用于移动边缘计算低时延多用户虚拟现实传输的太赫兹可重构全息表面收发器

Liangshun Wu、Wen Chen、Honghao Wang、Zhendong Li、Ying Wang

arXiv:2603.01888v1 [cs.IT] 2 Mar 2026

摘要—**本文研究了一种由太赫兹（THz）赋能、移动边缘计算（MEC）辅助的虚拟现实（VR）系统，该系统采用可重构全息表面（RHS）作为收发器，实现多用户波束成形和全息图案分多址（HDMA）。我们针对三维视场（FoV）生成流程建立端到端模型，并在内存和功率约束下优化内容预取与渲染卸载，同时通过调整全息图案权重进行波束赋形、通过馈源功率分配调整激励幅度，以实现适应用户移动的波束成形。对于同质视场，我们推导出预取二维或三维视场，或直接传输三维视场的闭式策略。对于异质视场，我们利用预取／渲染与快速RHS波束成形之间的时间尺度分离，将优化问题分解为渲染—预取组合优化问题和短时间尺度的波束成形凸优化问题。仿真结果表明，即使资源约束严格，也能显著降低时延。**

索引词—**太赫兹、可重构全息表面、移动边缘计算、虚拟现实、低时延**

I. 引言

IMmersive 360° 虚拟现实（VR）被视为一项关键的 6G 应用，可为娱乐、教育及其他领域带来身临其境的体验 [1]。然而，无线虚拟现实面临三大关键挑战：（i）数据速率需求极高，4K 360° 视频需要 70–300 Mbps [2], ，远超 Sub-6 GHz [3]所能提供的能力；（ii）超低时延，运动到光子时延需低于 20 ms（即 0.02 s）[4]–[6]，因此在渲染/解码后仅剩数毫秒用于传输；（iii）头戴式显示器（HMD）资源有限 [7]。这些瓶颈使得在当今网络上实现高质量无线虚拟现实并不切实际。

为解决这一问题，太赫兹（THz）通信正成为一种新兴的赋能技术，可提供多 Gbps 链路（0.1–10 THz） [8], [9]，适用于实时全景流传输、6G 虚拟现实及全息应用 [10]–[13]。

A. 基于移动边缘计算的虚拟现实传输

我们首先分析一种典型的 360° 虚拟现实视频制作流程（图1） [10]: i) 拼接：将多摄像机阵列拍摄的视频拼接起来，生成球面视频；ii)等距柱状投影：将球面视频展开为二维视频，并编码为图像组（GOP）由 I 帧、P 帧和 B帧组成。对于高分辨率帧，每帧会拆分为多个分块，每个分块再进一步划分为 $4 \times 6 = 24$ 图块。对于低分辨率帧，每帧则直接划分为 24 个图块。iii) 提取：根据用户的注视方向确定当前视场，并按照头戴式显示器跟踪到的视场，从二维视频中提取相应的二维视场；iv) 投影：将二维视场投影为三维视场；v)渲染：在头戴式显示器上渲染生成的三维视场。i) 拼接、ii) 等距柱状投影和 iii) 提取 由基于云的内容提供商执行，因为这些操作需要完整的 360° 视频作为输入。将这些任务卸载出去，可减轻移动边缘计算服务器（MECS）和头戴式显示器的渲染负担，并减少无线流量。内容提供商通过高容量光纤链路连接到 MECS 和太赫兹（THz）基站（BS）。基站向多名配备头戴式显示器的用户传输二维视场，而头戴式显示器执行 v)投影 和 v) 渲染 以显示沉浸式内容。

该系统利用移动边缘计算（MEC） [14]–[15]将渲染和解码任务卸载到头戴式显示器（HMD）之外，从而降低设备能耗和时延 [16], [17]。为进一步提高效率，系统在移动边缘计算服务器（MECS）预取二维和三维视场（FoV），其中三维视场由低复杂度投影模块离线生成。由于用户的视场请求呈现出明显的流行度模式，因此可以利用概率模型优化预取决策[18]。随着头戴式显示器的计算能力不断增强，投影模块可以迁移至头戴式显示器，使其能够根据预取的二维视场在本地生成三维视场，并将无线流量最多减少一半，因为立体渲染需要为双眼分别复制三维视场 [19]。然而，这种迁移会增加渲染时延，因此需要谨慎制定预取和计算策略。头戴式显示器内存可以存储二维或三维视场；三维预取能够降低时延和能耗，但会占用更多内存，因此如何权衡通信、计算和存储成为一项关键设计挑战。大多数现有的基于移动边缘计算的虚拟现实研究都过度简化了渲染过程，且缺乏严格的流水线级建模 [16], [17]–[20]，而头戴式显示器侧预取方法往往忽视虚拟现实的关键特性，或只关注二维视场 [21]–[22]。相比之下，本文对完整流程进行建模，并联合优化二维和三维视场预取与渲染。

在图 1 所示的系统中，我们采用可重构全息表面（RHS）作为基站收发器，以实现精确的波束成形，从而高效传输二维和三维视场

Liangshun Wu、Wen Chen 和 Honghao Wang 任职于中国上海的上海交通大学电子工程系（电子邮箱：wuliangshun@sjtu.edu.cn；wenchen@sjtu.edu.cn，mc25018@um.edu.mo）。Zhendong Li 任职于中国西安的西安交通大学信息与通信工程学院（电子邮箱：lizhendong@xjtu.edu.cn）。Ying Wang 任职于中国北京的北京邮电大学网络与交换技术国家重点实验室（电子邮箱：wangying@bupt.edu.cn）。

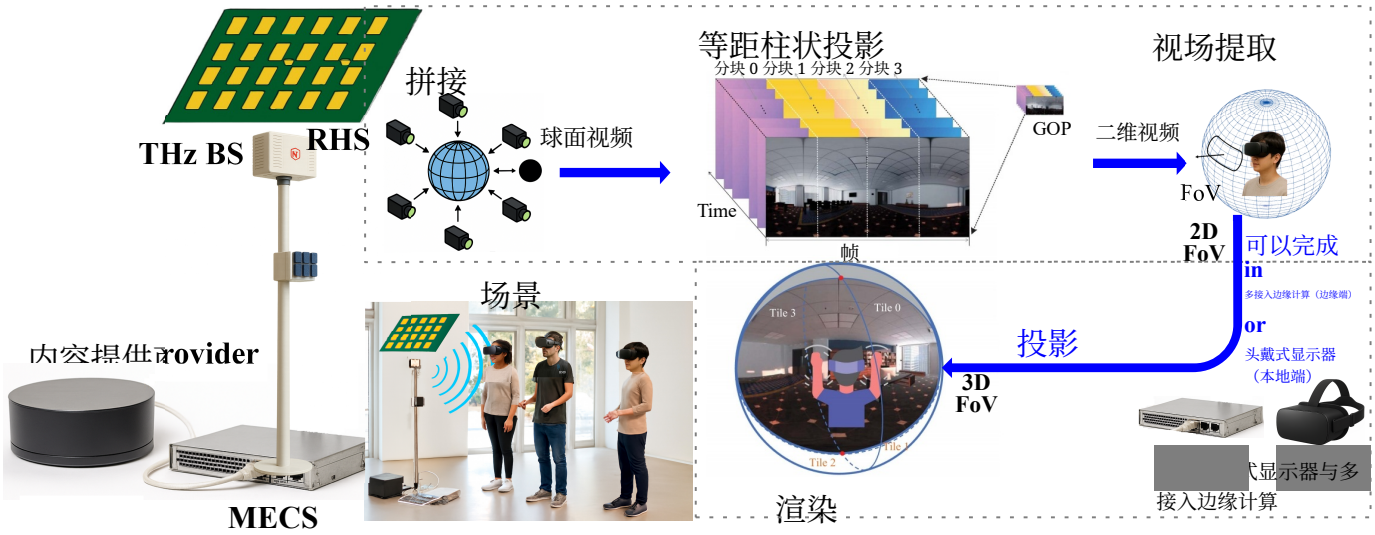


图1：典型的 360° 虚拟现实 采用太赫兹可重构全息表面收发器进行无线传输并结合移动边缘计算的视频制作流程。

从移动边缘计算服务器向多个头戴式显示器传输内容，并在用户移动时确保最佳覆盖。太赫兹技术可将二维视场从移动边缘计算服务器高速传输至头戴式显示器，实现沉浸式虚拟现实体验所必需的低时延、高带宽通信。

B. 可重构全息表面

可重构全息表面采用谐振特性经过精心设计的亚波长单元，无需针对不同频率设置独立结构，即可在数百 GHz 的频率范围内实现稳定的宽带全息调制。可重构全息表面通过空间相位分布控制波前，确保在宽带范围内实现一致的波前控制 [23]。借助电控单元（例如 PIN二极管、变容二极管），全息图案可在微秒内切换，从而实现实时波束转向和多用户成形 [24], ，非常适合动态沉浸式虚拟现实。采用超薄印制电路板制造工艺，且无需独立移相器（PA），因此硬件成本和功耗较低，便于构建包含数百或数千个元素的大规模阵列。此外，全息图案分割多址（HDMA） [25], [26] 利用与不同波束相关的全息图案叠加，为多个头戴式显示器提供服务。目前，可重构全息表面的研究主要关注硬件和波束成形，而其与移动边缘计算及太赫兹虚拟现实传输的集成仍未得到探索。

C. 主要贡献

本文的主要贡献如下：

1) 太赫兹与可重构全息表面辅助的虚拟现实传输架构：我们提出一种基于可重构全息表面（全息超表面）和太赫兹技术的虚拟现实传输架构，通过精确的波束控制和可电子转向的多波束支持多用户系统。该方案利用太赫兹的超高带宽和可重构全息表面的快速波束成形，以适应沉浸式虚拟现实环境中的多用户移动性，实现高效的虚拟现实数据传输。我们还考虑了互耦效应，该效应反映了可重构全息表面阵列中相邻辐射单元之间的相互作用。

2）基于移动边缘计算的虚拟现实框架及问题建模：我们设计了一种基于移动边缘计算的虚拟现实框架，将渲染任务卸载至移动边缘计算—头戴式显示器边缘，从而减少无线流量和时延。我们构建并求解联合预取-渲染-波束成形优化问题，在约束条件下最小化传输时延，并通过优化内容预取、渲染卸载以及结合全息图案权重调整和馈电功率分配的波束重新成形，使其适应用户移动性。3）面向联合预取-渲染优化的二元—连续混合优化与移动性感知可重构全息表面波束成形：该问题涉及一个具有挑战性的联合二元—连续优化问题。我们首先推导出以下情形的闭式解：预取并本地渲染二维视场、预取三维视场，或直接传输远端三维视场。我们将该框架扩展至异构视场设置，并考虑头戴式显示器移动性和快速波束重新转向。通过将问题分解为长短时间尺度子问题，我们结合基于多选择多维背包问题（MMKP）的启发式算法、二元松弛差分凸算法（DCA）和凸-凹过程（CCCP）、完整 DCA/CCCP 求解器，以及基于投影梯度（PG）的可重构全息表面波束成形优化器，高效求解该优化问题。

接下来，第二节介绍系统模型；第三节和第四节研究同质和异质情况；第五节给出仿真结果；第六节总结全文。

II. 系统模型

A. 太赫兹*RHS*信道模型

渲染卸载涉及 MECS 与 *L*HMDs 之间的通信。MECS 配备了基站和可重构全息表面，以支持高度定向且自适应的太赫兹波束成形。

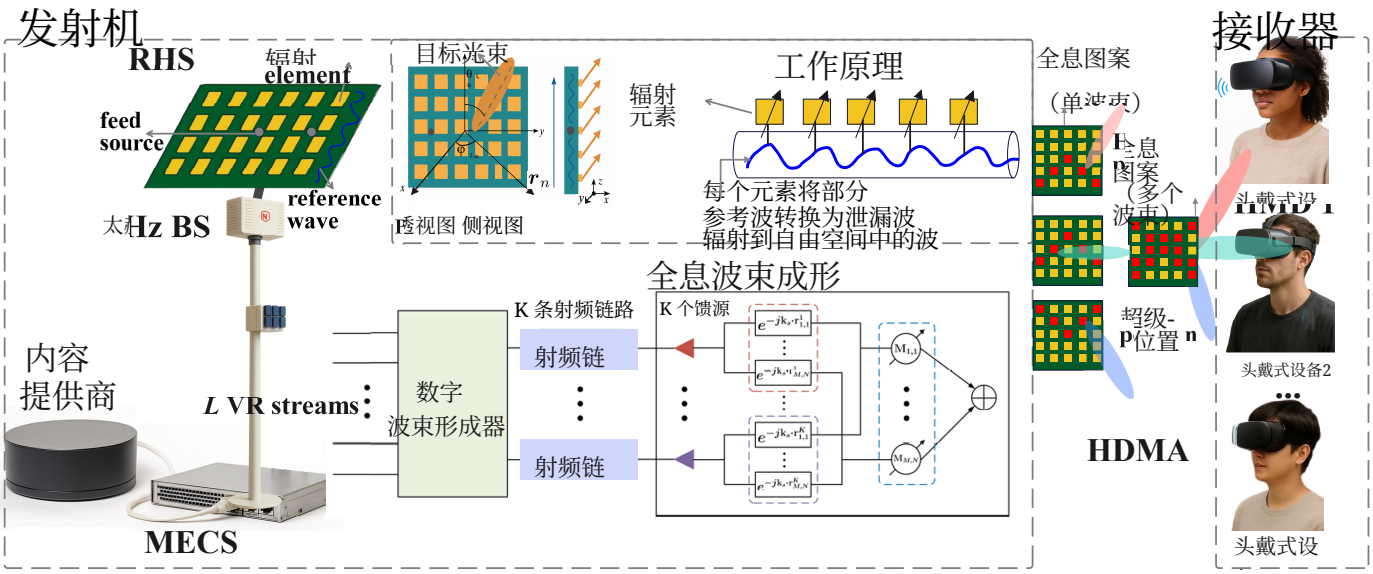


图2：RHS辅助的太赫兹多用户虚拟现实传输。

1）频率选择性太赫兹信道：由于载波频率极高，太赫兹信道表现出强烈的频率选择性、分子吸收和严重的扩散损耗。按照惯例，载波频率为$f_c = 0.3\text{THz} = 300\text{GHz}$，因此$\lambda_c = c/f_c = 1$ mm，其中$c = 3 \times 10^8$ m/s为光速。总带宽$B_c = 40\text{GHz}$被划分为$U = 4$个子带，每个子带带宽为$B_g = 10$ GHz，小于信道相干带宽。因此，每个子带都可以建模为窄带平坦衰落信道。令f_u表示第u个子带的中心频率，$f_u = f_c - \frac{B_c}{2} + \left(u - \frac{1}{2}\right)\frac{B_c}{U}$，即$f_u \in \{285, 295, 305, 315\}$ GHz。在子带窄带假设下，基站与第l个头戴式显示器（$l \in \mathcal{L}$，$|\mathcal{L}| = L$）之间的频率选择性太赫兹基带信道系数在子带u上表示为$h_{l,u} = \frac{c}{4\pi f_u d_l}\left(-\kappa\left(f_u\right)\frac{d_l}{2}\right)$，exp ( )，其中 ( ) 表示频率相关的分子吸收系数 [27]。d_l表示可重构全息表面与头戴式显示器l之间的欧几里得距离：$d_l = \|\mathbf{p}_l - \mathbf{p}_R\| = \sqrt{\Delta x_l^2 + \Delta y_l^2 + \Delta z_l^2}$，其中$\mathbf{p}_R = [x_R, y_R, z_R]^T$为可重构全息表面的位置，$\mathbf{p}_l = [x_l, y_l, z_l]^T$. 为头戴式显示器$l$的位置。$\Delta x_l = x_l - x_R$，$\Delta y_l = y_l - y_R$，$\Delta z_l = z_l - z_R$.

2）带间干扰（*IBI*）：尽管每个子带都是窄带的，但相邻子带间的频谱泄漏会导致带间干扰（IBI）。根据 [27], ，子带u上的带间干扰（IBI）可近似为一个方差为

$$I_u \sim \mathcal{N}\left(0, \int_{f_u} \sum_{v \neq u}^{U} P_v \left| G_v(f_u) \sum_{q \in \mathcal{Q}} \alpha_{v,q} \right|^2 df_u\right), \quad (1)$$

其中$\alpha_{v,q}$为泄漏系数，P_v为分配给子带v的功率，$G_v(\cdot)$为波形频谱。

3）基于*HDMA*的*RHS*全息波束成形：RHS是一种特殊的漏波天线，由馈源和大量超表面辐射单元组成。如图2所示，每个馈源都放置在RHS的边缘或背板上，并将输入信号转换为电磁参考波。RHS采用串馈架构：参考波在表面传播，并依次激励各辐射单元。每个单元将部分参考波转换为辐射到自由空间中的漏波，其中每个单元的辐射振幅均可独立调节，以实现全息波束成形。与传统的并馈相控阵相比，串馈RHS架构大幅简化了大规模MIMO部署的布线。

为说明全息原理，考虑一个具有K个馈源和$N = N_x \times N_y$个辐射单元的可重构全息表面（RHS），$N_x = N_y = \sqrt{N}$，其排列为具有$d = \lambda_c/2$单元间距的$\sqrt{N} \times \sqrt{N}$均匀平面阵列（UPA）。令$\Psi_{l,u}$表示子带u中朝向HMD l方向$(\theta_{l,u}, \varphi_{l,u})$辐射的期望波束，$\Psi_{\text{ref}}$表示由馈源$k$生成的参考波。在第$n$个辐射单元处，它们可分别写为$\Psi_{l,u} = e^{-j\boldsymbol{k}_f(\theta_{l,u}, \varphi_{l,u})\cdot \boldsymbol{r}_n}$，$\Psi^{\text{ref}}(\boldsymbol{r}_n^k) = e^{-j\boldsymbol{k}_s \cdot \boldsymbol{r}_n^k}$，其中，$\boldsymbol{k}_f(\theta_{l,u}, \varphi_{l,u})$为自由空间波矢，$\boldsymbol{r}_n$为第$n$个单元的位置，$\boldsymbol{k}_s$为参考波的传播矢量，$\boldsymbol{r}_n^k$为从馈源$k$到单元$n$的位移矢量。根据全息干涉，记录的全息图是干涉图样$\Psi_l^{\text{intf}}$、$_u = \Psi_{l,u}\Psi^{\text{ref}}$。当再次由参考波照射时，产生的漏波场满足$\Psi_l^{\text{intf}}$、$_u\Psi^{\text{ref}} \propto \Psi_{l,u}\left|\Psi^{\text{ref}}\right|^2$。从RHS到HMD l的视距路径出发方向可用球坐标中的仰角$\theta_{l,u}$和方位角$\varphi_{l,u}$参数化为$\varphi_{l,u} = \arctan 2(\Delta y_l, \Delta x_l)$、$\theta_{l,u} = \arctan 2\left(\Delta z_l, \sqrt{\Delta x_l^2 + \Delta y_l^2}\right)$；这些角度与子带$u$无关，完全由几何结构决定。

根据全息干涉原理，将对应的干涉场叠加，可得到HDMA中用于所有头戴式显示器的子带u复合全息图案L：$m_{u,n} = \sum_{l=1}^{L}\sum_{k=1}^{K} a_{l,u,k}^{\frac{\text{Re}\left[\Psi_{l,u}^{\text{intf}}\right]+1}{2}}$，其中$a_{l,u,k} \geq 0$表示从馈源指向用户$l$的波束幅度比，来自馈源$k$，权重满足$\sum_{l=1}^{L}\sum_{k=1}^{K} a_{l,u,k} =$

全息波束成形矩阵 $\boldsymbol{M}_u \in \mathbb{C}^{N\times K}$ $\boldsymbol{M}_{un}^{k} = \sqrt{\beta m_{u\,n}} e^{-\alpha\|\boldsymbol{r}_n^k\|} e^{-j\boldsymbol{k}_s \boldsymbol{r}_n^k}$ 中，来自馈源 k 的第 n 个辐射单元定义为，，其中 表示单元辐射效率，α 表示结构衰减系数。

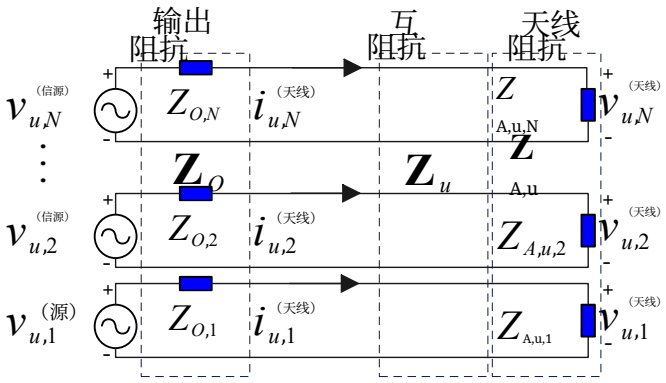


图 3：可重构全息表面发射端（Tx）互耦效应的等效多端口电路模型。

由于辐射单元之间的间距更小，因此会产生互耦，如 [28], [29] 中所述。发射机处可重构全息表面的等效多端口电路模型如图 3 [29] 所示。每个可重构全息表面元素均由一个输出阻抗为 Z_O 的电压源激励，由此产生的天线电流流经频率相关互阻抗网络 $\boldsymbol{Z}_u$，其中对角元素和非对角元素分别表示自阻抗和单元间互耦。因而，天线端电压 $\mathbf{v}_u^{(\mathrm{ant})}$由源激励、输出阻抗和互阻抗网络共同决定。天线阵元上的电压记为 $\mathbf{v}_u^{(\mathrm{ant})}$，其表示为 $\mathbf{v}_u^{(\mathrm{ant})} = \mathbf{Z}_u \mathbf{i}_u^{(\mathrm{ant})}$，其中 $\mathbf{Z}_u$ 表示频率 f_u处的频率相关互阻抗矩阵。互阻抗矩阵 $\mathbf{Z}_u \in \mathbb{C}^{N\times N}$ 按元素构造为 $[\mathbf{Z}_u]_{p,q} = Z_{A,u}$，适用于 $p = q$；而对于 $p \neq q$，其由相距 d_p 的两个电短偶极子之间的互阻抗给出，

$$= \text{表} \quad q \quad = -30\{\int_0^{l_d} \left(\frac{2\pi}{\lambda_u} z\right) + \int_{l_d}^{2l_d} \left[\frac{2\pi}{\lambda_u}\ 2l_d - z\right]\}$$

示为 $[\mathbf{Z}_u]_p$，sin sin ( ) (

$$\frac{-\mathrm{j}e^{-\mathrm{j}\frac{2\pi}{\lambda_u} r_1}}{r_1} + \mathrm{j}$$

$$\frac{e^{-\mathrm{j}\frac{2\pi}{\lambda_u} r_2}}{r_2} + 2\mathrm{j}\cos\left(\frac{2\pi}{\lambda_u} l_d\right) \frac{e^{-\mathrm{j}\frac{2\pi}{\lambda_u} r_0}}{r_0})\mathrm{d}z,$$

其中 $r_0 = \sqrt{d_{p,q}^2 + z^2}$，$r_1 = \sqrt{d_{p,q}^2 + (l_d - z)^2}$，$r_2 = \sqrt{d_{p,q}^2 + (l_d + z)^2}$，，且在 $N = N_x \times N_y$ 平面阵列索引下，单元间距为 $d_{p,q} = d\sqrt{(i_p - i_q)^2 + (j_p - j_q)^2}$；$l_d = \lambda_c/4$为电短偶极天线的长度。此处，$z$ 是表示沿天线长度方向位置的变量，用于计算天线不同位置处的距离。

根据基尔霍夫定律，存在互耦时，电压源产生的电压为 $\mathbf{v}_u^{(\mathrm{src})} = \mathbf{Z}_O \mathbf{Z}_u^{-1} \mathbf{v}_u^{(\mathrm{ant})} + \mathbf{v}_u^{(\mathrm{ant})}$，其中 $\mathbf{Z}_O = Z_O \mathbf{I}_N$ 表示电压源的输出阻抗矩阵，对于射频功率放大器，该值通常为$Z_O = 50\Omega$。不存在互耦时，互阻抗矩阵退化为天线阻抗矩阵，即 $\mathbf{Z}_u \xrightarrow{\text{无互耦}} \mathbf{Z}_{A,u} = Z_{A,u}\mathbf{I}_N$。由于可重构全息表面采用电短偶极子 ($l_d \ll \lambda_u$)，$Z_A$，$u$ 遵循[28]中的标准近似：Z_A，$_u \approx 80\left(\frac{l_d}{\lambda_u}\right)^2 - \mathrm{j}\frac{120}{\pi}\ln\left(\frac{\lambda_u}{2\pi l_d}\right)$。不存在互耦时，电压源产生的电压可表示为：$\mathbf{v}_u^{(\mathrm{src})} = \mathbf{Z}_O\mathbf{Z}_A^{-1}$，$_u\mathbf{v}_u^{(\mathrm{ant,nomc})} + \mathbf{v}_u^{(\mathrm{ant,nomc})}$，其中 $\mathbf{v}_u^{(\mathrm{ant,nomc})}$表示不存在互耦时天线阵元上的电压。因此，我们有 $\mathbf{Z}_O\mathbf{Z}_{A\,u}^{-1}$，$\mathbf{v}_u^{(\mathrm{ant,nomc})} + \mathbf{v}_u^{(\mathrm{ant,nomc})} = \mathbf{Z}_O\mathbf{Z}_u^{-1}\mathbf{v}_u^{(\mathrm{ant})} + \mathbf{v}_u^{(\mathrm{ant})}$，也就是说 $\left(\mathbf{I}_N + \mathbf{Z}_O\mathbf{Z}_u^{-1}\right)^{-1}\left(\mathbf{I}_N + \mathbf{Z}_O\mathbf{Z}_A^{-1}\right.$，$\left._u\right)\mathbf{v}_u^{(\mathrm{ant,nomc})} = \mathbf{v}_u^{(\mathrm{ant})}$。我们将耦合矩阵 $\boldsymbol{\Xi}_u \in \mathbb{C}^{N\times N}$ 定义为 $\boldsymbol{\Xi}_u = \left(\mathbf{I}_N + \mathbf{Z}_O\mathbf{Z}_u^{-1}\right)^{-1}\left(\mathbf{I}_N + \mathbf{Z}_O\right.$ $\left.\mathbf{Z}_A^{-1}\right.$，$\left._u\right)$。全息波束成形矩阵变为

$$\tilde{\boldsymbol{M}}_u \triangleq \boldsymbol{\Xi}_u \boldsymbol{M}_u. \tag{2}$$

令 $\boldsymbol{P}_u \in \mathbb{C}^{K\times 1}$ 为子带 u 的功率分配向量，满足 $Tr(\boldsymbol{P}_u\boldsymbol{P}_u^H) = P$ Tx_u，其中 $P_{Tx_u} = \frac{P^{\mathrm{Tx}}}{U}$ 是发射机在子带 u 上的总功率（由所有 U 个子带均分），而 $x_{l,u}$ 为广播信号。子带 u 上 *HMD* l 的接收信号 $\tilde{}$ 为 $y_{l,u} = \boldsymbol{H}_{l,u}\boldsymbol{M}_u\boldsymbol{P}_u x_{l,u} + n_l$，其中 n_l 为方差为 $B_g n_0$ 的高斯噪声，$\boldsymbol{H}_{l,u} \in \mathbb{C}^{1\times N}$ 为 *RHS-HMD*信道向量。假设所有元素到 *HMD* l 的距离近似相等（即 *RHS* 元素之间的间距相较于链路距离可以忽略不计），则 $\boldsymbol{H}_{l,u}$ 中所有元素的信道系数都相同。因此，第 n 个辐射单元的信道增益为 $h_{l,u,n} = h_{l,u}$.

第 u 个子带会受到 IBI I_u 的影响，这是由其他子带的频谱泄漏所致，如公式 (1) 所示。子带 u 上 HMD l 的瞬时信干噪比为

$$\gamma_{l,u} = \frac{G^{\mathrm{Tx}}G^{\mathrm{Rx}}\left|\boldsymbol{H}_{l,u}\tilde{\boldsymbol{M}}_u\boldsymbol{P}_u\right|^2}{G^{\mathrm{Tx}}G^{\mathrm{Rx}}I_u + B_g n_0}, \tag{3}$$

其中 G^{Tx} 和 G^{Rx} 分别表示发射天线增益和接收天线增益。

给定目标误码率（BER）$\epsilon_{l,\,u}$ 对于子带 u上的头戴式显示器 l，采用 M-QAM时可实现的频谱效率近似为$k_{l,\,u}$（比特/秒/赫兹）：$k_{l,\,u} = \log_2\left(1 - \frac{1.5\gamma_l}{(5\epsilon_{l,\,u})}\right)$ $\ln\,。$

头戴式显示器 l在全部 U 个子带上的单独传输速率为

$$R_l = B_g \sum_{u=1}^{U} k_{l,u} = B_g \sum_{u=1}^{U} \log_2\left(1 - \frac{1.5\gamma_{l,u}}{\ln\left(5\epsilon_{l,u}\right)}\right). \tag{4}$$

B. 采用移动边缘计算的虚拟现实传输模型

设 $\mathcal{V} \triangleq \{1, \cdots, V\}$ 表示视场空间，其中这些视场提取自二维视频，如图2所示。对于头戴式显示器 l，每个视场 $i \in \mathcal{V}$ 从二维视场到三维视场的投影表示为三元组（$Q_l^{2\mathrm{D}}$，$Q_l^{3\mathrm{D}}$，o），其中$Q_l^{2\mathrm{D}}$和 $Q_l^{3\mathrm{D}}$分别是头戴式显示器 l 需要处理的二维视场和三维视场的数据大小（单位：比特），o 是每个输入比特所需的渲染周期数（单位：周期/比特）。设 $\alpha_l \triangleq \frac{Q_l^{3\mathrm{D}}}{Q_l^{2\mathrm{D}}}$ 表示头戴式显示器 l处理的三维视场大小与二维视场大小之比。通常，$\alpha_l \geq 2$，这是创建双目立体视觉效果所必需的。

与 [30],类似，头戴式显示器 l 的请求流遵循独立引用模型（IRM）。假设

头戴式显示器 l 在每次请求中请求视场 $i \in \mathcal{V}$ 的概率为常数 $\pi_{l,i}$，且各请求相互独立。概率 $\pi_{l,i}$ 表示头戴式显示器 l 浏览场景时选择视场 i 的频率，并满足 $\sum_{i=1}^{V} \pi_{l,i} = 1$。此外，为避免晕动症和恶心，头戴式显示器 l 的每次请求都必须在最大允许时延 τ内得到服务。

首先，考虑头戴式显示器 l的预取布局。假设头戴式显示器 l 的内存容量为δQ_l^{2D},$_i$ （比特），其中δ 为整数，可用于存储特定视角的二维和三维视场。令 p_l^{2D},$_i \in \{0,1\}$ 表示头戴式显示器 l对视场 i的二维视场预取决策，其中 p_l^{2D},$_i = 1$ 表示二维视场 i已预取到头戴式显示器 l的内存中，而 p_l^{2D},$_i = 0$ 则表示未预取。令 p_l^{3D},$_i \in \{0,1\}$ 表示头戴式显示器 l对视场 i的三维视场预取决策，其中 p_l^{3D},$_i = 1$ 表示三维视场 i已预取到头戴式显示器 l的内存中，而 p_l^{3D},$_i = 0$则表示未预取。在内存约束下，必须满足以下条件：$\sum_{i=1}^{V} Q_l^{2D}$,$_i p_l^{2D}$,$_i + \alpha Q_l^{2D}$,$_i p_l^{3D}$,$_i \le \delta Q_l^{2D}$,$_i$ · 我们假设所有头戴式显示器的所有二维和三维视场均已预取到移动边缘计算服务器。这是合理的，因为移动边缘计算服务器的内存远大于任何头戴式显示器的内存。

接下来，考虑头戴式显示器l上的投影渲染决策。假设头戴式显示器 l 以给定的 CPU 周期频率 f_l （单位：周期/秒）运行，且每个渲染周期的功耗为 ζf_l^2，其中 ζ 是与芯片架构相关的有效开关电容，反映了头戴式显示器 [10]的 CPU 能效。我们假设头戴式显示器的电池容量有限。为确保用户的虚拟现实观看时间不低于要求的平均时长，同时考虑到其他应用也会消耗资源，且需要延长电池寿命，因此引入长期时间平均功耗约束，记为 P_l （单位：瓦特，W）。令 $r_{l,i} \in \{0,1\}$表示头戴式显示器 l 针对视场 i的渲染决策，其中$r_{l,i} = 1$ 表示将二维视场到三维视场 i 的投影在头戴式显示器 l上渲染，而 $r_{l,i} = 0$ 则表示不渲染。在功耗约束下，必须满足以下条件 [30]: $\zeta f_l^2 Q_l^{2D}$, $_{V\tau}{}^{o} \sum_{i=1}^{V} r_{l,i} \le \bar{P}_l$，即 $\sum_{i=1}^{V} r_{l,i} \le \frac{V\bar{P}_l\tau}{\zeta f_l^2 Q_l^{2D,i}{}^{o}}$，这实际上表示头戴式显示器 l 在请求过程中能够渲染的最大视场数量，称为头戴式显示器 l的渲染预算。为简单起见，我们假设 $\frac{V\bar{P}_l\tau}{\zeta f_l^2 Q_l^{2D,i}{}^{o}}$ 为整数。

最后，令 $(\boldsymbol{p}_l^{3D}, \boldsymbol{p}_l^{2D}, \mathbf{r}_l)$ 表示联合预取和 HMD l 对所有视场的渲染决策，其中 $\boldsymbol{p}_l^{3D} \triangleq (p_{l,i}^{3D})_{i\in\mathcal{V}}$ 表示所有视场的三维视场预取决策向量，而 $\boldsymbol{p}_l^{2D} \triangleq (p_{l,i}^{2D})_{i\in\mathcal{V}}$ 表示满足内存约束的二维视场预取决策向量。向量 $\mathbf{r}_l \triangleq (r_{l,i})_{i\in\mathcal{V}}$ 表示渲染决策向量。根据联合预取与渲染决策 $(\boldsymbol{p}_l^{3D}, \boldsymbol{p}_l^{2D}, \mathbf{r}_l)$，HMD l 对视场 $i \in \mathcal{V}$的请求可通过以下四种路径之一得到服务（如图4所示）：

i：设备端三维预取：如果 $p_{l,i}^{3D} = 1$，则可以直接从 HMD l 的内存中获取三维视场i，无需传输或渲染。因此，时延可以忽略不计：$T_{l,i}^{\langle 1\rangle} = 0.$

ii：预取二维视场的设备端渲染：如果 $p_{l,i}^{3D} = 0$，$r_{l,i} = 1$，且 $p_{l,i}^{2D} = 1$，则 HMD l 从内存中获取二维视场i，无需传输，并使用其本地CPU处理器将其渲染为三维视场。总时延为：$T_{l,i}^{\langle 2\rangle} = \frac{Q_{l,i}^{2D}{}^{o}}{f_l}$.

iii：设备端渲染与下载2D：如果 $p_{l,i}^{3D} = 0$、$r_{l,i} = 1$ 和 $p_{l,i}^{2D} = 0$，则 HMD l 会从 MECS 下载二维视场 i并在本地进行渲染。因此，总时延为：$T_{l,i}^{\langle 3\rangle} = \frac{Q_{l,i}^{2D}}{R_l} + \frac{Q_{l,i}^{2D}{}^{o}}{f_l}$，其中$\frac{Q_{l,i}^{2D}}{R_l}$是通过无线链路下载二维视场的传输时延，而 $\frac{Q_{l,i}^{2D}{}^{o}}{f_l}$是渲染时延。

iv：远程三维传输：如果 $p_{l,i}^{3D} = 0$ 且 $r_{l,i} = 0$，那么头戴式显示器 l 会从移动边缘计算服务器下载三维视场 i。总时延为：$T_{l,i}^{\langle 4\rangle} = \frac{Q_{l,i}^{3D}}{R_l}$.

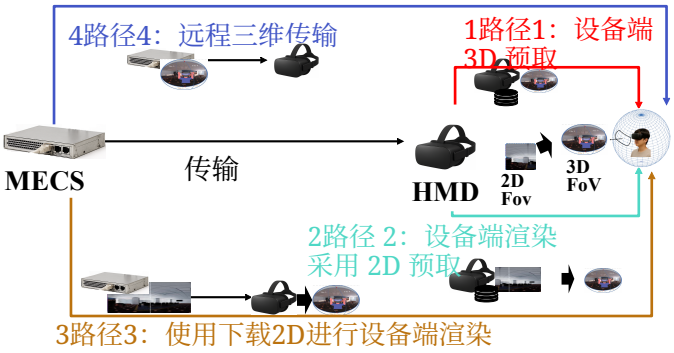


图4：四条投影路径。

C. 问题建模

对于头戴式显示器 l，处理视场 i 的总时延为 $T_{l,i} = p_{l,i}^{3D} T_{l,i}^{\langle 1\rangle} + (1 - p_{l,i}^{3D}) r_{l,i} p_{l,i}^{2D} T_{l,i}^{\langle 2\rangle} + (1 - p_{l,i}^{3D}) r_{l,i} (1 - p_{l,i}^{2D}) T_{l,i}^{\langle 3\rangle} + (1 - p_{l,i}^{3D})(1 - r_{l,i}) T_{l,i}^{\langle 4\rangle}$，代入 $T_{l,i}^{\langle j\rangle}$ 的四个表达式后，$j \in \{1,2,3,4\}$，可得，$p_{l,i}^{3D}(\cdot) \left[r_{l,i}(\cdot), (a), (\cdot), (a) \right]$，其中 $r_{l,i}\left(\frac{Q_{l,i}^{2D}{}^{o}}{f_l} + (1 - p_{l,i}^{2D},) Q_{l,R_l}^{2D}{}^{i}(a\ \boldsymbol{P})\right)$ 表示选择设备端渲染时产生的总时延。如果二维视场已预先缓存在内存中，则不会产生传输时延；否则，还需计入额外的传输时延 $Q_{l,R_l}^{2D}{}^{i}(a\ \boldsymbol{P})$。$(1 - r_{l,i}) Q_{l,R_l}^{3D}{}^{i}(a\ \boldsymbol{P})$ 对应于从移动边缘计算服务器下载三维视场的时延。最后，外部乘法因子 $(1 - p_{l,i}^{3D},)$ 表示仅当三维视场未预先缓存在内存中时才会产生这些时延。$R_l(a\ \boldsymbol{P})$ 表示头戴式显示器 l 的下行传输速率，该速率由可重构全息表面的波束成形参数决定：全息图案权重 a 和馈电功率分配 $\boldsymbol{P}$：$a \triangleq \{a_u\}_{u=1}^{U}, a_u = \{a_{l,u,k}\}_{l=1}^{L},_{k=1}^{K}$，以及 $\boldsymbol{P} \triangleq \{\boldsymbol{P}_u\}_{u=1}^{U}$，$\boldsymbol{P}_u = \{P_{u,k}\}_{k=1}^{K}$ $\boldsymbol{P}_u \boldsymbol{P}_u^H = P_u = \frac{P^{Tx}}{U}$，Tr( ) Tx。

我们现在构建一个联合优化问题，通过联合优化内容预取$(\boldsymbol{p}_l^{3D}, \boldsymbol{p}_l^{2D})$、渲染卸载$\mathbf{r}_l$和RHS全息波束成形（$a\ \boldsymbol{P}$）决策，最小化所有头戴式显示器和视场的平均视场处理时延。

问题1（联合时延最小化）。

$$
\begin{aligned}
(\mathcal{P}_1): \min_{\substack{\boldsymbol{p}_l^{3D},\boldsymbol{p}_l^{2D},\mathbf{r}_l,\\ \mathbf{a},\boldsymbol{P}}} \quad & \frac{1}{LV}\sum_{l=1}^{L}\sum_{i=1}^{V}\pi_{l,i}T_{l,i} \\
\text{s.t.} \quad & \textit{(C1)}\ \sum_{i=1}^{V}\left(Q_{l,i}^{2D}p_{l,i}^{2D}+Q_{l,i}^{3D}p_{l,i}^{3D}\right)\le\delta, \quad \forall l\in\mathcal{L}, \\
& \textit{(C2)}\ \sum_{i=1}^{V}\pi_{l,i}\frac{\zeta f_l^2 Q_{l,i}^{2D}o}{\tau}r_{l,i}\le\bar{P}_l, \quad \forall l\in\mathcal{L}, \\
& \textit{(C3)}\ \frac{Q_{l,i}^{2D}}{R_l(\mathbf{a},\boldsymbol{P})}+\frac{Q_{l,i}^{2D}o}{f_l}\le\tau, \quad \textit{if } r_{l,i}=1,\ p_{l,i}^{2D}=0, \\
& \textit{(C4)}\ \frac{Q_{l,i}^{3D}}{R_l(\mathbf{a},\boldsymbol{P})}\le\tau, \quad \textit{if } r_{l,i}=0, \\
& \textit{(C5)}\ p_{l,i}^{2D},\ p_{l,i}^{3D},\ r_{l,i}\in\{0,1\}, \quad \forall l\in\mathcal{L},\ i\in\mathcal{V}, \\
& \textit{(C6)}\ a_{l,u,k}\ge 0, \quad \sum_{l=1}^{L}\sum_{k=1}^{K}a_{l,u,k}=1, \forall u\in\{1,\dots,U\}, \\
& \textit{(C7)}\ P_{u,k}\ge 0,\ \boldsymbol{P}_u=\{P_{u,k}\}_{k=1}^{K},
\end{aligned}
\tag{5}
$$

III. 齐次性与静态性假设下的等价重构

我们假设所有视场均服从均匀分布，即$\pi_{l,\ i}=\frac{1}{V}$对每个 $i\in\mathcal{V}$均成立。我们还假设，对于任意头戴式显示器 l，所有视场大小相等，即$Q_{l}^{2D}{}_{,\ i}=Q^{2D}$,
$Q_{l\ i}^{3D}=Q^{3D}=\alpha Q^{2D}$ f_l f $p_{l\ i}^{2D}$ p_i^{2D} , , 并将标记为 ,. 标记为 ,$p_{l\ i}^{3D}$ p_i^{3D} $\bar{P}_l$ $\bar{P}$ $r_{l\ i}$ r_i , 标记为 , 标记为 , 并将. 标记为 。在静态性假设下，头戴式显示器保持静止，因此可重构全息表面与头戴式显示器之间的几何关系不会改变。因此，RHS波束成形配置固定，下行传输速率 R_l保持不变（标记为 R）。于是，头戴式显示器

l $1-$的总时延目标p_i^{3D} $1-p_i^{2D}$ $1-r_i$
可简化为$\sum_{i=1}^{V}$()
$\left[r_i\left(\frac{Q^{2D}o}{f}+\right.\right.$ $\left.\left.\frac{Q^{2D}}{R}\right)+\right.$ $\left.\frac{\alpha Q^{2D}}{R}\right]$ $=$() () () ()
$\sum_{i=1}^{V}1-p_i^{3D}\ Q^{2D}\left[r_i\left(\frac{o}{f}+\frac{1-p_i^{2D}}{R}\right)+1-r_i\ \frac{\alpha}{R}\right]$. $-$

相应的约束条件变为：s.t.

$$
\begin{aligned}
&\text{(C1e)} \quad \sum_{i=1}^{V}p_i^{2D}+\alpha p_i^{3D}\le\delta, \\
&\text{(C2e)} \quad \sum_{i=1}^{V}r_i\le\frac{V\bar{P}\tau}{\zeta f^2 Q^{2D}o}\triangleq Q_{\max}, \\
&\text{(C3e)} \quad \frac{Q^{2D}}{R}+\frac{Q^{2D}o}{f}<\tau, \\
&\text{(C4e)} \quad \frac{Q^{3D}}{R}\le\tau. \\
&\text{(C5e)} \quad p_i^{2D},p_i^{3D},r_i\in\{0,1\}, \quad \forall l\in\mathcal{L},i\in\mathcal{V}.
\end{aligned}
\tag{6}
$$

命题1（3D 预取消除了对 2D 预取和渲染的需求）。 对于任意视场 $i\in\mathcal{V}$，如果 $p_i^{3D}=1$，则在最优解中有 $p_i^{2D}=0$ 且 $r_i=0$。

证明。 如果 $p_i^{3D}=1$，则三维视场 i 已在本地预取，因此无需进行传输或渲染；等效时延近似为零。在这种情况下设置 $p_i^{2D}=1$ 和 $r_i=1$ 只会额外消耗内存或计算资源并增加时延。因此，最优解必须满足 $p_i^{2D}=0$ 和 $r_i=0$。 □

命题2（2D 预取必须伴随本地渲染）。 对于任意视场 $i\in\mathcal{V}$，若 $r_i=0$，则不失最优性，可令 $p_i^{2D}=0$。

证明。 若 $r_i=0$，则不会对视场 i 进行本地渲染，因此永远不会使用二维视场。该视场必须通过远程三维传输提供，其时延与二维预取决策无关。设置 $p_i^{2D}=1$ 只会占用不必要的内存，增加时延。因此，设置 $p_i^{2D}=0$ 为最优。 □

命题3（最优解处内存约束取等）。 在最优解处，内存约束满足 $\sum_{i=1}^{V}p_i^{2D}+\alpha p_i^{3D}=\delta$，且不失最优性。

证明。考虑任意一个未充分利用可用内存的可行解。假设存在一个视场 j当前通过远程三维传输路径提供服务（即 $p_j^{3D}=0$ 和 $r_j=0$），其时延为 $\frac{Q^{3D}}{R}$。若仍有剩余内存，则可将其用于存储该三维视场（即设置 $\overline{p_j^{3D}}=1$），从而将其时延从$\frac{Q^{3D}}{R}$ 降至近似为零。这会严格降低总目标值。因此，在最优性条件下，不可能在约束不紧时仍存在未使用的内存。故内存约束必须以等式成立。

□

综合命题1、2和3，我们可以得到一个以聚合变量表示的等价问题。将头戴式显示器的聚合决策变量定义为 $p^{3D}\triangleq\sum_{i=1}^{V}p_i^{3D}$ $p^{2D}\triangleq\sum_{i=1}^{V}p_i^{2D}$ $r\triangleq\sum_{i=1}^{V}r_i$, , ,

我们可以“在不损失最优性的前提下”假设如下结构:

- 视场 $i=1,\dots,p^{3}{}_{\mathrm{D}}$：设备端三维预取;
- 接下来的 $q\triangleq\min\{p^{2}{}_{\mathrm{D}}r\}$ 个视场：预取二维视场的设备端渲染;
- 接下来的 $r-q$ 个视场：下载2D视场的设备端渲染;
- 剩余的 $V-p^{3}{}_{\mathrm{D}}-r$ 个视场：远程三维传输。

令 $T^{\langle 2\rangle}\triangleq\frac{Q^{2D}o}{f}$ $T^{\langle 3\rangle}\triangleq\frac{Q^{2D}}{R}+\frac{Q^{2D}o}{f}$ $T^{\langle 4\rangle}\triangleq\frac{Q^{3D}}{R}=$,,。
$\frac{\alpha Q^{2D}}{R}$
平均时延可写为: $\bar{T}=$
$\frac{1}{V}\left[qT^{\langle 2\rangle}+\ r-q\ T^{\langle 3\rangle}+\left(V-p^{3D}-r\right)T^{\langle 4\rangle}\right]$ $q=$(),其中

$\min\{p^{2D},r\}$。根据命题3，内存约束取等号: $p^{2D}+\alpha p^{3D}=\delta\Rightarrow p^{3D}=\frac{\delta-p^{2D}}{\alpha}$. 平均功率约束变为: $r\le Q_{\max}$, $Q_{\max}\triangleq\frac{V\bar{P}\tau}{\zeta f^2 Q^{2D}o}$. 我们还假设 $\frac{\delta}{\alpha}+Q_{\max}\le V$,，因此通过预取和本地渲染服务的视场总数不超过 V。

我们现在得到等价问题:

问题1'（在齐次性和静态性假设下等价于（$\mathcal{P}_1$））。

$$
\begin{aligned}
(\mathcal{P}_1')\ &\min_{p^{2\mathrm{D}},p^{3\mathrm{D}},r} \bar{T}\left(p^{2\mathrm{D}},p^{3\mathrm{D}},r\right)=\min_{p^{2\mathrm{D}},r}\frac{1}{V}\\
&\left[qT^{\langle 2\rangle}+(r-q)T^{\langle 3\rangle}+\left(V-\frac{\delta-p^{2\mathrm{D}}}{\alpha}-r\right)T^{\langle 4\rangle}\right]\\
\text{s.t.}\quad & p^{2\mathrm{D}}\in\{0,1,\ldots,\delta\}, p^{3\mathrm{D}}=\frac{\delta-p^{2\mathrm{D}}}{\alpha}, p^{3\mathrm{D}}+r\le V,\\
& r\in\{0,1,\ldots,Q_{\max}\}, q=\min\left\{p^{2\mathrm{D}},r\right\}
\end{aligned}
\tag{7}
$$

核心比较是"下载2D后进行设备端渲染"（路径3）的时延与"远程三维传输"（路径4）的时延：$T^{\langle 3\rangle}$？$\gtrless T^{\langle 4\rangle} \Leftrightarrow \frac{Q^{2\mathrm{D}}}{R}+\frac{Q^{3\mathrm{D}}}{f}\gtrless\frac{\alpha Q^{2\mathrm{D}}}{R}$·将两边同时除以$Q^{2\mathrm{D}}$:——$\frac{1}{R}+\frac{o}{f}\gtrless\frac{\alpha}{R}\Longleftrightarrow\frac{o}{f}\gtrless\frac{\alpha-1}{R}$·_ _ _ _ ——

因此，我们定义两个区域。

定义1（远程三维传输优先区）。远程三维传输优先的区域由以下条件定义：$T^{\langle 3\rangle}\ge T^{\langle 4\rangle}\Leftrightarrow\frac{o}{f}\ge\frac{\alpha-1}{R}$，其中本地渲染（不进行预取）的效率低于远程三维传输或速度更慢。

定义2（设备端渲染优先区域）。本地渲染优先的区域由以下条件定义：$T^{\langle 3\rangle}<T^{\langle 4\rangle}\Longleftrightarrow o\le\frac{\alpha-1}{f}$，其中本地渲染速度快于远程三维传输。

以下命题4和命题5给出了 $\mathcal{P}_1'$ 的最优策略。

命题4（远程 3D 传输优先区的最优策略为闭式解）。假设对于某个头戴式显示器，常数 $R\quad f\quad\frac{o}{f}\ge$*and*满足，$\frac{\alpha-1}{R}\quad T^{\langle 3\rangle}\ge T^{\langle 4\rangle}$即，且内存和功率预算满足 $\frac{\delta}{\alpha}+Q_{\max}\le V$. 则，$\mathcal{P}_1'$ 的最优联合策略为 $p^{2\mathrm{D}^*}=$最小 $\{\delta,Q_{\max}\}$, $r^*=p^{2\mathrm{D}^*}$, $p^{3\mathrm{D}^*}=\frac{\delta-p^{2\mathrm{D}^*}}{\alpha}$。相应的最小平均时延为：

$\bar{T}^*=T^{\langle 4\rangle}-\frac{T^{\langle 4\rangle}}{V}$·，其中$\frac{\delta}{\alpha}-\frac{T^{\langle 4\rangle}-T^{\langle 2\rangle}}{V}\cdot p^{2\mathrm{D}^*}\quad T^{\langle 2\rangle}=\frac{Q^{2\mathrm{D}o}}{f}\quad T^{\langle 4\rangle}=\frac{Q^{3\mathrm{D}}}{R}$ ——*and*。证明。参见附录 A。 □

命题5（设备端渲染优先区域的最优策略具有闭式解）。假设对于某个头戴式显示器，常数 $R\quad f\quad\frac{o}{f}<\frac{\alpha-1}{R}$ —— $T^{\langle 3\rangle}<$和满足 $T^{\langle 4\rangle}$，即，且内存和功率预算满足 $\frac{\delta}{\alpha}+Q_{\max}\le V$.那么，$\mathcal{P}_1'$ 的最优联合策略为 $p^{2\mathrm{D}^*}=$最小化 $\{\delta,Q_{\max}\}$, $r^*=Q_{\max}$ $p^{3\mathrm{D}^*}=\frac{\delta-p^{2\mathrm{D}^*}}{\alpha}$。相应的最小平均时延为：

$\bar{T}^*=T^{\langle 4\rangle}-\frac{T^{\langle 4\rangle}}{V}\cdot$ $\frac{\delta}{\alpha}-\frac{T^{\langle 4\rangle}-T^{\langle 2\rangle}}{V}p^{2\mathrm{D}^*}-\frac{T^{\langle 4\rangle}-T^{\langle 3\rangle}}{V}\left(Q_{\max}-p^{2\mathrm{D}^*}\right)$.

证明。见附录B。 □

对于同质静态多用户情形，由于每个视场都相同，头戴式显示器的内存和计算约束彼此独立，且下行传输速率保持不变，因此我们通过以下方式求解：

i: 对于每个头戴式显示器 l，我们可以写出一个单独的问题1'。

ii: 根据头戴式显示器 $\left(R_l,f_l,\bar{P}_l\right)$ 的参数，我们确定其属于远程 3D 传输优先区还是设备端渲染优先区域。

iii: 相应地应用命题4或命题5，以得到 $\left(p_l^{3\mathrm{D}^*},p_l^{2\mathrm{D}^*},r_l^*\right)$。

iv: 整个多用户系统的最优联合策略是所有头戴式显示器的最优策略的集合。

IV. 异质性与非静态假设下的等效重构

在异构场景中，各头戴式显示器的视场参数各不相同，而太赫兹下行速率取决于RHS波束成形。由于RHS能够实时重新配置其全息波束，用户移动（例如，头戴式显示器在室内移动）需要频繁进行短时间尺度的波束更新，而预取和渲染决策的变化则慢得多。这种自然的时间尺度分离促使我们将联合问题分解为一个利用长期等效速率的长时间尺度预取–渲染子问题（$\mathcal{P}$1-1），以及一个能够快速适应用户移动的短时间尺度RHS波束成形子问题（$\mathcal{P}$1-2）。$\mathcal{P}$1-1和 $\mathcal{P}$1-2通过下行速率向量$\mathbf{R}=(R_1,\ldots,R_L)$相互耦合。

问题1-1（预取-渲染优化（给定 $\mathbf{R}$））。我们假设物理层已在更快的时间尺度上接收到长期等效速率向量$\mathbf{R}$，则：

$$
(\mathcal{P}1\text{-}1):\min_{\boldsymbol{p},\mathbf{r}}\bar{T}=\min_{\boldsymbol{p},\mathbf{r}}\frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{V}\pi_{l,i}T_{l,i}. \tag{8}
$$

我们为每个头戴式显示器l、视场 i和路径 $j\in\{1,2,3,4\}$引入路径选择变量$x_{l,i,j}$，其中$x_{l,i,j}\in\{0,1\}$；若$x_{l,i,j}=1$，则表示视场（l,i）沿路径 j，否则不沿该路径。对于每个（l,i），必须恰好选择一条路径：$\sum_{j=1}^{4}x_{l,i,j}=1,\forall l,i.$四条路径对应的时延为：$T_l^{\langle 1\rangle},^i=0$, $T_l^{\langle 2\rangle},^i=\quad,^{io}$, $T_l^{\langle 3\rangle},^i=\quad,\frac{Q_l^{2\mathrm{D}}}{R_l}{}^i+\frac{Q_l^{2\mathrm{D}},^{io}}{f_l}$, $T_{l\,i}^{\langle 4\rangle}=\frac{Q_{l\,i}^{3\mathrm{D}}}{R_l}$. —— t ,,因此，头戴式显示器的平均时延可写为：$\sum_{i=1}^{V}\pi_{l,i}\sum_{j=1}^{4}T_l^{\langle j\rangle},^i\,x_{l,i,j}$。将所有用户相加并除以 L，得到总时延目标：$\bar{T}=\frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{V}\pi_{l,i}\sum_{j=1}^{4}T_l^{\langle j\rangle},^i\,x_{l,i,j}$。

对于每条路径，定义内存消耗 $C_{l\,i}^{(\mathrm{m})},_{,j}$ 和功耗 $C_{l\,i}^{(\mathrm{p})},_{,j}$：

$$
C_{l,i,j}^{(\mathrm{m})}\triangleq\begin{cases}Q_{l,i}^{3\mathrm{D}}, & j=1,\\ Q_{l,i}^{2\mathrm{D}}, & j=2,\\ 0, & j=3,4,\end{cases}\quad C_{l,i,j}^{(\mathrm{p})}\triangleq\begin{cases}\pi_{l,i}\frac{\zeta f_l^2 D_{l,i}^{2\mathrm{D}}o}{\tau}, & j=2,3,\\ 0, & j=1,4.\end{cases} \tag{9}
$$

该问题等价于以下形式：

问题 1-1-H1 (多重选择多维背包问题)。

$$
\begin{aligned}
&(\mathcal{P}1\text{-}1\text{-}H1)\\
&\min_{\{x_{l,i,j}\}}\quad \bar{T}=\tfrac{1}{L}\textstyle\sum_{l=1}^{L}\sum_{i=1}^{V}\pi_{l,i}\sum_{j=1}^{4}T_{l,i}^{\langle j\rangle}x_{l,i,j}\\
&\text{s.t.}\quad \textstyle\sum_{i=1}^{V}\sum_{j=1}^{4}C_{l,i,j}^{(\mathrm{m})}x_{1,i,j}\le\delta Q_{l,i}^{2\mathrm{D}},\ \forall l\in\mathcal{L},\\
&\quad\ \textstyle\sum_{i=1}^{V}\sum_{j=1}^{4}C_{l,i,j}^{(\mathrm{p})}x_{1,i,j}\le\bar{P}_l,\forall l\in\mathcal{L},\\
&\quad\ \textstyle\sum_{j=1}^{4}x_{l,i,j}=1,\ \forall l\in\mathcal{L},i\in\mathcal{V},\\
&\quad\ x_{l,i,j}\in\{0,1\},\forall l\in\mathcal{L},i\in\mathcal{V},j\in\{1,2,3,4\}.
\end{aligned}
\tag{10}
$$

这是一个标准的多选择多维背包问题 (MMKP)：每个 $(l,\ i)$ 都需要从四个选项中选择一个（多选择）；该问题包含两个资源维度，即内存和功率（多维）。

令路径4（远程三维传输）为基准，并将字段 i 的基准时延定义为 $T_{l\,i}^{\langle 4\rangle}$ ，。“时延增益”（减少量）定义为路径 j 的:

$$g_{l,i,j} \triangleq \begin{cases} \pi_{l,i}\left(T_{l,i}^{\langle 4\rangle} - T_{l,i}^{\langle 1\rangle}\right), & j=1, \\ \pi_{l,i}\left(T_{l,i}^{\langle 4\rangle} - T_{l,i}^{\langle 2\rangle}\right), & j=2, \\ \pi_{l,i}\left(T_{l,i}^{\langle 4\rangle} - T_{l,i}^{\langle 3\rangle}\right), & j=3, \\ 0, & j=4. \end{cases} \tag{11}$$

¯因此，总平均时延可表示为: $T = \frac{1}{L}\sum_{l\,i}\pi_{l\,i}T_{l\,i}^{\langle 4\rangle} - \frac{1}{L}\sum_{l\,i}\sum_{j=1}^{4} g_{l\,i\,j}x_{l\,i\,j}$ ，。第一项为常数，不受决策 $x_{l,\ i,\ j}$ 影响；因此，¯最小化 T 等价于最大化总时延增益：$\max_{\{x_{l,i,j}\}}\sum_{l,\ i}\sum_{j=1}^{4} g_{l,\ i,\ j}x_{l,\ i,\ j}$。

接下来，我们介绍 $\mathcal{P}$1问题1-1-H2。

问题1-1-H2 （二元松弛与 IQP 问题）。

$$\begin{aligned} (\mathcal{P}\text{1-1-H2}) \min_{\{x_{l,i,j}\}} & \sum_{l,i}\sum_{j=1}^{4} -g_{l,i,j}x_{l,i,j} \\ \text{s.t.} \quad & \sum_{j} x_{l,i,j} = 1, \forall l \in \mathcal{L}, i \in \mathcal{V}, j \in \{1,2,3,4\} \\ & 0 \le x_{l,i,j} \le 1, \forall l \in \mathcal{L}, i \in \mathcal{V}, j \in \{1,2,3,4\} \\ & \sum_{l,i,j} x_{l,i,j}\left(1 - x_{l,i,j}\right) \le 0 \end{aligned} \tag{12}$$

在此，我们将二元决策变量松弛到连续域，即 $x_{l,i,j} \in [0,1], \forall l, i, j$。在该松弛下，原始二元约束可通过二次不等式等价地强制满足。具体而言，整数性条件 $x_{l,i,j} \in \{0,1\}$ 等价于 $x_{l,i,j}(1 - x_{l,i,j}) = 0$，并可进一步改写为聚合形式 $\sum_{l,i,j} x_{l,i,j}(1 - x_{l,i,j}) \le 0$. 该重构将原始二元约束转化为线性约束与凹二次约束的组合（因为 $x_{l,i,j}(1 - x_{l,i,j})$ 是凹的），从而得到一种适用于差分凸技术的整数二次规划（IQP）结构。

现在，我们可以添加一个惩罚项 $\lambda x_{l,i,j}(1 - x_{l,i,j})$。

问题1-1-H3 （惩罚项与凹凸过程问题）。

$$\begin{aligned} (\mathcal{P}\text{1-1-H3}) \min_{\{x_{l,i,j}\}} & \sum_{l,i}\sum_{j=1}^{4}\left(-g_{l,i,j}x_{l,i,j} - \lambda x_{l,i,j}\left(x_{l,i,j} - 1\right)\right) \\ \text{s.t.} \quad & \sum_{j}^{4} x_{l,i,j} = 1, \forall l \in \mathcal{L}, i \in \mathcal{V} \\ & 0 \le x_{l,i,j} \le 1, \forall l \in \mathcal{L}, i \in \mathcal{V}, j \in \{1,2,3,4\} \end{aligned} \tag{13}$$

其中，$\lambda > 0$ 是惩罚参数。

目标函数由线性项 $-g_{l,i,j}x_{l,i,j}$ 和凹二次惩罚项 $-\lambda x_{l,i,j}(x_{l,i,j} - 1)$ 构成；因此，它具有差分凸（DC）结构，可以使用差分凸算法 (DCA) 或凸-凹过程 (CCCP) 求解。

算法 1 $\mathcal{P}$1**-1-H1（MMKP）的启发式算法 需要：** $\{\pi_{l,i}\}$、$\{T_{l,i}^{\langle j\rangle}\}$、资源代价 $\{C_{l,i,j}^{(\mathrm{m})}, C_{l,i,j}^{(\mathrm{p})}\}$、内存 ¯预算 δ、功率预算 $\{P_l\}$、避免除零的常数 ϵ、内存代价加权系数 α_1、功率代价加权系数 α_2**确保：可行的二元解** $\{x_{l,i,j}\}$**1：对所有** (l,i)**，设置** $x_{l,i,4} \leftarrow 1$ **（远程三维传输基准）；其余设为 0。记录初始资源使用量** $\mathbf{mem}_l$ **和** $\mathbf{pow}_l$**。2：对于每个** (l,i) **和** $j \in \{1,2,3\}$**，执行 3：时延增益：** $g_{l,i,j} \leftarrow \pi_{l,i}(T_l^{\langle 4\rangle}{}_{,i} - T_{l\,i}^{\langle j\rangle}{}_{,})$ **4：得分：** $\rho_{l,i,j} \leftarrow g_{l,i,j}/(\epsilon + \alpha_1 C_{l\,i}^{(\mathrm{m})}{}_{,} + \alpha_2 C_{l\,i\,j}^{(\mathrm{p})}{}_{,})$ **5：结束循环 6：按** $\rho_{l,i,j}$ **降序排列所有候选项** (l,i,j)**。7：依次处理排序后的** (l,i,j) **8：令** j' **为** (l,i) **的当前路径；计算暂定资源使用量** $\tilde{}\mathbf{mem}_l = \mathbf{mem}_l - C_{l\,i}^{(\mathrm{m})}{}_{,} + C_{l\,i\,j}^{(\mathrm{m})}{}_{,}$，$\tilde{}\mathbf{pow}_l = \mathbf{pow}_l - C_{l\,i}^{(\mathrm{p})}{}_{,} + C_{l\,i}^{(\mathrm{p})}{}_{,}$ **9：如果** $\tilde{}\mathbf{mem}_l \le \delta$ **且** $\tilde{}\mathbf{pow}_l \le P_l$**，则 10：** $x_{l,i,j'} \leftarrow 0, x_{l,i,j} \leftarrow 1$**；更新** $\mathbf{mem}_l$、$\mathbf{pow}_l$**。11：结束条件判断 12：结束循环 13：返回** $\{x_{l,i,j}\}$

算法 2 二元松弛与整数二次规划，用于 $\mathcal{P}$1-1-H2

要求： 增益 $\{g_{l,i,j}\}$，成本 $\{C_{l,i,j}^{(\mathrm{m})}, C_{l,i,j}^{(\mathrm{p})}\}$，预算 $\delta Q_{l\,i}^{2\mathrm{D}}{}_{,}$，$\{P_l\}$，惩罚项 $\lambda > 0$

确保： 松弛解 $\{x_{l\quad\quad,i,j}\} \in [0,1]$

1：使用算法 1 获得一个可行二元点 $\{x_{l,i,j}^{(0)}\}$**。2：将** $x_{l,i,j} \in \{0,1\}$ **松弛为** $0 \le x_{l,i,j} \le 1, \sum_j x_{l,i,j} = 1$**，并保留所有内存/功率约束。3：考虑惩罚后的整数二次规划目标函数**

$f\ x\ = \sum_{l\ i\ j}\Big(-(\ )_{,}(\ )_{,}(\ )_{,}$ **其凹部分** $g_{l,i,j}x_{l,i,j} - \lambda x_{l,i,j}\ x_{l\ i\ j} - 1\ \Big)$

为 $-\lambda x_{l,\quad i\ j\quad i\quad i\ j\quad i\ j}-_{,}(x_{l,}{}_{,j} - 1)$**。在** $x^{(0)}$ **处对其线性化：** $-\lambda x_{l,}{}_{,}(x_{l,}{}_{,})(\ )_{,}(\ )(\ )_{,}(\ )_{,}(\ )$ $\Big)$**.：去掉与 无关的常数，并求解以下凸二次规划：** $\min_{\{x_{l,i,j}\}}$ $\sum_{l\ i\ j}\Big(-g_{l,i,j}x_{l,i,j} -_{,}(\ )_{,}(\ )_{,}$ **subject to** $\lambda\ 2x_{l\ i\ j}^{(0)} - 1\ x_{l\ i\ j}\Big)$ $\sum_j x_{l\ i\ j} = 1\ 0 \le x_{l\ i\ j} \le_{,}$，$\sum_{i\ j}^{1} C_{l\ i\ j}^{(\mathrm{m})}x_{l\ i\ j} \le \delta Q_{l\ i}^{2\mathrm{D}}\ \sum_{i\ j} C_{l\ i\ j}^{(\mathrm{p})}x_{l\ i\ j} \le \bar{P}_l\ \ \forall l._{,,,,,,,,,,,,,,}$ **：记最优解为** $\{x_{l,i,j}\}$**，并将其作为松弛的 H2 解（如有需要，也可用作算法 3 的初始值）。6：返回** $\{x_{l,i,j}\}$

首先通过贪心时延增益资源成本启发式（算法 1）获得一个可行二元解，随后进行二元松弛并采用单次 DCA/CCCP 精化（算法 2），最后使用带舍入的完整 DCA/CCCP 求解器（算法 3）。

假设预取-渲染策略（$\boldsymbol{p}$，$\boldsymbol{r}$）已

$$(\mathcal{P}1\text{-}2):\max_{\mathbf{a},\boldsymbol{P}}\sum_{l=1}^{L}\frac{R_l}{\omega_l}=B_g\sum_{l=1}^{L}\frac{\sum_{u=1}^{U}k_{l,u}}{\omega_l}=\max_{\mathbf{a},\boldsymbol{P}}B_g\sum_{l=1}^{L}\left(\sum_{u=1}^{U}\log_2\left(1-\frac{1.5\frac{G^{\mathrm{Tx}}G^{\mathrm{Rx}}\left|\boldsymbol{H}_{l,u}\tilde{\boldsymbol{M}}_u\boldsymbol{P}_u\right|^2}{G^{\mathrm{Tx}}G^{\mathrm{Rx}}I_u+B_gn_0}}{\ln\left(5\epsilon_{l,u}\right)}\right)\right)/\omega_l,\qquad(14)$$

算法 3 用于 $\mathcal{P}$1-1-H3的DCA/CCCP算法

要求：增益 $\{g_{l,i,j}\}$**、惩罚项** $\lambda>0$**、容差** $\varepsilon>0$**、最大迭代次数** $T_{\max}$**确保：近似二元解** $\{x_{l,i,j}\}$

1 **：获得一个初始可行点** $\{x_{l,i,j}^{(0)}\}$ **（例如，由算法 1 得到）；设置** $t\leftarrow 0$。**2：当** $t<T_{\max}$ **时执行 3：对于所有 (**l,i,j**)，计算凹项在** $x^{(t)}$ **处的** $\nabla_{x_{l\,i\,j}}\left[\,-\lambda x_{l,i,j}\;x_{l\,i\,j}\;-\;1\;\right]\Big|_{x^{(t)}}=-\lambda\big(2x_{l\,i\,j}^{(t)}-1\big)$.

梯度：, , (, ,) , , 通过线性化凹部分构建凸替代目标函数： $\tilde{f}^{(t)}\;x\;=\sum_{l\,i\,j}\Big(-\mathbf{(}\mathbf{)},,g_{l,i,j}x_{l,i,j}-\lambda\;2x_{l\,i\,j}^{(t)}-1\;\;x_{l\,i\,j}\Big)$ **(, ,) , , , 其中省略了关于** x **的常数项。4：求解以下凸线性规划/二次规划子问题：** $\mathbf{min}_{\{x_{l,i,j}\}}\;\tilde{f}^{(t)}(x)$**，满足** $\sum_j x_{l,i,j}=1,0\le x_{l,i,j}\le 1$**，以及对所有** l,i,j **的内存/功率约束，并将解记为** $\{x_{l,i,j}^{(t+1)}\}$。**5：如果** $\mathbf{max}_{l,i,j}\left|x_{l\,i}^{(t+1)},,_j-x_{l\,i}^{(t)},,_j\right|\le\varepsilon$**，则 6：跳出 7：结束 if 8：**$t\leftarrow t+1$**9：结束 while 10：对于每个 (**$l,i$**)，(舍入)** $j^*\leftarrow\mathbf{arg\,max}_j\,x_{l\,i}^{(t)},,_j$**；对于所有** $j\neq j^*$**，设置** $x_{l,i,j^\star}\leftarrow 1,x_{l,i,j}\leftarrow 0$ **11：返回** $\{x_{l,i,j}\}$

算法4 用于 $\mathcal{P}$2–2的投影梯度（PG）求解器

要求：信道 $\boldsymbol{H}_{l,u}$**、干扰项** I_u**、步长**η_P,η_a**、停止容差** ε**确保：馈电功率** $\{\boldsymbol{P}_u\}_{u=1}^{U}$**、全息权重：初始化 (**$\boldsymbol{P}^0$**)**$_u$**，使得** $\|\boldsymbol{P}(\boldsymbol{)}_u\|^2=P^{\mathbf{Tx}_u}$**、**$P^{\mathbf{Tx}_u}=P^{\mathbf{Tx}}/U$**，并初始化** $a_{l\,u}^{(0)},,_k$**，计算** $\boldsymbol{M}(\mathbf{0})_u$**，对所有** u**均成立。设置** $i\leftarrow 0$。**2：当**$\left|J^{(i+1)}-J^{(i)}\right|>\varepsilon$**时循环 3：使用附录 C 计算梯度** $\nabla_{\boldsymbol{P}_u^*}J^{(i)}$ **和** $\nabla_{a_{l,u,k}}J^{(i)}$。**梯度上升更新：** $\boldsymbol{P}_u\leftarrow\boldsymbol{P}_u+\mathbf{(}_{i+1}\mathbf{)}\,\mathbf{(}_i\mathbf{)},,,,,,,,\eta_P\nabla_{\boldsymbol{P}_u^*}J^{(i)}\;\;a_{l\,u\,k}^{(i+1)}\leftarrow a_{l\,u\,k}^{(i)}+\eta_a\nabla_{a_{l\,u}}{}_kJ^{(i)}.$**：将 (**$\boldsymbol{P}^{i+1}$**)**$_u$**投影到** $\|\boldsymbol{P}_u\|^2=P^{\mathbf{Tx}_u}$**；将** $\{a_{l,u,k}^{(i+1)}\}$**投影到单纯形** $\{a_{l,u,k}\ge 0,\;\sum_{l,k}a_{l,u,k}=1\}$。**5：对所有** u**更新** $\boldsymbol{M}(_{i+1})_u$。$i\leftarrow i+1$。**6：结束循环**

已确定。我们引入问题1-2。

问题1-2 (可重构全息表面全息波束成形（给定预取-渲染策略）)。 物理层的目标是优化全息图案权重 $\mathbf{a}$ 和馈电功率分配向量 $\boldsymbol{P}$，以最大化总速率。总速率基于多个子带上的信干噪比，如式(14)所示，

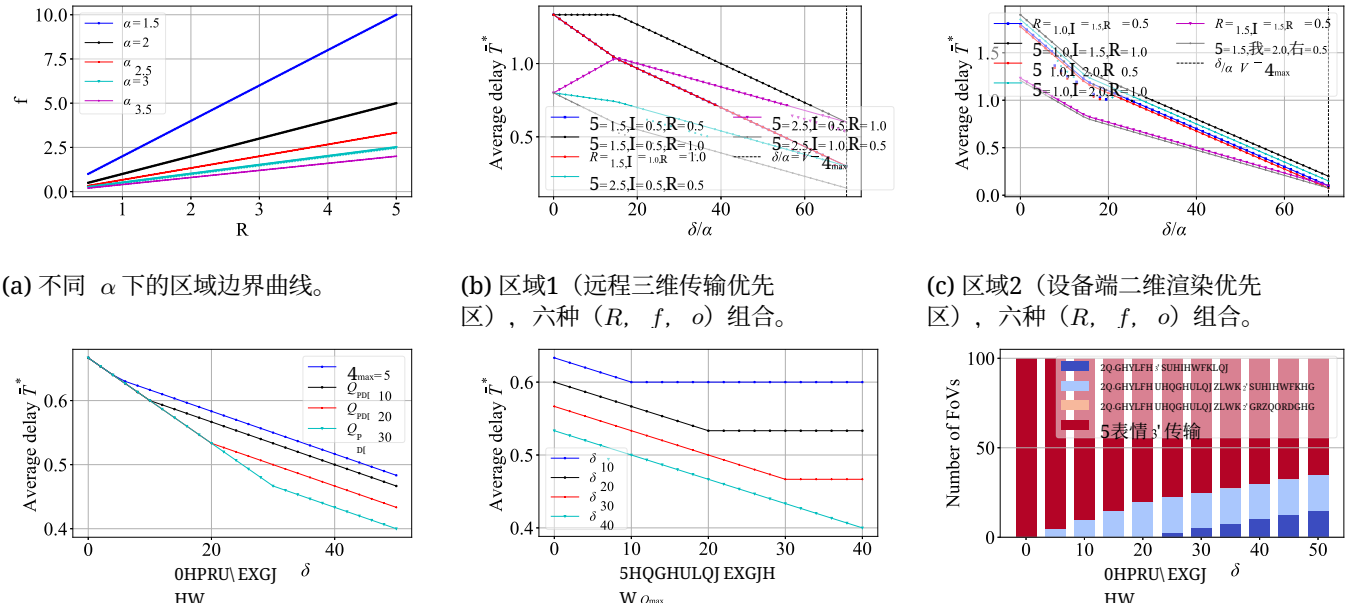
满足约束条件 *(C6)* 和 *(C7)*。

此处，$\omega_l=\sum_{i=1}^{V}\pi_{l\,i}^{l}\Big(1-p_{l,i}^{3\mathrm{D}}\Big)\left[r_{l,i}\left(1-p_{l,i}^{2\mathrm{D}}\right)Q_{l,i}^{2\mathrm{D}}+\;1-r_{l\,i}\;\;Q_{l\,i}^{3\mathrm{D}}\right]$ 由 $\mathcal{P}$1-1的最优解导出：,(,) ,，其本质上表示在当前预取-渲染策略下，仍需通过无线链路获取的关键数据预期量。该值越高，物理层应分配的速率就越大。

$\boldsymbol{H}_{l,u}\tilde{\boldsymbol{M}}_u\boldsymbol{P}_u$ 是从可重构全息表面传输到头戴式显示器 l 的信号，$\boldsymbol{H}_{l,\;u}\in\mathbb{C}^{1\times N}$，$\tilde{\boldsymbol{M}}_u\in\mathbb{C}^{N\times K}$，$\boldsymbol{P}_u\in\mathbb{C}^{K\times 1}$，则：

$$\boldsymbol{H}_{l,u}\tilde{\boldsymbol{M}}_u\boldsymbol{P}_u=\sum_{n=1}^{N}h_{l,u,n}\left(\sum_{m=1}^{N}\Xi_{u,n,m}\sum_{k=1}^{K}\boldsymbol{M}_{u\,m}^{\;k}P_{u,k}\right),\qquad(15)$$

其中 $\Xi_{u,\;\;\;



(a) 不同 α 下的区域边界曲线。 (b) 区域1（远程三维传输优先区），六种（R, f, o）组合。 (c) 区域2（设备端二维渲染优先区），六种（R, f, o）组合。

（d）区域2（设备端二维渲染优先区）： $\bar{T}^*$ 与 δ对比。（e）区域2（设备端二维渲染优先区）： $\bar{T}^*$ 与 $Q_{\max}$对比。（f）区域1（远程三维传输优先区），视场与 δ对比。

Fig. 5：同质设置（$\mathcal{P}$1'）下预取与渲染卸载的仿真结果 .

o越大或 f 越小，曲线越陡。位于 $\delta/\alpha = V - Q_{\max}$ 处的竖直虚线标示了所有视场均可在本地提供服务的点，而该边界附近的折点与命题4中的结构相吻合。

图5c考察了“设备端渲染”严格优于“远程三维传输”的情况。与图5b相比，曲线下降得更快，尤其是对于 $\delta/\alpha \leq 20$，这反映出当“本地渲染”占主导时，预取的边际收益很高。中央处理器的能力和内容复杂度对曲线排序有显著影响；更高的 f 或更小的 o 都能大幅降低时延。竖直虚线再次与命题5给出的边界对齐。

图 5d 展示了在“设备端渲染优先”区域内， T^* 如何随内存大小 δ 变化。预取仍是提升性能的主要手段：预取更多二维视场，会增加对快速的“预取二维视场的设备端渲染”路径的依赖，从而减少“远程三维传输”。较大的 $Q_{\max}$ 能改善基准时延，而当 $Q_{\max}$ 较小时，预取带来的边际收益更为显著，这凸显了二者的互补性。

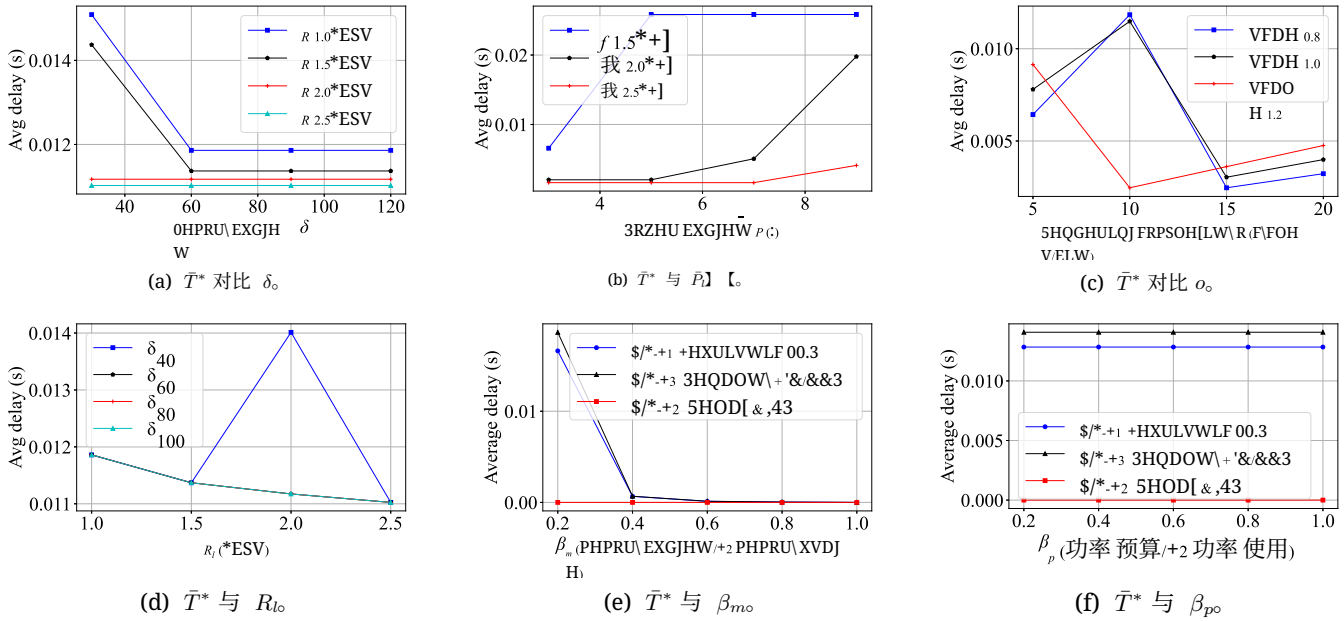
图 5e 展示了在不同 δ 下，时延如何随 $Q_{\max}$ 增大而降低。随着 $Q_{\max}$ 增大，内存匮乏（ δ 较小）时的时延降幅最大，因为更强的渲染能力降低了对“远程三维传输”的依赖；而内存充足（ δ 较大）时，曲线趋于平缓，性能瓶颈转为预取。

图5f展示了在“远程三维传输区域”内，随着内存预算 δ 增加，各服务模式分配到的视场如何变化。随着 δ 增大，视场会从“远程三维传输”转向“设备端三维预取”和“渲染”： δ 较小时，“远程传输”占主导； δ适中时，开始出现“设备端渲染”； δ 较大时，“远程三维传输”基本被“设备端三维预取”和“设备端二维渲染”取代，这与命题4中的最优结构完全一致。

B. 异构场景下预取与渲染卸载的仿真（$\mathcal{P}$*1-1*）

1）设置：我们考虑一个多用户虚拟现实系统，其中包含 $L = 4$ 个头戴式显示器，每个头戴式显示器对应 $V = 100$ 个视场。对于每个用户 l 和视场 i，异质视场大小服从 $Q_{li}^{2D} \sim \mathcal{U}[1,4]$ $Q_{li}^{3D} \sim \mathcal{U}[2,8]$, Mbit, Mbit，渲染复杂度服从 $o \sim \mathcal{U}[5,15]$ cycles/bit。视场请求概率服从指数为 $\gamma = 1.2$ 的齐夫分布。时延阈值为 $\tau = 20$ ms。本地CPU频率具有异质性，为 $f_l \in \{1.5, 2.0, 2.5, 3\}$ GHz，开关电容为 $\zeta = 10^{-27}$ J·s·cycle^{-3}。长期功率预算 P_l 在不同实验中有所变化。每次运行期间下行链路速率保持恒定，取值为 $R \in \{1.0, 1.5, 2.0, 2.5\}$ Gbps，以便评估异质信道条件下的性能。

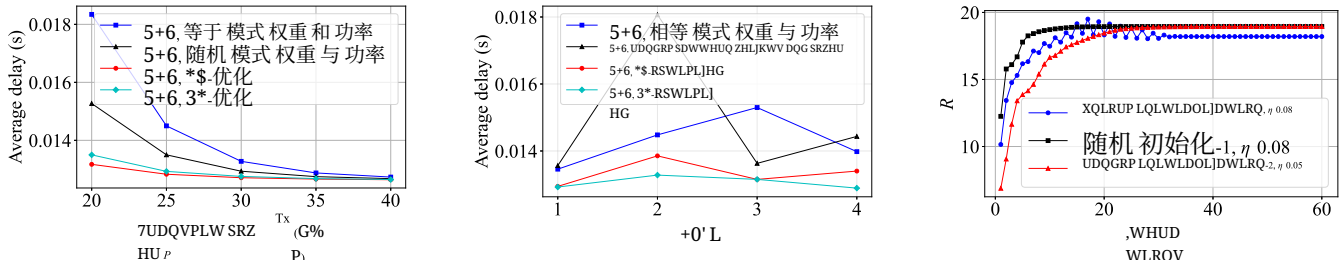
2) 结果与讨论：图 6a–6f 综合展示了关键系统参数的影响，以及三种算法在异质视场下的有效性。图 6a–6d 显示，增大内存预算 δ、功率预算 P_l、渲染能力（增大 f_l 或减小 o），或无线速率 R_l，均会降低最优平均时延 T^*；在相应的瓶颈区间（ R_l较小、 P_l较小、 o较大或 δ较小）中，增益最为显著。图 6e 和图 6f 进一步比较了内存与功率预算比 β_m 和 β_p 变化时 ALG-H1、ALG-H3 和 ALG-H2 的表现，其中 β_m （分别地， β_p）表示以松弛下界解 ALG-H2 的用量为基准归一化后的可用内存（分别为功率）。ALG-H2 给出了恒定的性能下界，而 ALG-H1 的时延最大，尤其是在 β_m 或 β_p 较小时。ALG-H3 的表现始终优于 ALG-H1，并且随着 或 增大而迅速



(a) $\bar{T}^*$ 对比 δ。 (b) $\bar{T}^*$ 与 $\bar{P}_l$】【。 (c) $\bar{T}^*$ 对比 o。

(d) $\bar{T}^*$ 与 R_l。 (e) $\bar{T}^*$ 与 β_m。 (f) $\bar{T}^*$ 与 β_p。

图 6：异构场景下预取和渲染卸载的仿真结果（$\mathcal{P}$1-1）。



(a) $\bar{T}^*$ 与 P_{Tx}。 (b) 头戴式显示器对比。 (c) PG优化器的收敛性。

图 7：RHS全息图案权重调整与馈源功率分配的仿真结果（$\mathcal{P}$1-2）。

趋近 ALG-H2界，证实所提出的松弛与DC/CCP细化能够有效利用额外的缓存和功率资源，在实际约束条件下逼近理想解。

C. 异构场景下的*RHS*波束成形仿真（$\mathcal{P}$*1-2*）

1) 系统设置：我们考虑一个配备可重构全息表面（RHS）的单个太赫兹基站（THz BS），其表面包含 $N = N_x N_y = 256$ 个元素，$N_x = N_y$ =16 个元素放置在一个配有 $K = 3$ 个馈源的均匀平面阵列（UPA）上。$Z_O = 50$ $_O = Z_O$ $_N$ $_N \in \mathbb{C}^{256\times256}$ $l_d = \lambda_c/4 =$ 0.25 $Z_{A\ u} = 80\left(\frac{l_d}{\lambda_u}\right)^2 - \frac{120}{\pi}\left(\frac{\lambda_u}{2\pi l_d}\right)$ $_A$ Ω，以及 **ZI,I**。毫米，和 ,j ln ,**Z**, $_u =$ $Z_{A\ u}$ $_N$ $_{\mathrm{R}} = [0, 0, 2]^\top$,**I**。RHS 的相位中心位于室内房间墙壁上的 **p** 米处，房间尺寸为 6 米 × 6 米 × 3 米。每个 $L = 4$ 个头戴式显示器（HMD）用户均位于 $\mathbf{p}_l = [x_l, y_l, 1.6]^\top$ 米处，其中 (x_l, y_l) 从 $[1, 5] \times [1, 5]$ 中均匀抽取。分子吸收系数 $\kappa(f_u)$ $= 3 \times 10^{-3}$ 米$^{-1}$；发射功率为 $P^{\mathrm{Tx}} = 30$ dBm，并在各子带间均匀分配；等效地，通过缩放 P_{Tx}，使信噪比从 -10 dB 扫描至 20 dB。波形频谱 $G_v(f)$ 满足 $\int |G_v(f)|^2 df = 1$。索引集 $\mathcal{Q}$ 包含来自相邻子带的主导泄漏分量，并设为 $\mathcal{Q} = \{1\}$，泄漏系数为$\alpha_{v,\ q} = 0.05$，对所有 $v \neq u$ 均成立。发射和接收天线增益分别设为$G^{\mathrm{Tx}} = 30$ dBi 和 $G^{\mathrm{Rx}} = 10$ dBi，接收机噪声功率谱密度为 $n_0 = -174$ dBm/Hz。RHS 采用结构衰减系数 $\alpha = 0.1\mathrm{m}^{-1}$、辐射效率$\beta = 0.8$，且表面波矢量与 x 轴对齐。全息图案权重 $a_{l,\ u,\ k}$ 满足 $\sum_{l,\ k} a_{l,\ u,\ k} = 1$。通过步长为 $\eta_{\mathrm{P}} = \eta_a = 0.05$、最大迭代次数为 $N_{\mathrm{iter}} = 200$的投影梯度上升法求解 $\mathcal{P}$1-2，并以加权和速率的相对改进量为依据设置停止容差 $\varepsilon = 10^{-4}$；每次梯度更新后，向量 $\{P_n\}_{n=1}^N$ 和 $\{a_{l,u,k}\}^L_{,\ l=1}{}^K_{,\ k=1}$ 均投影到各自对应的概率单纯形上。

2) 结果与讨论：图7显示，与采用等权或随机图案权重和功率的方案相比，联合优化RHS波束成形（通过遗传算法（GA）和基于PG的优化器进行全息图案权重优化和馈源功率分配）显著降低了平均时延，在中等发射功率下尤为明显。对于不同的头戴式显示器，优化方案不仅实现了更低的时延，还带来了更均衡的

性能，避免了等权设计下部分用户出现的严重性能下降。收敛曲线进一步表明，所提出的PG优化器可在几十次迭代内快速收敛，并且对不同的初始化和步长具有较强的鲁棒性。

VI. 结论

本文提出了一种框架，用于在多用户无线虚拟现实系统中联合优化预取、渲染、通过调整全息图案权重实现的RHS赋能的太赫兹波束成形，以及馈源功率分配。此外，模型还纳入了相邻单元之间的互耦效应，从而确保对波束成形过程进行更准确的优化。对于同质视场，我们推导出闭式结构策略，揭示了两个运行区域：远程三维传输优先区和设备端渲染优先区，并展示了内存、计算能力和内容复杂度如何决定最优策略。对于异质视场，我们利用缓慢变化的预取/渲染决策与用户移动情况下快速进行的RHS重新波束成形之间的时间尺度分离，提出了三阶段解决方案。基于DC/CCP的方法在内存和功率约束下实现了近乎最优的时延。我们进一步构建了一个平滑RHS波束成形问题，并开发出一种高效的投影梯度求解器，其表现始终优于未优化基线。在所有场景下进行的仿真均表明，该框架显著降低了虚拟现实时延。未来工作包括探索更多基于启发式方法的求解器、用于实时移动性处理的基于强化学习的方法，以及不进行问题分解的完全联合优化。

参考文献

[1] R. Huang、V. W. Wong 和 R. Schober，“面向智能反射面辅助多用户虚拟现实流媒体的速率分割”，*IEEE Journal on Se- lected Areas in Communications*，第 41 卷，第 5 期，第 1516–1535 页，2023 年。[2] R. Fantacci 和 B. Picano，“基于边缘的虚拟现实：通过 6G 太赫兹信道实现”，*IEEE Network*，第 35 卷，第 5 期，第 28–33 页，2021 年。[3] F. Zhinuk、A. Gaydamaka、D. Moltchanov 和 Y. Koucheryavy，“面向亚太赫兹通信的基于频谱的主动阻塞检测”，*IEEE*通信快报，第 28 卷，1707，2024 年 7，1703 第 – 页。[4] B. Picano，“异构流量和不同调度策略下 6G 太赫兹网络中虚拟现实服务的端到端时延上界”，《数学》，第 9 卷，第 14 期，第 1638 页，2021 年。[5] S. Zhao、W. Jing、F. R. Yu、X. Wen 和 Z. Lu，“太赫兹无线网络中面向增强现实任务的移动性感知计算卸载：一种离线强化学习方法”，*IEEE*车辆技术汇刊，2024 年。[6] Y. K. Tun、G. D´an、Y. M. Park 和 C. S. Hong，“太赫兹辅助、移动边缘计算赋能的空天地一体化网络中的无人机部署与资源分配联合优化”，*IEEE*移动计算汇刊，2024 年。[7] J. Du、F. R. Yu、G. Lu、J. Wang、J. Jiang 和 X. Chu，“太赫兹无线网络中移动边缘计算辅助的沉浸式虚拟现实视频流媒体：一种深度强化学习方法”，*IEEE*物联网期刊，第 7 卷，第 10 期，第 9517–9529 页，2020 年。[8] I. F. Akyildiz、C. Han、Z. Hu、S. Nie 和 J. M. Jornet，“太赫兹频段通信：重审一个老问题并探讨未来十年的研究方向”，*IEEE*通信汇刊，第 70 卷，第 6 期，第 4250–4285 页，2022年。[9] W. Jiang、Q. Zhou、J. He、M. A. Habibi、S. Melnyk、M. El-Absi、B. Han、M. Di Renzo、H. D. Schotten、F.-L. Luo 等，“面向 6G 及以后的太赫兹通信与感知：全面综述”，*IEEE*通信调查与教程，第 26 卷，第 4 期，第 2326–2381 页，2024 年。

[10] Y. Sun, Z. Chen, M. Tao, and H. Liu, “移动虚拟现实中的通信、缓存与计算：建模及权衡,” IEEE通信汇刊, vol. 67, no. 11, pp. 7573–7586, 2019年。[11] X. Liu, Y. Deng, C. Han, and M. D. Renzo, “RIS辅助太赫兹网络中交互式虚拟现实的基于学习的预测、渲染与传输,” IEEE通信领域精选期刊, vol. 40, no. 2, pp. 710–724, 2022年。[12] Y. Wang, M. Chen, Z. Yang, W. Saad, T. Luo, S. Cui, and H. V. Poor, “THz/VLC无线虚拟现实网络中用于可靠通信的元强化学习,” IEEE无线通信汇刊, vol. 21, no. 9, pp. 7778–7793, 2022年。[13] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, and P. Popovski, “太赫兹能否为无线虚拟现实提供高传输速率、可靠、低时延通信？” IEEE物联网期刊, vol. 9, no. 12, pp. 9712–9729, 2022年。[14] L. Wu, C. Zhang, B. Zhang, J. Du, and J. Qu, “迈向能效提升：将MATD3强化学习方法用于RIS辅助的无人机移动边缘计算环境中的计算卸载,” IEEE物联网期刊, vol. 12, no. 14, pp. 26 582–26 595, 2025年。[15] L. Wu, J. Qu, S. Li, C. Zhang, J. Du, X. Sun, and J. Zhou, “基于NOMA的车载移动边缘计算卸载中的注意力增强型MADDPG,” IEEE物联网期刊, vol. 11, no. 16, pp. 27 000–27 014, 2024年。[16] Y. Liu, J. Liu, A. Argyriou, L. Wang, and Z. Xu, “多小区移动边缘计算网络中的渲染感知型虚拟现实视频缓存,” IEEE车辆技术汇刊, vol. 70, no. 3, pp. 2728–2742, 2021年。[17] Z. Gu, H. Lu, and C. Zou, “移动边缘计算赋能的小蜂窝网络中虚拟现实内容交付的水平与垂直协作,” IEEE通信快报, vol. 26, no. 3, pp. 627–631, 2022年。[18] C. Zheng, J. Yin, F. Wei, Y. Guan, Z. Guo, and X. Zhang, “STC：面向移动平台、支持视场跟踪的高质量16K虚拟现实视频流传输,” IEEE视频技术电路与系统汇刊, vol. 32, no. 4, pp. 2396–2410, 2021年。[19] N. Deng, Z. He, J. Ye, B. Duinkharjav, P. Chakravarthula, X. Yang, and Q. Sun, “FOV-NeRF：面向虚拟现实的注视点感知神经辐射场,” IEEE可视化与计算机图形学汇刊, vol. 28, no. 11, pp. 3854–3864, 2022年。[20] M. R. Soleimani, Z. Nasiri-Gheidari, F. Tootoonchian, and H. Oraee, “基于磁等效电路模型的外转子旋转变压器优化设计,” IEEE传感器期刊, vol. 25, no. 3, pp. 4440–4447, 2025年。[21] Y. Ma, K. Ota, and M. Dong, “多RIS辅助太赫兹无线网络中虚拟现实服务的QoE优化,” IEEE通信领域精选期刊, vol. 42, no. 3, pp. 538–551, 2024年。[22] Y. Wang, M. Chen, Z. Yang, W. Saad, T. Luo, S. Cui, and H. V. Poor, “THz/VLC无线虚拟现实网络中用于可靠通信的元强化学习,” IEEE无线通信汇刊, vol. 21, no. 9, pp. 7778–7793, 2022年。[23] R. Deng, B. Di, H. Zhang, Y. Tan, and L. Song, “可重构全息表面赋能的多用户无线通信：幅度控制全息波束成形,” IEEE无线通信汇刊, vol. 21, no. 8, pp. 6003–6017, 2022年。[24] J. Struye, F. Lemic, and J. Famaey, “Covrage：面向移动交互式虚拟现实的毫米波波束成形,” IEEE无线通信汇刊, vol. 22, no. 7, pp. 4828–4842, 2023年。[25] R. Deng, B. Di, H. Zhang, and L. Song, “HDMA：全息图案分多址,” IEEE通信领域精选期刊, vol. 40, no. 4, pp. 1317–1332, 2022年。[26] Z. Yutong, Z. Haobo et al., “全息无线电：可重构全息表面赋能超大规模MIMO的新范式,” 《电子学报》, vol. 50, no. 12, pp. 2984–2995, 2022年。[27] J. Du, J. Gong, X. Chu, Z. Xiong, X. Chen, M. Dong, and F. R. Yu, “移动边缘计算赋能的元宇宙中的QoE保障优化：一种主动推理深度强化学习方法,” IEEE认知通信与网络汇刊, 2025年。[28] Y. Guo, X. Guo, and Y. Wang, “缓解宽带全息MIMO通信中的频率相关互耦,” 收录于2025年IEEE第36届个人、室内和移动无线电通信国际研讨会（PIMRC）, pp. 1–6, 2025年。[29] Y. Guo, X. Guo, and Y. Wang, “全息MIMO系统中互耦矩阵的快速哈希推导方法,” 收录于 2025 年 IEEE 第 36 届个人、室内和移动无线电通信国际研讨会（PIMRC）, pp. 1–6, 2025年。[30] C. Liu, K. Wang, H. Zhang, X. Li, and H. Ji, “基于视场预测、面向移动边缘计算辅助的无线虚拟现实服务的渲染图块复用方案,” IEEE网络科学与工程汇刊, vol. 10, no. 3, pp. 1709–1721, 2023年。

附录 A：命题4的证明

证明。考虑以下条件。在远程三维传输优先区 $T^{\langle 3\rangle} \geq T^{\langle 4\rangle}$ 中，以下结论成立：对于任何未预取二维内容的视场，若采用本地渲染（路径3），其时延 $T^{\langle 3\rangle}$ 不小于直接采用远程三维传输（路径4）的时延，即 $T^{\langle 4\rangle}$。因此，在最优解中，绝不会出现“将未预取二维内容的视场分配给本地渲染”的情况。这意味着，在最优点，我们可以设定$r \leq p^{2\mathrm{D}}$，这等价于 $q = \min\{p^{2\mathrm{D}}, r\} = r$。因此，所有本地渲染的视场都必须已将二维视场预取到内存中，对应于路径2。否则，应将“未预取二维内容的设备端渲染”视场改为“远程三维传输”，这样不会增加总时延，反而会使其降低。

在约束 $r \leq p^{2\mathrm{D}}$ 下，最优解为 $r^* = p^{2\mathrm{D}^*}$。当 $q = r$ 时，平均时延为：

$$\bar{T} = \frac{1}{V}\left[rT^{\langle 2\rangle} + \left(V - p^{3\mathrm{D}} - r\right)T^{\langle 4\rangle}\right]. \tag{16}$$

关于 r 的系数为：$\frac{\partial \bar{T}}{\partial r} = \frac{T^{\langle 2\rangle} - T^{\langle 4\rangle}}{V}$. 在典型的 VR 设置中（渲染时间短于下载 3D 内容，即 $T^{\langle 2\rangle} < T^{\langle 4\rangle}$），$\bar{T}$ 会随着 r 的增加而严格减小。由于 $r \leq p^{2\mathrm{D}}$，可行区域内 r 的最优值必定为 $r^* = p^{2\mathrm{D}}$。因此，我们可以将 $\bar{T}$ 改写为仅关于 $p^{2\mathrm{D}}$ 的函数。

代入内存约束，得到关于 $p^{2\mathrm{D}}$ 的线性目标函数。此时，由于 $p^{3\mathrm{D}} = \frac{\delta - p^{2\mathrm{D}}}{\alpha}$，我们有：

$$\begin{aligned}\bar{T} &= \frac{1}{V}\left[p^{2\mathrm{D}}T^{\langle 2\rangle} + \left(V - \frac{\delta - p^{2\mathrm{D}}}{\alpha} - p^{2\mathrm{D}}\right)T^{\langle 4\rangle}\right] \\ &= T^{\langle 4\rangle} - \frac{T^{\langle 4\rangle}}{V}\cdot\frac{\delta}{\alpha} + \frac{p^{2\mathrm{D}}}{V}\left(T^{\langle 2\rangle} - T^{\langle 4\rangle} + \frac{T^{\langle 4\rangle}}{\alpha}\right).\end{aligned} \tag{17}$$

注意到 $T^{\langle 4\rangle} = \frac{\alpha Q^{2\mathrm{D}}}{R}$，因此

$$T^{\langle 4\rangle} - \frac{T^{\langle 4\rangle}}{\alpha} = T^{\langle 4\rangle}\left(1 - \frac{1}{\alpha}\right) = \frac{(\alpha - 1)Q^{2\mathrm{D}}}{R}. \tag{18}$$

通常，在虚拟现实场景中，我们可以假设 $T^{\langle 2\rangle} < T^{\langle 4\rangle}\left(1 - \frac{1}{\alpha}\right)$，即，预取二维视场的设备端渲染比“将部分内存分配给三维”带来的时延降低更多。因此，括号中的系数为负，且 T 随$p^{2\mathrm{D}}$单调递减。因此，在约束条件 $p^{2\mathrm{D}} \leq \delta$，$p^{2\mathrm{D}} \leq Q_{\max}$，下，最优解为 $p^{2\mathrm{D}^*} = \min\{\delta, Q_{\max}\}$，$r^* = p^{2\mathrm{D}^*}$. 由于 $p^{3\mathrm{D}^*} = \frac{\delta - p^{2\mathrm{D}^*}}{\alpha}$，这自然给出了用于三维预取的剩余内存量。将其代入，T 的表达式可得命题中所述的最小平均时延闭式解。

□

附录B：命题5的证明

证明。在设备端二维渲染优先区中，当$T^{\langle 3\rangle} < T^{\langle 4\rangle}$时，即使其内存中没有二维预取，本地渲染也快于远程三维传输。因此，本地渲染始终有助于降低时延，无论是否已预取二维视场。因此，最优解必须满足$r^* = Q_{\max}$，这意味着全部本地内存预算都被使用。此外，若 $p^{2\mathrm{D}} \leq r$：部分本地内存没有二维预取并遵循路径3；若 $p^{2\mathrm{D}} \geq r$：所有本地视场都有二维预取并遵循路径 2。对于这两种情况，我们可以分别表示 T。我们可以证明：在 $p^{2\mathrm{D}} \leq r$所在区域中，$\bar{T}$ 随 $p^{2\mathrm{D}}$ 和 r单调递减。因此，最优解为 $p^{2\mathrm{D}^*} = \min\{r, \delta\}$ 与 $r^* = Q_{\max}$；在 $p^{2\mathrm{D}} \geq r$所在区域中，最优点位于边界 $p^{2\mathrm{D}^*} = r^*$，且 $r^* = Q_{\max}$。两个区域在边界 $p^{2\mathrm{D}} = r$处取得相同的值。因此，全局最优解为 $r^* = Q_{\max}$，$p^{2\mathrm{D}^*} = \min\{\delta, Q_{\max}\}$，$p^{3\mathrm{D}^*} = \frac{\delta - p^{2\mathrm{D}^*}}{\alpha}$。

将上述最优策略代入：

$$\bar{T} = \frac{1}{V}\left[qT^{\langle 2\rangle} + (r - q)T^{\langle 3\rangle} + \left(V - p^{3\mathrm{D}} - r\right)T^{\langle 4\rangle}\right], \tag{19}$$

其中 $q = \min\left\{p^{2\mathrm{D}}, r\right\}$。我们可以分组计算以下视场的时延：本地内存中已预取二维内容的视场（路径 2）、未从远程移动边缘计算服务器下载2D内容的视场（路径3），以及远程三维传输（路径4）。我们得到以下结果：

i：$p^{2\mathrm{D}^*}$ 沿路径 2、时延为 $T^{\langle 2\rangle}$ 的视场。ii：$Q_{\max} - p^{2\mathrm{D}^*}$ 沿路径3、时延为 $T^{\langle 3\rangle}$ 的视场。iii：$V - p^{3\mathrm{D}*} - Q_{\max}$沿路径 4、时延为 $T^{\langle 4\rangle}$ 的视场。

展开并化简后，我们得到命题中所述的 T^* 的闭式表达式。 □

附录C：$\mathcal{P}$2–2的梯度计算

将目标函数定义为

$$J(\mathbf{a}, \boldsymbol{P}) = B_g \sum_{l=1}^{L}\frac{1}{\omega_l}\sum_{u=1}^{U}\log_2\left(1 - B\left|\boldsymbol{H}_{l,u}\tilde{\boldsymbol{M}}_u\boldsymbol{P}_u\right|^2\right), \tag{20}$$

其中 $B = \frac{G^{\mathrm{Tx}}G^{\mathrm{Rx}}}{1.5\ln(5\epsilon_{l,u})\,G^{\mathrm{Tx}}G^{\mathrm{Rx}}I_u + B_g n_0}$是 $(l,\ u)$ 和 $\tilde{\boldsymbol{M}}_u \triangleq \boldsymbol{\Xi}_u\boldsymbol{M}_u$ 的常数。使用 Wirtinger微积分，$\left|\boldsymbol{H}_{l,u}\tilde{\boldsymbol{M}}_u\boldsymbol{P}_u\right|^2$ 关于 $\boldsymbol{P}_u^*$ 的梯度为 $\frac{\partial}{\partial \boldsymbol{P}_u^*}\left|\boldsymbol{H}_{l,u}\tilde{\boldsymbol{M}}_u\boldsymbol{P}_u\right|^2 = \tilde{\boldsymbol{M}}_u^H\boldsymbol{H}_{l,u}^H\left(\boldsymbol{H}_{l,u}\tilde{\boldsymbol{M}}_u\boldsymbol{P}_u\right)$.。应用链式法则，可得 J 关于 $\boldsymbol{P}_u^*$ 的梯度：

$$\nabla_{\boldsymbol{P}_u^*}J = -\frac{B_g}{\ln 2}\sum_{l=1}^{L}\frac{\eta_{l,u}}{\omega_l}\frac{\tilde{\boldsymbol{M}}_u^H\boldsymbol{H}_{l,u}^H\left(\boldsymbol{H}_{l,u}\tilde{\boldsymbol{M}}_u\boldsymbol{P}_u\right)}{1 - B\left|\boldsymbol{H}_{l,u}\tilde{\boldsymbol{M}}_u\boldsymbol{P}_u\right|^2}. \tag{21}$$

全息波束成形矩阵可表示为 $\boldsymbol{M}_u = \sqrt{\beta}\,\mathrm{diag}(\boldsymbol{m}_u)\,\boldsymbol{A}_u$，其中 $\boldsymbol{m}_u = [m_{u,1}, \ldots, m_{u,N}]^T$，$\boldsymbol{A}_u$ 的 $(n,\ k)$ 项 $\triangleq$ 且 $(e^{-\alpha\|\boldsymbol{r}_n^k\|}e^{-j\boldsymbol{k}_s\boldsymbol{r}_n^k})$，$\boldsymbol{A}_u \in \mathbb{C}^{N\times K}$. $\boldsymbol{A}_u$ 与 $a_{l,\ u,\ k}$ 无关。由于 $m_{u,\ n} = \sum_{l'=1}^{L}\sum_{k'=1}^{K}a_{l',\ u,\ k'}\,\phi_{l',\ u,\ k'}(m)$，$\frac{\partial \tilde{\boldsymbol{M}}_u}{\partial a_{l,\ u,\ ,\ k}} = \boldsymbol{\Xi}_u\frac{\partial \boldsymbol{M}_u}{\partial a_{l,\ u,\ ,\ k}} = \sqrt{\beta}\,\boldsymbol{\Xi}_u\mathrm{diag}(\boldsymbol{\phi}_{l,u,k})\,\boldsymbol{A}_{u'}$ 其中 $\boldsymbol{\phi}_{l,\ u,\ k}(m) = \left(\mathrm{Re}\{\Psi_{l,u}^{\mathrm{intf}}(\boldsymbol{r}_n^k)\} + 1\right)/2$。应用链式法则，$J$ 关于 $a_{l,\ u,\ k}$ 的梯度为 $\nabla_{a_{l,\ u,\ k}}J =$

$$-\frac{2B_g}{\ln 2}\sum_{q=1}^{L}\frac{\eta_{q,u}}{\omega_q}\frac{\mathrm{Re}\left\{\left(\boldsymbol{H}_{q,u}\tilde{\boldsymbol{M}}_u\boldsymbol{P}_u\right)^*\boldsymbol{H}_{q,u}\left(\frac{\partial \tilde{\boldsymbol{M}}_u}{\partial a_{l,u,k}}\right)\boldsymbol{P}_u\right\}}{1 - \eta_{q,u}\left|\boldsymbol{H}_{q,u}\tilde{\boldsymbol{M}}_u\boldsymbol{P}_u\right|^2}. \tag{22}$$