面向大规模MIMO的子空间选择与量化细化分裂式预编码

Yasaman Khorsandmanesh，IEEE 学生会员，Emil Bj¨ornson，IEEE 会士，以及 Joakim Jald´en，IEEE 高级会员

arXiv:2601.00616v1 [eess.SP] 2 Jan 2026

摘要——前传容量有限是大规模多输入多输出（MIMO）5G架构中的实际瓶颈。在这类架构中，基站（BS）由先进天线系统（AAS）和与之相连的基带单元（BBU）组成。传统下行链路设计将全部预编码计算放在BBU执行，并通过前传传输高维预编码矩阵，导致显著的量化损失和信令开销。本文提出一种分裂式预编码架构，将设计任务分配给AAS和BBU。AAS在本地执行子空间选择，以降低信道维度；BBU则基于由此得到的有效信道，计算经过优化的量化精细化预编码。数值结果表明，与传统单级预编码相比，所提分离式预编码策略可实现更高的和频谱效率。

索引词——分裂式预编码、大规模MIMO、子空间选择、量化感知预编码和有限前传。

I. 引言

与传统单天线系统相比，大规模多输入多输出（MIMO）通过空间复用同时服务多个用户设备（UE），显著提升了频谱效率（SE） [1]。在实际的 5G 部署中，基站（BS）通常由集中式基带处理单元（BBU）和分布式先进天线系统（AAS）组成，两者通过容量有限的前传链路连接。BBU 执行与接收上行数据和发送下行数据相关的数字基带处理。AAS 是一个集成了天线单元和射频单元的设备。这种分离式架构符合开放无线接入网（O-RAN）的原则 [2]。

随着 AAS 中天线数量的增加，传统的前传传输成为主要瓶颈。用户设备的数据符号和预编码矩阵都必须通过前传链路传输，而有限的链路容量要求进行量化。因此，在大规模MIMO系统中，以高分辨率传输预编码信号并不可行。这推动了替代架构的发展，以便在维持系统性能的同时降低前传负载。

已有研究探讨了有限分辨率预编码，其中预编码信号在前传传输之前于基带处理单元处进行量化 [3]。然而，在大规模MIMO系统中，这种方法会带来显著的前传开销。文献 [4], 提出了一种更高效的方法，开发了适用于前传受限系统的量化感知预编码方法。该方法并非对.....进行量化预编码信号，这种方法无需量化即可传输数据，并通过混合整数形式优化量化预编码矩阵，以最小化均方误差之和（MSE）。虽然这类单级预编码设计适用于小规模系统，但在大规模场景中，由于计算开销高，而且通过前传传输全维预编码矩阵成本高昂，因此并不实用，也限制了架构的灵活性。随着分布式处理技术的最新进展，AAS 现在能够在本地执行部分物理层计算 [5]。例如， [6] 提出了一种预滤波架构，由 AAS 压缩前传数据的维度，在信道状态信息（CSI）完美的情况下实现近乎无损的性能；然而，该研究关注的是信道估计，而非预编码。混合波束成形中的两级预编码结构已在 [7]，中得到探讨，但在存在前传约束的全数字下行链路系统中仍鲜有研究。一个相关思路是 [8]，中的两级降维码本，它使用统计信道状态信息进行模拟波束成形，并使用量化有效信道进行数字精细化。此外， [9] 提出了一种基于长期信道统计信息、适用于集中式无线接入网（C-RAN）架构的分层预编码策略。

在本文中，我们提出在 AAS 和基带处理单元之间分布计算任务。AAS 设置一种能够捕获主导信号方向的预编码，从而提供一个高增益、低维的有效信道；而基带处理单元则设计量化预编码，以抑制多用户干扰。这可以降低前传负载，同时保持 AAS 侧预编码的灵活性。本文的主要贡献如下：

- 我们提出一种适用于前传受限大规模MIMO的分裂式预编码架构，其中，AAS通过半酉预编码执行子空间选择，而基带处理单元则计算最优的量化精细化预编码，以最小化均方误差。
- 为最大化信号增益，我们考虑在AAS处采用基于最大比传输（MRT）的子空间选择，包括其格拉姆–施密特正交化变体；此外，我们还探索将离散傅里叶变换（DFT）预编码作为一种实用的低复杂度替代方案。
- 基带处理单元的量化感知预编码通过在量化约束下求解均方误差最小化问题，并采用基于施诺尔–欧伊希纳球形译码（SESD）的算法来实现。
- 我们给出了数值结果，证明在相同前传分辨率下，所提出的分阶段设计优于单阶段、考虑量化的基线方案，尤其是在高信噪比（SNR）下。

Y. Khorsandmanesh、E.Bj¨ornson 和 J. Jald´en 任职于瑞典斯德哥尔摩的瑞典皇家理工学院电气工程与计算机科学学院（电子邮件： {yasamank、emilbjo、jalden}@kth.se）。

本研究得到克努特和爱丽丝·瓦伦堡基金会以及瑞典创新署通过 SweWINcenter（2023-00572）提供的支持。

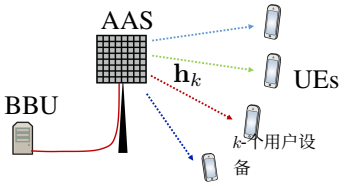


图 1：所考虑的具有分裂式预编码和前传约束的下行大规模多输入多输出系统。

II. 系统模型

我们考虑单小区大规模多输入多输出系统的下行链路，其中基站通过一个公共时频资源块向 K 单天线用户设备传输数据。基站包含两个关键组件：AAS，由大量$M \geq K$ 天线集成无线电设备组成；以及基带处理单元（BBU），二者通过容量受限的前传链路连接，如图 1 所示。基带处理单元与 AAS 之间交换的任何信号都必须量化到有限分辨率。下文中，预编码与预编码矩阵这两个术语可互换使用。

由于数据符号取自信道编码码本，因此可以通过前传链路传输，而不会产生量化误差。调制映射在 AAS 处执行。相比之下，预编码矩阵由复数值元素组成，必须在前传传输之前进行量化。该矩阵在一个信道相干间隔内保持不变。为聚焦于分裂式预编码的算法开发，我们假设 AAS 和基带处理单元具有完美的信道状态信息。

A. 分裂预编码下行传输

下行系统模型如下所示

$$\mathbf{y} = \mathbf{H}\mathbf{P}_{\mathrm{A}}\mathbf{P}_{\mathrm{B}}\mathbf{s} + \mathbf{n}, \tag{1}$$

其中，$\mathbf{y} = [y_1, \ldots, y_K]^{\mathrm{T}} \in \mathbb{C}^K$ 包含所有用户设备接收到的信号，y_k 表示第 k 个用户设备接收到的信号。下行信道矩阵 $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K]^{\mathrm{T}} \in \mathbb{C}^{K \times M}$ 为平坦衰落窄带信道。向量 $\mathbf{n} = [n_1, \ldots, n_K]^{\mathrm{T}} \in \mathbb{C}^K$ 为加性高斯白噪声，其元素独立同分布，为 $n_k \sim \mathcal{CN}(0, \sigma_0^2)$。这里，$\mathbf{s} = [s_1, \ldots, s_K]^{\mathrm{T}} \in \mathcal{O}^K$ 包含数据符号，其中 s_k 表示发送给用户设备 k 的单位功率符号，$\mathcal{O}$ 表示有限星座集合，例如正交振幅调制。AAS预编码$\mathbf{P}_{\mathrm{A}} \in \mathbb{C}^{M \times N}$ 由 AAS 本地计算，并执行子空间选择。基于可用的有效信道（信道乘以 AAS预编码），基带处理单元在低维子空间中计算基带处理单元预编码$\mathbf{P}_{\mathrm{B}} \in \mathcal{P}^{N \times K}$，并在通过前传传输之前对其进行量化。这种分离式结构降低了前传负载，因为$\mathbf{P}_{\mathrm{B}}$ 的维度低于完整的 $M \times K$ 预编码矩阵，因此在给定前传容量下可以采用更高的量化分辨率。

量化字母表 $\mathcal{P}$ 是 $\mathbb{C}$ 的离散子集：

$$\mathcal{P} = \{l_R + jl_I : l_R, l_I \in \mathcal{L}\}, \tag{2}$$

其中，$\mathcal{L} = \{l_0, \ldots, l_{L-1}\}$包含 L 个实数量化级别。每个实数维度的量化比特数为$B = \log_2(L)$。实部和虚部均采用均匀量化，具体细节见第 II-B 节。对于为简洁起见，我们仅考虑矩形量化网格，但所提出的方法很容易扩展到极坐标网格。

备注 1。假设总前传链路预算为 C_{Total}。采用所提比特拆分架构时，前传负载随 $2B_{\mathrm{Split}}NK$ 缩放，而传统单级设计需要 $2B_{\mathrm{One\text{-}stage}}MK$。由于 $K \leq N < M$，所提架构产生的前传负载更低，因此能够支持更高的量化分辨率。对于固定的总前传预算 C_{Total}，相应的比特分配满足 $\frac{M}{N} = \frac{B_{\mathrm{Split}}}{B_{\mathrm{One\text{-}stage}}}$。

AAS 计算发射向量 $\mathbf{x} = \mathbf{P}_{\mathrm{A}}\mathbf{P}_{\mathrm{B}}\mathbf{s}$，且平均发射功率必须满足

$$\mathbb{E}\left[\|\mathbf{x}\|_2^2\right] = \|\mathbf{P}_{\mathrm{A}}\mathbf{P}_{\mathrm{B}}\|_{\mathrm{F}}^2 \leq q, \tag{3}$$

其中，$\|\cdot\|_{\mathrm{F}}$ 表示 Frobenius 范数，q 表示最大平均发射功率。由于对数据符号作单位功率假设，即 $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_K$，等式成立。我们建议将 $\mathbf{P}_{\mathrm{A}}$ 选为半酉矩阵，此时功率约束可化简为 $\|\mathbf{P}_{\mathrm{B}}\|_{\mathrm{F}}^2 \leq q$。

由此得到的预编码信号向量通过下行信道无线传输至用户设备。每个用户设备通过线性均衡估计其目标数据符号，表示为$\hat{s}_k = \beta_k y_k$，其中 $\beta_k \in \mathbb{C}$ 是复数接收增益，其取值旨在使均方误差 $\mathbb{E}[|s_k - \hat{s}_k|^2]$ 最小。系统设计旨在通过最小化总均方误差，公平对待所有用户设备：

$$\mathbb{E}\left[\|\mathbf{s} - \hat{\mathbf{s}}\|_2^2\right] = \mathbb{E}\left[\|\mathbf{s} - \mathbf{B}\mathbf{y}\|_2^2\right], \tag{4}$$

其中，$\hat{\mathbf{s}} = [\hat{s}_1, \ldots, \hat{s}_K]^{\mathrm{T}}$ 包含估计的数据符号，$\mathbf{B} = \mathrm{diag}(\boldsymbol{\beta})$ 是一个收集接收机增益 $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_K]^{\mathrm{T}}$ 的对角矩阵，这些增益可以与预编码矩阵联合优化。

计算连续单阶段预编码的经典方法包括迫零（ZF）、正则化迫零（RZF）和最大比传输（MRT）[10]。其中，正则化迫零预编码最为理想，因为它能使式（4）中的均方误差最小；但在存在量化的情况下，这种方法并非最优。为考虑前传约束，我们采用量化正则化迫零（QRZF）设计，其对应的未归一化连续预编码矩阵为

$$\bar{\mathbf{P}}^{\mathrm{QRZF}} = \mathbf{H}^{\mathrm{H}}\left(\mathbf{H}\mathbf{H}^{\mathrm{H}} + \mu\mathbf{I}_K\right)^{-1}, \tag{5}$$

其中 $\mu = \frac{K}{(1-\eta)^2}\left(\frac{\sigma_0^2}{q} + \eta(1-\eta)M\right)$ 是量化后的正则化因子，η是取决于量化分辨率 [11]的失真因子。

B. 离线均匀量化器函数

我们采用一种数据驱动方法，直接根据未量化的 BBU 预编码$\mathbf{P}$的实现来优化对称中升均匀量化器的步长 Δ，使未量化的 BBU 预编码实部和虚部的期望均方误差失真最小。令 r表示一个一般的预编码元素，$\mathbb{Q}(\cdot)$表示一个实值量化器。在最大熵假设下，最优步长满足

$$\Delta^\star = \arg\min_{\Delta > 0} \mathbb{E}\left[\,|r - \mathbb{Q}(r)|^2\,\right]. \tag{6}$$

由于该期望无法以闭式形式求得，实际中通常使用由一组独立生成的 BBU 预编码实现计算出的样本均值来近似。

对于实数 x， L级中升量化器定义为

$$\mathbb{Q}(x) = \Delta\left(o(x) - \frac{L-1}{2}\right), \tag{7}$$

$$\text{for} \quad o(x) = \min\left\{\max\left\{\left\lfloor \frac{x}{\Delta} + \frac{L}{2} \right\rfloor, 0\right\}, L-1\right\}, \tag{8}$$

而最小化 $\Delta^\star$ 可通过简单的一维搜索获得。经过此次离线优化后，对于后续所有预编码矩阵，量化器均保持不变。

C. 问题表述

我们提出对AAS预编码$\mathbf{P}_{\mathrm{A}}$、量化BBU预编码$\mathbf{P}_{\mathrm{B}}$以及接收机缩放向量β进行联合优化。由于干扰抑制由BBU阶段处理，AAS旨在选择一个 N维子空间，以最大化期望信号增益。此外，由于量化BBU预编码无法完全抑制多用户干扰，AAS还可以在子空间选择阶段就引入干扰缓解，从而进一步提升性能。由此得到以下问题:

$$\begin{aligned} &\underset{\mathbf{P}_{\mathrm{A}} \in \mathbb{C}^{M\times N}}{\text{maximize}} \ \left\|\mathbf{H}\mathbf{P}_{\mathrm{A}}\right\|_{\mathrm{F}}^2 \\ &\text{subject to} \quad \mathbf{P}_{\mathrm{A}}^{\mathrm{H}}\mathbf{P}_{\mathrm{A}} = \mathbf{I}_N. \end{aligned} \tag{9}$$

为抑制AAS所选子空间内的残余干扰，BBU在满足式（3）中的总功率约束的条件下，设计一种可最小化总均方误差的量化预编码，并显式考虑前传量化感知的$\mathbf{P}_{\mathrm{B}}$。[1] 相应的优化问题为

$$\begin{aligned} &\underset{\boldsymbol{\beta}\in\mathbb{C}^K, \mathbf{P}_{\mathrm{B}}\in\mathcal{P}^{N\times K}}{\text{minimize}} \ \mathbb{E}\left\{\|\mathbf{s} - \mathbf{B}\mathbf{y}\|^2\right\} \\ &\text{subject to} \quad \|\mathbf{P}_{\mathrm{B}}\|_{\mathrm{F}}^2 \le q, \end{aligned} \tag{10}$$

其中，$\mathbf{B} = \mathrm{diag}(\boldsymbol{\beta})$ 是对角矩阵，其第 k 个对角元素记为 β_k。

III. 所提分离式预编码

下文将介绍选择 AAS 和基带处理单元预编码的不同方法。

A. AAS预编码

AAS 选择接收信号强度最大的子空间。

1）方法 I–格拉姆–施密特最大比传输：当 AAS 预编码维度满足 $N = K$ 时，我们从最大比传输方向构造一个标准正交基，作为式 (9) 的解。令 $\mathbf{h}_i$表示下行信道矩阵 $\mathbf{H}$ 的第 i 列。AAS 预编码通过对最大比传输向量进行Gram–Schmidt正交化得到:

$$\mathbf{P}_{\mathrm{A}}^{\text{GS-MRT}} = \mathrm{GS}\big(\mathbf{h}_1^*, \ldots, \mathbf{h}_N^*\big), \tag{11}$$

其中，GS(·)表示经典的格拉姆–施密特过程[12]。所得半酉基可使最大比传输向量去相关，并通过减轻干扰来改善BBU细化。虽然也可以通过选择信道矩阵的主导右奇异向量（利用瞬时奇异值分解（SVD））获得类似的正交子空间，但GS-MRT能够以显著更低的计算复杂度实现相同的性能。

2) 方法 II– MRT：采用归一化 MRT 可以获得一种用于最大化信号增益的低复杂度解决方案。所得的 AAS预编码为

$$\mathbf{P}_{\mathrm{A}}^{\text{MRT}} = \left[\frac{\mathbf{h}_1^*}{\|\mathbf{h}_1\|}, \ldots, \frac{\mathbf{h}_N^*}{\|\mathbf{h}_N\|}\right], \tag{12}$$

该方案是在不满足约束 $\mathbf{P}_{\mathrm{A}}^{\mathrm{H}}\mathbf{P}_{\mathrm{A}} = \mathbf{I}_N$ [2] 的情况下，对其进行 (9) 的计算高效近似。

3) 方法 III– 基于 DFT 的选择：前述方法要求在 AAS 处进行瞬时信道处理。为进一步降低复杂度，AAS 可以利用 DFT矩阵 $\mathbf{F} \in \mathbb{C}^{M\times M}$ 将信道变换到角度波束空间

$$\widetilde{\mathbf{H}} = \mathbf{H}\mathbf{F}. \tag{13}$$

随后，我们选择能够使 $\mathbf{H}$ 的列具有最大 ℓ_2-范数的 N DFT列~，这些列捕获了用户设备的主导角度方向。该方法以最小的计算负担 [13],提供了一个半酉子空间，但可能需要使 N 大于 K 才能获得良好的性能。

B. 基带处理单元预编码

问题（10）是非凸的，因为优化变量在目标函数中相互耦合，且其中一个变量是离散的。为解决该问题，我们采用坐标下降算法。

令有效信道为 $\mathbf{H}_{\mathrm{eff}} = \mathbf{H}\mathbf{P}_{\mathrm{A}}$。采用拉格朗日方法，式（10）中的优化问题可写为

$$\begin{aligned} &\mathfrak{L}(\mathbf{P}, \boldsymbol{\beta}, \lambda) = \lambda\big(\mathrm{tr}(\mathbf{P}_{\mathrm{B}}\mathbf{P}_{\mathrm{B}}^{\mathrm{H}}) - q\big) \\ &+ \mathrm{tr}\Big(\mathbf{I}_K - \mathbf{B}\mathbf{H}_{\mathrm{eff}}\mathbf{P}_{\mathrm{B}} - \mathbf{P}_{\mathrm{B}}^{\mathrm{H}}\mathbf{H}_{\mathrm{eff}}^{\mathrm{H}}\mathbf{B}^{\mathrm{H}} + \mathbf{B}\mathbf{H}_{\mathrm{eff}}\mathbf{P}_{\mathrm{B}}\mathbf{P}_{\mathrm{B}}^{\mathrm{H}}\mathbf{H}_{\mathrm{eff}}^{\mathrm{H}}\mathbf{B}^{\mathrm{H}}\Big), \end{aligned} \tag{15}$$

其中， λ 为拉格朗日乘子。固定 $\mathbf{P}_{\mathrm{B}}$ 和 $\mathbf{P}_{\mathrm{A}}$，通过计算维尔廷格导数可得到最优接收机增益 β_k ，其表达式为

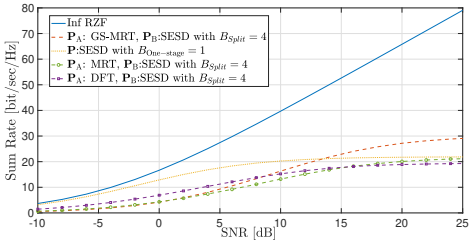
$$\beta_k^{\text{opt}} = \frac{[\mathbf{P}_{\mathrm{B}}^H\mathbf{H}_{\mathrm{eff}}^H]_{k,k}}{[\mathbf{P}_{\mathrm{B}}^H\mathbf{H}_{\mathrm{eff}}^H\mathbf{H}_{\mathrm{eff}}\mathbf{P}_{\mathrm{B}}]_{k,k} + \sigma_0^2}. \tag{16}$$

由于迭代联合优化的计算开销较大， β被固定为将 $\mathbf{P}$ 设为式（5）中的 QRZF解。对于固定的 λ，剩余问题是在 P 的量化搜索空间上求解一个整数凸规划。虽然精确求解器（例如结合 Gurobi 的 CVX）对于中等规模的系统是可行的，但一种高效的替代方法是将该问题重新表述为整数最小二乘问题，并使用 SESD

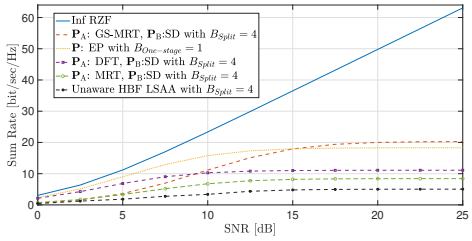
[1]对于和速率最大化，可以利用其与加权和均方误差最小化之间的标准等价关系；因此，我们采用和均方误差准则作为核心设计目标。

[2]我们注意到，这种方法可能会违反式（3）中的功率约束；但是，为了确保各算法之间进行公平比较，最终的预编码矩阵始终会进行缩放，使其严格满足功率约束。

$$\underset{\mathbf{a}_i \in \mathcal{P}^{N\times 1}, i=1,\ldots,K}{\text{minimize}} \quad \sum_{i=1}^{K} \left(\mathbf{a}_i^{\mathrm{H}} \left(\mathbf{H}_{\text{eff}}^{\mathrm{H}} \mathbf{B}^{\mathrm{H}} \mathbf{B} \mathbf{H}_{\text{eff}} + \lambda \mathbf{I}_N \right) \mathbf{a}_i - \mathbf{h}_{\text{eff}_i}^{\mathrm{T}} \mathbf{a}_i - \left(\mathbf{h}_{\text{eff}_i}^{\mathrm{T}} \mathbf{a}_i \right)^{\mathrm{H}} \right) \tag{14}$$



(a) 具有 $M = 32$ 和 $K = 8$的独立同分布瑞利衰落信道。



(b) 具有 $M = 128$ 和 $K = 8$的毫米波信道。

图 2：不同预编码方案的平均和速率随信噪比变化的曲线。

算法求解。该算法在量化点上执行结构化树搜索，并剪枝次优分支，从而以显著更低的复杂度实现最优均方误差。

按照 [4], 中的相同步骤，我们可以得到本页顶部的式（14），其中 $\mathbf{a} = \text{vec}(\mathbf{P}_{\text{B}})$ 和 $\mathbf{h}_{\text{eff}} = \text{vec}\left((\mathbf{B}\mathbf{H}_{\text{eff}})^{\mathrm{T}} \right)$。定义 $\mathbf{V} = \mathbf{H}_{\text{eff}}^{\mathrm{H}} \mathbf{B}^{\mathrm{H}} \mathbf{B} \mathbf{H}_{\text{eff}} + \lambda \mathbf{I}_N$后，可以将式（14）中目标函数各项的等价形式写为

$$\mathbf{a}_i^{\mathrm{H}} \mathbf{V} \mathbf{a}_i - \mathbf{h}_{\text{eff}_i}^{\mathrm{T}} \mathbf{a}_i - \left(\mathbf{h}_{\text{eff}_i}^{\mathrm{T}} \mathbf{a}_i \right)^{\mathrm{H}} = \| \mathbf{e}_i - \mathbf{R} \mathbf{a}_i \|_2^2 - \mathbf{e}_i^{\mathrm{H}} \mathbf{e}_i, \tag{17}$$

其中，$\mathbf{R} \in \mathbb{C}^{N\times N}$ 由 Cholesky 分解 $\mathbf{V} = \mathbf{R}^{\mathrm{H}} \mathbf{R}$ 得到，且 $\mathbf{e}_i = (\mathbf{h}_{\text{eff}_i}^{\mathrm{T}} \mathbf{R}^{-1})^{\mathrm{H}}$。由于$\mathbf{e}_i \in \mathbb{C}^{N\times 1}$ 与优化变量无关，我们可以将关于$\mathbf{a}_i$ 的子问题改写为

$$\underset{\mathbf{a}_i \in \mathcal{P}^{N\times 1}}{\text{minimize}} \quad \| \mathbf{e}_i - \mathbf{R} \mathbf{a}_i \|_2^2. \tag{18}$$

现在，对于 λ的固定值，可以使用 SESD算法求解该最小化问题。随后，我们通过二分搜索 [14] 在 λ 上查找该值，使所得解近似满足功率约束 $\mathbf{k}\mathbf{P}_{\text{B}}\mathbf{k}_{\text{F}}^2 \le q$。

IV. 数值结果

本节评估所提分裂预编码方案的性能，并将其与单级预编码进行比较，包括采用RZF预编码的无限分辨率（Inf）情况，以及量化感知单级SESD基线 [4]。由于和速率相比优化中使用的总均方误差能够更直观地衡量通信效率，因此采用和速率作为性能指标。平均和速率计算如下：

$$\sum_{k=1}^{K} \mathbb{E} \left[\log_2 \left(1 + \frac{\left| [\mathbf{H}\mathbf{P}]_{k,k} \right|^2}{\sum_{i=1, i\neq k}^{K} \left| [\mathbf{H}\mathbf{P}]_{k,i} \right|^2 + \sigma_0^2} \right) \right], \tag{19}$$

其中，$\mathbf{P}$ 表示各方法得到的预编码矩阵，期望值是通过蒙特卡洛仿真对信道实现取平均得到的。我们考虑一种归一化场景，其中所有用户设备具有相同的大尺度衰落系数 γ。假设接收机拥有完美的信道状态信息。接下来，我们考虑两种信道模型：独立同分布瑞利衰落和毫米波信道。

除非另有说明，基准配置为：一个配备 $M = 32$ 根天线的基站为 $K = 8$ 个用户设备提供服务，前传比特分辨率为 $B_{\text{Split}} = 4$ 和 $N = K$。如备注 1 所述，在此配置下，单级预编码的分辨率应设为 $B_{\text{one-stage}} = 1$ [4]。所有用户设备均经历相同的信噪比，该信噪比定义为单天线、单用户设备情况下的 SNR= $\frac{q\gamma}{\sigma_0^2}$。

图 2 展示了不同预编码方案的平均和速率随信噪比变化的情况。在图 2a 中，信道矩阵 $\mathbf{H}$ 遵循独立同分布瑞利衰落模型。“Inf RZF”曲线随信噪比（以 dB 为单位）几乎线性增长，因为它未经过量化。在实际方案中，GS–MRT 在高信噪比下实现了最高的和速率。在低信噪比下，其行为与最大比传输相似，随后由于残余用户间干扰而趋于饱和。单阶段 SESD 基线 [4] 的斜率相近，但在高信噪比下，其性能低于前传容量相同的 GS–MRT 分离式架构。在低信噪比下，单阶段 SESD 与 GS–MRT 之间存在权衡：分离方法的计算复杂度较低，而单阶段 SESD 的和速率更高。最大比传输和 DFT有源天线阵列预编码的速率明显较低，且很早就趋于饱和，反映出它们的干扰抑制能力有限。由于前传约束，基带处理单元预编码无法消除所有干扰；因此，有源天线阵列预编码必须选择近正交子空间，以实现稳健性能。在独立同分布瑞利衰落中，DFT结构无法提供足够正交的波束，这解释了其性能下降。总的来说，这些结果表明，尽管最大比传输和 DFT 等简单分离方案因复杂度低而颇具吸引力，但量化感知或结构优化设计（例如单阶段 SESD 或 GS–MRT 分离方案）能够带来显著增益，尤其是在高信噪比下。图 2b 将比较扩展至毫米波信道，采用经过修改的 Saleh–Valenzuela模型，其中包含具有 10 dB 莱斯因子和 64 个子载波的 4 抽头莱斯衰落信道。随着 M 增加，单阶段方法的复杂度将变得难以承受；因此，对于这一基线，我们采用期望传播（EP）方法 [15]。总体性能趋势与瑞利衰落情形一致，但有一个显著例外：在毫米波信道下，DFT

在高信噪比下优于最大比传输，这是因为它天然契合稀疏角度结构。我们还纳入了“非感知 HBF LSAA”基准 [16], ，其中混合波束成形设置中的基带预编码在未考虑量化效应的情况下进行量化。该曲线显示出的明显性能损失凸显了基带处理单元进行量化感知预编码的重要性。

图 3 展示了基于 DFT 的分离式预编码下，不同系统配置的平均和速率随信噪比变化的情况。信道为独立同分布瑞利衰落。结果表明，性能深受基带处理单元预编码维度 N 和量化分辨率 B_{Split} 的影响。同时增大维度和量化分辨率（例如， $N = 2K$ 和$B_{\text{Split}} = 4$）可显著提升速率，使其接近所提出的 GS–MRT 的性能。相比之下，采用 $B_{\text{Split}} = 1$ 的配置很快便达到饱和，这表明粗量化会严重限制高信噪比下的多用户干扰抑制。 $N = M$ 配合 $B_{\text{Split}} = 1$ 的情况可获得中等速率，表现与 SESD 单级基线相近，体现了设计简洁性与量化精度之间的权衡。对于固定的 B_{Split}，从$N = K$ 增至 $N = 2K$ 可在低至中等信噪比范围内带来明显增益，但高信噪比下仍会饱和，因为量化引起的失真最终会占据主导地位。该图表明，联合增大有效信道维度和量化分辨率对于提升吞吐量至关重要；而在严格的前传约束下，只改善其中一个参数会带来递减的收益。从计算角度来看，采用基于 SESD 的基带处理单元预编码的单级架构，对于 $0 \le \gamma \le 1$ [4] 需要 $O(KL^{2\gamma M})$ 次运算。在分离式架构中，基带处理单元复杂度降低至$O(KL^{2\gamma N})$；尽管仍呈指数级，但在 N 较小时显著降低。AAS 侧复杂度取决于子空间选择方法：GS–MRT 需要 $O(MN^2)$，而采用快速傅里叶变换的基于 DFT 的实现可将计算量降至 $O(M \log M)$，且仅需执行一次。总之，最大比传输和 DFT 的复杂度最低，但性能损失明显；GS–MRT 分离式设计则在计算效率与可达速率之间实现了良好平衡。

五、结论

本文提出了一种面向前传受限下行大规模 MIMO 系统的量化感知分裂式预编码方案。AAS 通过选择相关子空间来执行第一级预编码，而基带处理单元则采用量化感知 MMSE 预编码来优化传输，该问题通过 SESD 求解。仿真结果表明，在两种方案均满足相近前传分辨率约束的情况下，所提设计在高信噪比下优于传统单级方法。GS-MRT 和 SESD 单级基线方案的性能接近理想水平，而 DFT 和最大比传输方案则是实用的低复杂度替代方案。对于基于 DFT 的方法，增加基带处理单元预编码维度和量化分辨率可进一步提升和速率。总体而言，所提架构在前传效率与计算复杂度之间实现了颇具吸引力的权衡，兼顾了

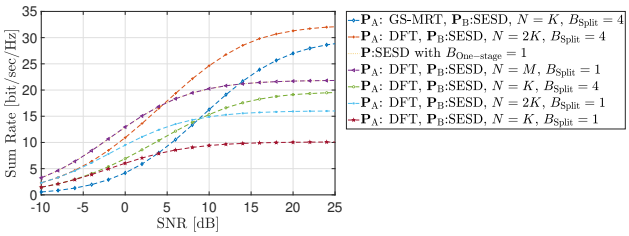


图 3：采用 $M = 32$ 和 $K = 8$的不同配置下，平均和速率随信噪比的变化。

可扩展性与量化鲁棒性，适用于大规模 MIMO 部署。

参考文献

[1] E. Bj¨ornson、J. Hoydis 和 L. Sanguinetti，《大规模MIMO网络：频谱效率、能效与硬件效率》，《信号处理基础与趋势》，第 11 卷，第 3–4 期，第 154–655 页，2017 年。[2] M. Peng、C. Wang、V. Lau 和 H. V. Poor，《前传受限云无线接入网络：见解与挑战》，《IEEE 无线通信汇刊》，第 22 卷，第 2 期，第 152–160 页，2015 年。[3] O. Casta˜neda、S. Jacobsson、G. Durisi、T. Goldstein 和 C. Studer，《毫米波大规模多用户MIMO系统的有限字母维纳滤波器预编码》，载于 2019 年第 53 届阿西洛马信号、系统与计算机会议。IEEE，2019 年，第 178–183 页。[4] Y. Khorsandmanesh、E. Bj¨ornson 和 J. Jald´en，《采用前传量化的多用户MIMO优化预编码》，《IEEE 无线通信汇刊》，第 22 卷，第 11 期，第 7102–7115 页，2023 年 11 月。[5] V. Karthikeyan、K. Gopalakrishnan 和 S. S. Raj，《6G先进天线技术》，载于《面向现代社会的以人为本的6G赋能技术融合》。Elsevier，2025 年，第 71–86 页。[6] W. Chang、T. Xie、F. Zhou、J. Tian 和 X. Zhang，《面向大规模MIMO、具有压缩链路数据速率的预滤波 C-RAN 架构》，载于 2016 年 IEEE 第 83 届车辆技术会议（VTC 春季会议）。IEEE，2016 年，第 1–6 页。[7] F. Sohrabi 和 W. Yu，《面向大规模天线阵列的混合数字与模拟波束成形设计》，《IEEE 信号处理选题期刊》，第 10 卷，第 3 期，第 501–513 页，2016 年。[8] A. Al mradi、M. Matthaiou 和 V. F. Fusco，《多用户MIMO系统的两阶段有限反馈混合预编码》，载于 ICC 2020——2020 年 IEEE 国际通信会议（ICC）。IEEE，2020 年，第 1–7 页。[9] J. Kang、O. Simeone、J. Kang 和 S. Shamai，《全维MIMO C-RAN系统的分层下行预编码》，《IEEE 车辆技术汇刊》，第 66 卷，2016 年，第 3 期，第 2170–2182 页。[10] E. Bj¨ornson、M. Bengtsson 和 B. Ottersten，《最优多用户发射波束成形：一个难题及其简单的解结构 [讲义],》，《IEEE 信号处理杂志》，第 31 卷，第 4 期，第 142–148 页，2014 年。[11] L. Fan、S. Jin、C.-K. Wen 和 H. Zhang，《采用低分辨率模数转换器的大规模 MIMO系统上行可达速率》，《IEEE 通信快报》，第 19 卷，第 12 期，第 2186–2189 页，2015 年。[12] K. Matsumura 和 T. Ohtsuki，《多用户MIMO下行系统中采用 Gram–Schmidt 正交化的正交波束成形》，《Eurasip 无线通信与网络期刊》，第 2011 卷，第 1 期，第 41 页，2011 年。[13] O. El Ayach、S. Rajagopal、S. Abu-Surra、Z. Pi 和 R. W. Heath，《毫米波MIMO系统中的空间稀疏预编码》，《IEEE 无线通信汇刊》，第 13 卷，第 3 期，第 1499–1513 页，2014 年。[14] R. L. Burden 和 J. D. Faires，《2.1 二分算法》，《数值分析》，第 46–52 页，1985 年。[15] H. Wang、A. Kosasih、C.-K. Wen、S. Jin 和 W. Hardjawana，《超大规模大规模MIMO的期望传播检测器》，《IEEE 无线通信汇刊》，第 19 卷，第 3 期，第 2036–2051 页，2020 年。[16] F. Sohrabi 和 W. Yu，《面向毫米波OFDM大规模天线阵列的混合模拟与数字波束成形》，《IEEE 通信选题领域期刊》，第 35 卷，第 7 期，第 1432–1443 页，2017 年。