7] S. S. Shinde、D. Naseh 和 D. Tarchi，《6G地面与非地面融合网络中基于机器学习的智能 O-RAN 控制》，载于*2024*年*IEEE Globecom Workshops (GC Wkshps)*，2024年，第1–6页。[8] R. Campana、C. Amatetti 和 A. Vanelli-Coralli，《基于 O-RAN 的非地面网络：趋势与挑战》，载于*2023*年欧洲网络与通信联合会议暨*6G*峰会 *(Eu- CNC/6G Summit)*，2023年，第264–269页。[9] M. Rihan、T. D¨ue、M. Vakilifard、D. W¨ubben 和 A. Dekorsy，《面向地面与非地面融合6G网络的 RAN功能拆分选项》，载于*2023*年第*11*届日本-非洲电子、通信与计算国际会议 *(JAC-ECC)*，2023年，第152–158页。[10] O-RAN工作组 2，《O-RAN人工智能/机器学习工作流程说明与要求 1.03》，O-RAN联盟技术报告，2021年10月。 [在线]。可访问：https://specifications.o-ran.org/download?id=158 [11] F. Firouzi、N. Bleier、B. Farahani、T. Seyfi 和 F. Afghah，《O^2-RAN：面向非地面网络与基于空间的边缘计算的轨道开放无线接入网》，载于*2025*年*IEEE*全层智能系统国际会议 *(COINS)*，2025年，第1–9页。[12] S. Mahboob、J. Dai、A. Soysal 和 L. Liu，《通过 O-RAN赋能的非地面网络变革未来6G网络》，*IEEE Commun. Mag.*，第63卷，第3期，第76–82页，2025年。[13] C. T. Nguyen、Y. M. Saputra、N. V. Huynh、T. N. Nguyen、D. T. Hoang、D. N. Nguyen、V.-Q. Pham、M. Voznak、S. Chatzinotas 和 D.-H. Tran，《6G非地面网络新兴技术：从学术研究到工业应用》，*IEEE Open J. Commun. Soc.*，第5卷，第3852–3885页，2024年。[14] M. M. H. Qazzaz、L. Kułacz、A. Kliks、S. A. Zaidi、M. Dryjanski 和 D. McLernon，《用于 O-RAN网络动态资源分配的基于机器学习的 xApp》，载于*2024*年*IEEE*通信与网络机器学习国际会议 *(ICMLCN)*，2024年，第492–497页。[15] A. Bhattacharyya、S. M. Nambiar、R. Ojha、A. Gyaneshwar、U. Chadha 和 K. Srinivasan，《机器学习与深度学习赋能的卫星通信：使能技术、应用、开放挑战与未来研究方向》，《国际卫星通信与网络期刊》，第41卷，588，2023 6，539no. pp. – 。[16] L. Gong、Q. Chen、L. Yang、Z. Yin 和 Y. Wang，《基于卫星时空特征映射的低地球轨道卫星物联网自主流量预测》，*IEEE Internet Things J.*，第12卷，第14期，第27 021–27 032页，2025年。[17] M. Du、K. Wang、Z. Xia 和 Y. Zhang，《边缘计算环境下无线大数据训练模型的差分隐私保护》，*IEEE Trans. Big Data*，第6卷，第2期，第283–295页，2020年。[18] S. Yao、Y. Lin、M. Wang、K. Xu、M. Xu、C. Xu 和 H. Zhang，《LEOEdge：面向大型低地球轨道星座人工智能推理的星地协作平台》，*IEEE J.Sel. Areas Commun.*，第43卷，第1期，第36–50页，2025年。



DANIELE TARCHI （IEEE高级会员）1975年出生于意大利佛罗伦萨。2000年和2004年，他分别获得佛罗伦萨大学（意大利佛罗伦萨）的电信工程理学硕士学位和信息学与电信工程博士学位。

2004年至2010年，他在意大利佛罗伦萨大学担任研究副职。2010年至2019年，他担任助理教授；2019年至2024年，他在意大利博洛尼亚大学担任副教授。自2024年起，他一直担任意大利佛罗伦萨大学副教授。他已发表170篇论文，刊载于国际期刊和会议论文集。他的研究主要聚焦于无线通信与网络、卫星通信与网络、边缘计算、分布式学习和优化技术。他参与了众多国内外研究项目，并与国外多家研究机构开展合作。

Tarchi教授自2012年起成为IEEE高级会员。他担任IEEE车辆技术汇刊、IEEE通信学会开放期刊和IET通信的编委会成员。他曾担任IEEE WCNC 2011、IEEE Globecom 2014、IEEE Globecom 2018和IEEE ICC 2020的专题研讨会共同主席，并担任IEEE ICC 2015的研讨会共同主席。