

利用深度学习在时分复用大规模 MIMO 系统下完成信道感知和混合预编码

Kareem M. Attiah, Foad Sohrabi, and Wei Yu

Department of Electrical and Computer Engineering

University of Toronto, Toronto, Ontario M5S 3G4, Canada

Emails: {kattiah, fsohrabi, weiyu}@ece.utoronto.ca

摘要：本文提出了一种利用深度学习在射频链路数量有限，且运行于时分双工模式下的毫米波大规模多输入多输出系统中完成信道感知、下行链路混合模拟和数字波束成形的方法。传统的下行链路预编码设计基于两步：第一步通过信道感知矩阵，来接收上行导频，并以此估计高维信道；然后根据检测到的信道来设计预编码矩阵。可是，这种方法并不总是最佳方法。比如当导频长度很短时，此方法就不是最优方法。本文表明，通过直接从接收的导频设计模拟感测和下行链路预编码矩阵，而无需中间信道估计步骤，可以显著提高整体系统性能。具体来说，我们提出了一种信道感知和混合预编码方法，该方法将导频阶段分为模拟阶段和数字训练阶段。在第一阶段利用深度神经网络来设计上行链路信道感知和下行链路模拟波束形成器。随后，我们调整模拟波束形成器，并基于等效低维信道，设计数字预编码器。本文提出的深度学习架构的一个关键特点是：它分解为独立并行的单用户深度神经网络，所以可以推广到具有任意用户数的系统。数值比较表明，与基于信道恢复的对应方法相比，本文所提出的方法所需的训练开销明显更少，并且可以利用相对较少的导频来获得和具有完整信道状态信息的系统一样的性能。

I. 介绍

大规模多输入多输出 (MIMO) 通信由于其出色的抵抗严重路径损耗的能力，已被公认为下一代蜂窝网络中毫米波 (mmWave) 系统的关键促成因素之一 [1]。大规模 MIMO 技术在时分双工 (TDD) 系统中的使用特别吸引人，因为当它的信道互易性与上行链路信道估计结合使用时，MIMO 系统中基站 (BS) 的天线数量是可扩展的 [2]。但是，运用大规模 MIMO 技术的一个关键问题是由射频 (RF) 组件带来的高功率消耗和高成本，这使部署常规的全数字波束成形 MIMO 架构变得不可行。为了减轻这种硬件限制，在 [3], [4] 中提出了所谓的混合波束成形架构，其中全数字波束成形器由模拟波束成形器代替，该模拟波束成形器使用低成本的模拟移相器实现，然后再采用低维数字波束成形器。

即使在理想的信道状态信息 (CSI) 假设下，混合预编码器的设计通常也是一项艰巨的任务，因为通过移相器阵列实现模拟预编码器会对设计参数施加非凸的恒定模数约束，这使得优化预编码器的计算量变得非常大。更重要的是，混合架构还严重限制了上行信道的感知能力。然而，大多数当前最新的混合预编码方案仍遵循将预编码过程分成两个独立步骤的常规通信系统设计方法：(i) 信道感测和估计，以及 (ii) 下行链路预编码设计。在信道估计步骤中，通常利用毫米波信道的空间相关性 [5]-[8]。特别地，由于 mmWave 信道在角域中具有稀疏表示，所以现有的信道估计方法通常使用压缩感知技术，来基于例如均方误差度量去恢复信道模型参数。随后，假设估计到的 CSI 是完美的，BS 再构造模拟和数字预编码矩阵 [9]-[11]。

本文的主要观察结果是，上述的分离信道估计和混合预编码方法不一定是最佳方法。尤其是当导频序列长度不足以确保 CSI 准确恢复时，该方法性能不是最优。在这种情况下，BS 处仅可获得 CSI 的噪声估计，并且现有的预编码方法可能无法提供良好的性能。本文的主要贡献在于指出，旨在针对某个任意距离度量（例如，均方误差）来重构通道的常规两步设计，不一定是设计最大速率的混合预编码器的最佳方案。我们可以绕过信道估计，直接从基带接收到的导频中设计预编码矩阵，以获得整体系统性能的显著改善。

本文提出使用深度学习方法，来基于接收的导频直接设计信道感知和混合预编码矩阵。但是，由于混合预编码器由模拟和数字组件组成，因此以数据驱动的方式同时进行直接设计是非常重要的。但是，训练复杂度和缺少对用户数量的通用处理让这个设计变得不现实。在本文中，我们提出了一种替代的基于深度学习的方案，该方案克服了上述限制，但仍绕过了基于模型的显式信道估计。为此，我们建议将上行链路导频训练分为两个单独的阶段，其中在第一阶段中，我们采用深度神经网络 (DNN)，该网络结合了感测矩阵的设计并将基带导频观测结果直接映射到模拟预编码矩阵中。在第二阶段，我们利用观察结果，一旦确定了模拟波束成形矩阵，就可以将端到端混合预编码系统

转换为低维全数字系统，从而设计出数字预编码矩阵。数值模拟表明，尤其是在短导频长度的情况下，这种方案是有利的，并且相对于基于常规信道恢复的方案而言，可以获得明显更好的性能。

值得注意的是，许多现有工作已经提出在具有混合架构的 mmWave 大规模 MIMO 系统中使用 DNN 来取代关键设计组件，例如用 DNN 取代信道估计 [12]，取代具有完美/不完美 CSI 的预编码设计 [13]-[15]，或取代以上两方面 [16]。但是，所有这些工作都采用了传统的信道估计和预编码分离方法，并未充分利用 DNN 的潜力来设计端到端下行链路预编码系统。

II. 系统模型

A. 下行链路传输

我们考虑一种 TDD 大规模 MIMO 系统，该系统在逐块衰落且频率平坦 mmWave 环境中运行，在该环境中，具有 M 个天线和 N_{RF} 个 RF 链的 BS 为 K 个单天线用户提供服务。使用 $\mathbf{s} \in \mathbb{C}^K$ 表示要在下行链路中发送给用户的预期信号的集合，第 k 个用户的复基带接收信号由下式给出：

$$\begin{aligned} y_k &= \mathbf{h}_k^H \mathbf{V} \mathbf{s} + n_k, \\ &= \mathbf{h}_k^H \mathbf{v}_k s_k + \sum_{j \neq k} \mathbf{h}_k^H \mathbf{v}_j s_j + n_k, \end{aligned} \quad (1)$$

其中 $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_K] \in \mathbb{C}^{M \times K}$ 是下行链路预编码矩阵， $\mathbf{h}_k \in \mathbb{C}^M$ 是从 BS 到用户 k 的复数下行链路信道增益的向量，而 $n_k \sim \mathcal{CN}(0, \sigma^2)$ 是用户 k 处的加性高斯白噪声。此外，我们假设 BS 采用的混合波束成形架构具有有限数量的 RF 链， N_{RF} ，其中 $K \leq N_{RF} \leq M$ 。因此，整个预编码矩阵必须采用以下形式：

$$\mathbf{V} = \mathbf{V}_{RF} \mathbf{V}_D, \quad (2)$$

其中 $\mathbf{V}_D \in \mathbb{C}^{N_{RF} \times K}$ 是低维数字预编码矩阵， $\mathbf{V}_{RF} \in \mathbb{C}^{M \times N_{RF}}$ 是移相器的模拟预编码矩阵，其项满足恒定模数约束，即 $[\mathbf{V}_{RF}]_{ij} = e^{j\phi_{ij}}$ ，其中 j 是虚数单位。最后，我们通过要求 $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_K$ 和 $\|\mathbf{V}\|_F^2 \leq P_D$ 对发射信号施加功率约束。

B. 信道模型

我们考虑散射体数量有限的毫米波传播环境。因此，第 k 个用户的信道向量 \mathbf{h}_k 遵循由 L_p 主导路径组成的稀疏信道模型：

$$\mathbf{h}_k = \frac{1}{\sqrt{L_p}} \sum_{\ell=1}^{L_p} \alpha_{\ell,k} \mathbf{a}_t(\theta_{\ell,k}), \quad (3)$$

其中 $\alpha_{\ell,k} \sim \mathcal{CN}(0, 1)$ 是 BS 和用户 k 之间的第 ℓ 条路径的复数增益， $\theta_{\ell,k}$ 是对应的离开角 (AoD)，而在 $\mathbf{a}_t(\cdot)$ 是阵列响应向量。假设在 BS 处有 M 个天线单元组成的均匀线性阵列，我们有：

$$\mathbf{a}_t(\theta) = \left[1, e^{j\frac{2\pi}{\lambda} d \sin(\theta)}, \dots, e^{j\frac{2\pi}{\lambda} d(M-1) \sin(\theta)} \right]^T, \quad (4)$$

其中， λ 是波长， d 是天线间隔。

C. 上行链路中基于导频的信道感应

我们假设 BS 没有信道的先验感知，并且必须基于 TDD 操作中的信道互易性，通过上行链路导频训练来得到 CSI 的估计（显式或隐式）[17]。为此，我们考虑由 L 个时间帧组成的上行链路导频训练，其中用户在每个时间帧中重复发送 K 个相互正交的导频序列。由于 RF 链有数量限制，BS 必须使用以下形式的模拟感测（或组合）矩阵来感测信道：

$$\mathbf{W}_{RF}^{(\ell)} \in \mathbb{C}^{N_{RF} \times M}, \quad \ell \in \{1, \dots, L\}, \quad (5)$$

其一元项必须满足恒模约束 $[\mathbf{W}_{RF}^{(\ell)}]_{ij} = e^{j\psi_{ij}}$ 。

令 $\mathbf{X} \in \mathbb{C}^{K \times K}$ 为一元矩阵，其行对应于用户发送的正交导频，则第 ℓ 个时间帧在 BS 处的 $N_{RF} \times K$ 接收信号由下式给出：

$$\mathbf{R}^{(\ell)} = \sqrt{P_U} \mathbf{W}_{RF}^{(\ell)} \mathbf{H} \mathbf{X} + \mathbf{W}_{RF}^{(\ell)} \mathbf{N}^{(\ell)}, \quad \ell \in \{1, \dots, L\}, \quad (6)$$

其中 $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K] \in \mathbb{C}^{M \times K}$ 和 $\mathbf{N}^{(\ell)} = [\mathbf{n}_1^{(\ell)}, \dots, \mathbf{n}_K^{(\ell)}] \in \mathbb{C}^{M \times K}$ 是所有用户的信道和噪声矩阵， P_U 是在单个时间帧内为每个用户分配的功率。通常 $N_{RF} \ll M$ ，因此我们需要多个时间帧（即 $L > 1$ ）来捕获用户信道。

由于导频是正交的，因此我们有 $\mathbf{X}\mathbf{X}^H = \mathbf{I}_K$ ，因此 BS 可以将接收到的信号乘以 \mathbf{X}^H 来获得：

$$\tilde{\mathbf{Y}}^{(\ell)} = \sqrt{P_U} \mathbf{W}_{RF}^{(\ell)} \mathbf{H} + \mathbf{Z}^{(\ell)}, \quad \ell \in \{1, \dots, L\}, \quad (7)$$

其中 $\mathbf{Z}^{(\ell)} = \mathbf{W}_{RF}^{(\ell)} \mathbf{N}^{(\ell)} \mathbf{X}^H$ 是第 ℓ 个时间帧中对应的噪声矩阵。用 $\tilde{\mathbf{y}}_k^{(\ell)}$ 表示 $\tilde{\mathbf{Y}}^{(\ell)}$ 的第 k 列，我们可以得到：

$$\tilde{\mathbf{y}}_k^{(\ell)} = \sqrt{P_U} \mathbf{W}_{RF}^{(\ell)} \mathbf{h}_k + \mathbf{z}_k^{(\ell)}, \quad \ell \in \{1, \dots, L\}. \quad (8)$$

通过垂直堆叠 $\tilde{\mathbf{y}}_k^{(\ell)}$ 可获得在整个导频传输阶段中，由于第 k 个用户的传输而产生的 $N_{RF}L$ 接收矢量（在乘以 \mathbf{X}^H 后）：

$$\tilde{\mathbf{y}}_k \triangleq \begin{bmatrix} \tilde{\mathbf{y}}_k^{(1)} \\ \vdots \\ \tilde{\mathbf{y}}_k^{(L)} \end{bmatrix} = \sqrt{P_U} \begin{bmatrix} \mathbf{W}_{RF}^{(1)} \\ \vdots \\ \mathbf{W}_{RF}^{(L)} \end{bmatrix} \mathbf{h}_k + \mathbf{z}_k, \quad (9)$$

其中 $\mathbf{z}_k \in \mathbb{C}^{N_{RF}L}$ 是相应的噪声矢量。

D. 设计目标

现在可以如下说明信道感知和混合预编码问题。假设根据 (3) 描述的信道模型，我们需要确定感知矩阵 $\{\mathbf{W}_{\text{RF}}^{(\ell)}\}$ ，从而可以将模拟和数字预编码矩阵 \mathbf{V}_{RF} 和 \mathbf{V}_{D} 设计为接收导频信号 $\tilde{\mathbf{y}}_k$ 的函数。更具体地说，我们选择下行链路总速率作为总体目标函数，即

$$R = \sum_k R_k, \quad (10)$$

其中 R_k 是用户 k 可达到的速率，由下式给出：

$$R_k = \log_2 \left(1 + \frac{|\mathbf{h}_k^H \mathbf{v}_k|^2}{\sum_{j \neq k} |\mathbf{h}_k^H \mathbf{v}_j|^2 + \sigma^2} \right). \quad (11)$$

注意，总速率表达式对感知矩阵 $\{\mathbf{W}_{\text{RF}}^{(\ell)}\}$ 的选择的依赖性隐含的。特别地，从 (9) 中我们观察到，感知矩阵 $\{\mathbf{W}_{\text{RF}}^{(\ell)}\}$ 的设计至关重要，因为它们起着汇总用户信道信息的重要作用，因此会直接影响我们构造的混合预编码矩阵的质量。

III. 建议的编码方案

A. 动机

常规的通信系统设计通常依赖于分析方法来分别执行信道估计和预编码设计。在这样的方案中，通常假设一些信道模型。然后根据一些距离度量来估计模型参数。然而，问题在于，这些距离度量可能与使整体系统性能最大化的最终目标不完全匹配。我们认为这种可能的失配是限制传统系统设计的关键，但是在其他方面是必要的，以确保传统的设计问题是可处理的。现在，数据驱动技术的出现使不同的范例成为可能。不再需要遵守上述分离，即可以完全避开 CSI 估计，并从基带接收的导频直接设计混合预编码器可能更好。换句话说，由于 DNN 作为通用函数逼近器的强大功能，现在可以实现同时包含信道感知/估计和混合预编码的端到端设计。更准确地说，设计问题变成了学习感知矩阵 $\{\mathbf{W}_{\text{RF}}^{(\ell)}\}$ 以及直接映射下面的矩阵 $\mathcal{F}(\cdot)$ ：

$$(\mathbf{V}_{\text{RF}}^*, \mathbf{V}_{\text{D}}^*) = \mathcal{F}(\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_K), \quad (12)$$

其中 $\tilde{\mathbf{y}}_k$ 是接收到的导频，而 \mathbf{V}_{RF}^* 和 \mathbf{V}_{D}^* 是速率最大的预编码矩阵。再次强调，从 (9) 开始， $\tilde{\mathbf{y}}_k$ 对感知矩阵 $\{\mathbf{W}_{\text{RF}}^{(\ell)}\}$ 具有依赖性。

但是，这种方法的复杂性较高。实际上，由于模拟预编码矩阵和数字预编码矩阵之间的乘法交互作用，使得即使在用户数量少的时候，这种 DNN 的训练也相当繁琐。此外，我们希望第 II-A 节中描述的多用户系统能为不同数量的用户提供服务，但是 DNN 要输出数字预编码矩阵，其

输出维数必须等于用户数量。为了支持这种动态操作，理想方法是基于每个用户设计 DNN。为此，我们在下面开发了一种新颖的 DNN 架构。所提出的架构与 [16] 中的架构有些相似，但是关键的区别在于我们的方法绕过了中间信道估计步骤。

B. 混合预编码的学习方法

本文提出的基于学习的方案将模拟预编码矩阵和数字预编码矩阵的设计解耦。具体来说，我们建议将 L 帧的总体导频训练阶段划分为 L_a 帧的模拟训练阶段和 L_d 帧的数字训练阶段，且 $L = L_a + L_d$ 。在模拟训练阶段，使用 DNN 将在最初的 L_a 帧中接收到的基带信号直接映射到模拟预编码矩阵中。在数字训练阶段，模拟波束形成器是固定的，并且基于在剩下的 L_d 帧中接收到的导频信号获得数字预编码矩阵。

1) 模拟训练阶段：我们将在最初的 L_a 个时间帧里从第 k 个用户接收到的导频表示为 $\tilde{\mathbf{y}}_k^a \triangleq \left[(\tilde{\mathbf{y}}_k^{(1)})^T, \dots, (\tilde{\mathbf{y}}_k^{(L_a)})^T \right]^T$ ，然后由 (9)，我们可以得到：

$$\tilde{\mathbf{y}}_k^a = \sqrt{P_{\text{U}}} \mathbf{W}_{\text{RF}}^a \mathbf{h}_k + \mathbf{z}_k^a, \quad (13)$$

其中 $\mathbf{W}_{\text{RF}}^a \triangleq \left[(\mathbf{W}_{\text{RF}}^{(1)})^T, \dots, (\mathbf{W}_{\text{RF}}^{(L_a)})^T \right]^T$ 。在模拟训练阶段，我们旨在设计 DNN 以学习感知矩阵 \mathbf{W}_{RF}^a 以及从 $\{\tilde{\mathbf{y}}_k^a\}_{k=1}^K$ (直接) 映射到 \mathbf{V}_{RF} 。但是，我们不是将全部 DNN 用于所有 k 个用户，我们提出的架构将整个 DNN 分解为 K 个并行单用户 DNN，这些 DNN 学习感知矩阵 \mathbf{W}_{RF}^a 。还分解为 K 个从 $\tilde{\mathbf{y}}_k^a$ 到 \mathbf{V}_{RF} 第 k 列的并行映射 $\mathcal{G}^{(k)}(\cdot)$ ，即

$$\mathbf{v}_{\text{RF}}^{(k)} = \mathcal{G}^{(k)}(\tilde{\mathbf{y}}_k^a), \quad \forall k \in \{1, \dots, K\}. \quad (14)$$

本文的想法是模拟预编码矩阵 \mathbf{V}_{RF} 由 K 个用户中每个用户的模拟预编码器 \mathbf{v}_{RF}^k 组成，可以分别学习。我们在本节中假设，当系统完全加载 $N_{\text{RF}} = K$ 时，此方法有效。

理想情况下，我们想学习映射 $\mathcal{G}^{(k)}(\cdot)$ 以实现最大化系统求和速率的目标。但是系统求和速率取决于尚未设计的数字波束形成器。因此，我们仅根据模拟预编码器估算每个用户可达到的速率，而无需考虑干扰。这是有道理的，因为随后的数字阶段将使用迫零 (ZF) 预编码器 (具有相等的功率分配) 来消除干扰。在这种情况下，损失函数变为

$$\mathcal{L} = - \sum_{k=1}^K \log_2 \left(1 + \frac{P_{\text{D}}}{MK\sigma^2} \left| \mathbf{h}_k^H \mathbf{v}_{\text{RF}}^{(k)} \right|^2 \right). \quad (15)$$

请注意，上述损失函数由 K 个独立项组成，每个项都是第 k 个用户的 $\mathcal{G}^{(k)}(\cdot)$ 的输出的函数。一个关键的结果是，整

个训练过程现在可以分解为学习 K 个独立的单用户映射。

我们还可以将信道感知矩阵 \mathbf{W}_{RF}^a 的训练合并到 DNN 架构中。从 (13) 中可以看出, \mathbf{W}_{RF}^a 可以作为可训练线性层合并到每个单用户 DNN 中。注意, 感知矩阵对所有用户都是公用的。因此, 通过将此线性层的参数与单用户 DNN 结合在一起, 可以轻松设计出公用的信道感知矩阵。

为了进一步方便训练, 我们还可以在用户之间共享 $\mathcal{G}^{(k)}(\cdot)$ 的映射的参数, 称之为 $\mathcal{G}(\cdot)$ 。在这种情况下, 所有单用户 DNN 现在具有相同的结构, 即以用户信道作为输入, 并以相应的模拟预编码器作为输出, 并且具有相同的损耗函数和相同的参数。于是, 使用一批独立的信道来训练整个 DNN 现在等同于仅训练一个具有 K 倍信道采样的单用户 DNN。关键优势在于, 通过采用 K 个合理训练的单用户 DNN, 我们可以得到适用于拥有任意用户数量的系统的通用体系结构。

2) 数字训练阶段: 确定 \mathbf{V}_{RF} 后, 我们将继续确定 \mathbf{V}_{D} 。一个关键的理论是, 由于信道互易性, 设计出的下行链路模拟预编码矩阵也必须是合适的上行链路信道感知矩阵。此外, 如果将模拟感知矩阵设为模拟预编码矩阵的厄米特矩阵 (Hermitian), 即 $\mathbf{W}_{\text{RF}} = \mathbf{V}_{\text{RF}}^H$, 那么端到端低维全数字系统也将具有上行链路-下行链路互易性。可以通过剩下的 L_d 个导频获得此低维信道。

更准确地说, 由 (1) 和 (2) 给出的下行链路等效低维信道 $\mathbf{H}_{\text{eq}} \in \mathbb{C}^{N_{\text{RF}} \times K}$ 为:

$$\mathbf{H}_{\text{eq}} = \mathbf{V}_{\text{RF}}^H \mathbf{H}. \quad (16)$$

我们可以通过设:

$$\mathbf{W}_{\text{RF}}^{(\ell)} = \mathbf{V}_{\text{RF}}^H, \quad \forall \ell \in \{L_a + 1, \dots, L\}, \quad (17)$$

来在上行链路估计 \mathbf{H}_{eq} 。并通过观察 (7), 即

$$\tilde{\mathbf{Y}}^{(\ell)} = \sqrt{P_U} \overbrace{\mathbf{V}_{\text{RF}}^H \mathbf{H}}^{\mathbf{H}_{\text{eq}}} + \mathbf{Z}^{(\ell)}, \quad \ell \in \{L_a + 1, \dots, L\}, \quad (18)$$

注意 $\tilde{\mathbf{Y}}^{(L_a+1)}, \dots, \tilde{\mathbf{Y}}^{(L)}$ 恰好是通过上述等效信道重复发送的导频。现在, 我们可以得到 \mathbf{H}_{eq} 的线性最小均方误差 (LMMSE) 的估计量。假设 BS 天线元件得到的是不相关的信道 (即 $\mathbb{E}[\mathbf{h}_k \mathbf{h}_k^H] = \mathbf{I}_M, \forall k$), 那么等效信道的 LMMSE 估计由下式给出:

$$\hat{\mathbf{H}}_{\text{eq}} = \frac{\sqrt{P_U}}{P_U L_d + \sigma^2} \sum_{\ell=L_a+1}^L \tilde{\mathbf{Y}}^{(\ell)}. \quad (19)$$

最后, 我们应用传统的线性预编码方案 (例如 ZF) 来确定数字预编码器:

$$\mathbf{V}_{\text{D}} = \gamma_{\text{ZF}} \hat{\mathbf{H}}_{\text{eq}} (\hat{\mathbf{H}}_{\text{eq}}^H \hat{\mathbf{H}}_{\text{eq}})^{-1}, \quad (20)$$

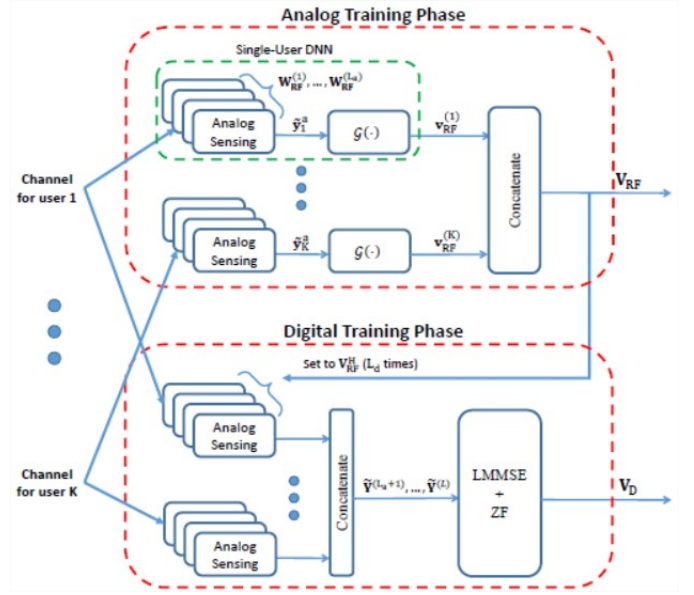


图 1: 设计模拟预编码矩阵和数字预编码矩阵的方法框图

其中 γ_{ZF} 是确保满足功率约束的标量。图 1 说明了提出的方案。

需要注意的是本文提出的体系结构克服了上一节中讨论的限制, 特别是:

- 本文所提出的方法直接从接收的导频确定混合预编码矩阵, 而不必首先估计高维信道。
- 与 (12) 中介绍的直接设计相比, 映射矩阵 $\mathcal{G}(\cdot)$ 现在非常容易学习。
- 单用户 DNN 的输出维度不是用户数量的函数, 因此, 该方案可推广到具有任意用户数量的系统。

C. 单用户 DNN 架构

现在我们讨论实现映射 $\mathcal{G}(\cdot)$ 和感知矩阵 \mathbf{W}_{RF}^a 的单用户 DNN 的细节。具体而言, 图 2 中针对第 k 个用户提出的体系结构包括传感层, 中间层和相位映射层。在 DNN 的离线训练中, 信道和噪声实现 $\{\mathbf{h}_k, \mathbf{n}_k\}$ 被馈送到第一层 (即传感层), 并且相应的模拟预编码向量 $\mathbf{v}_{\text{RF}}^{(k)}$ 由最后一层 (即相位映射层) 产生。训练后, 我们使用 K 个相同的单用户 DNN 进行操作, 其中在上行链路导频阶段将传感层用作模拟感知矩阵, 并将接收到的导频信号 $\tilde{\mathbf{y}}_k^a$ 馈入中间层以产生下行链路模拟预编码器。

1) 传感层: 可以将传感矩阵视为 DNN 的可训练线性层, 其中第 k 个用户的信道实现以及噪声作为输入。确实, 从 (13) 中可以看出, \mathbf{W}_{RF}^a 只是线性神经网络层中的权重, 整体上具有附加的单位模量约束。顺便一提, 现有的深度学习库 (例如 TensorFlow [18]) 允许创建具有可训练参数的自定义层。在这种情况下, 我们将可训练参数作为感知矩阵 \mathbf{W}_{RF}^a 的相位。

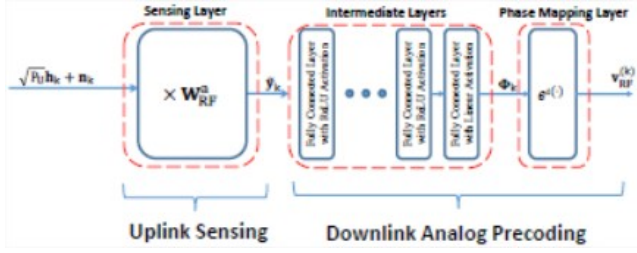


图 2: 用于上行链路感知和下行链路模拟预编码设计的单用户 DNN 的体系结构

2) 中间层: 第二级的输入是基带接收的导频矢量 $\tilde{\mathbf{y}}_k^a$, 而该级的输出是包含相应模拟预编码矢量的相位的矢量, 用 Φ_k 表示。我们考虑一个包含 R 层组成的中间级, 除了最后一层具有线性激活外, 其余每一层都与整流线性单元 (ReLU) 激活完全相连。在数学上, 中间层的输出为:

$$\Phi_k = \mathbf{W}_R \mathcal{R}(\dots \mathcal{R}(\mathbf{W}_1 \tilde{\mathbf{y}}_k^a + \mathbf{b}_1) \dots) + \mathbf{b}_R, \quad (21)$$

其中 $\mathcal{R}(\cdot) = \max(\cdot, 0)$ 是 ReLU 的激活函数, 而 $\{\mathbf{W}_r, \mathbf{b}_r\}$ 是第 r 层的一组可训练参数 (即权重和偏差)。

3) 相位映射层: 该层的输出是用户 k 的模拟预编码向量, 因此我们需要强制执行恒模约束。为了实现这一点, 我们采用 [13] 中的思想, 其中 (21) 的中间层的输出为模拟组合矢量的相位。在这种情况下, 该层使用进入式转换 $[\mathbf{v}_{\text{RF}}^{(k)}]_j = e^{j[\Phi_k]_j}$, 其中 $[\cdot]_j$ 表示向量的第 j 个入口。可以使用现有深度学习库中的 *Lambda* 层来实现这种转换层。

4) 损失函数: 如前所述, 第 k 个单用户 DNN 的损失函数可以近似视为单用户可达到的速率

$$\mathcal{L}^{(k)}(\mathbf{h}_k, \mathbf{v}_{\text{RF}}^{(k)}) = -\log_2 \left(1 + \frac{P_D}{MK\sigma^2} \left| \mathbf{h}_k^H \mathbf{v}_{\text{RF}}^{(k)} \right|^2 \right). \quad (22)$$

IV. 数值结果

在本节中, 我们将本文所提出的混合预编码方法的性能与现有的几种方案进行比较。为此, 我们考虑在 TDD 操作中的单小区大规模 MIMO 系统, 其中 BS 配备了均匀的线性阵列, 有 $M = 64$ 根发射天线, 且天线之间的距离为半波长。每个用户的信道遵循第 II-A 节介绍的稀疏信道模型, $L_p = 6$, 并且路径的 AoDs 遵循区间为 $[0, 2\pi]$ 的均匀分布。最后, 我们将上行链路和下行链路的信噪比设置为 $\text{SNR} = 10 \log P/\sigma^2 = 20\text{dB}$, 其中 $P = P_U = P_D$ 。在我们的模拟中, 我们考虑与以下方案进行比较:

1) 具有完整 CSI 的混合设计: 该设计遵循文献 [10] 中的工作, 假设该信道是 BS 完全已知的, 并被用作其他方案的基准。在这种设计中, 模拟预编码矩阵是通过匹配信道矩阵的相位获得的 (即 $[\mathbf{V}_{\text{RF}}]_{ij} = [\mathbf{H}]_{ij}/|[\mathbf{H}]_{ij}|$), 而数字预编码矩阵则是通过对等效信道做 ZF 算法得到的。

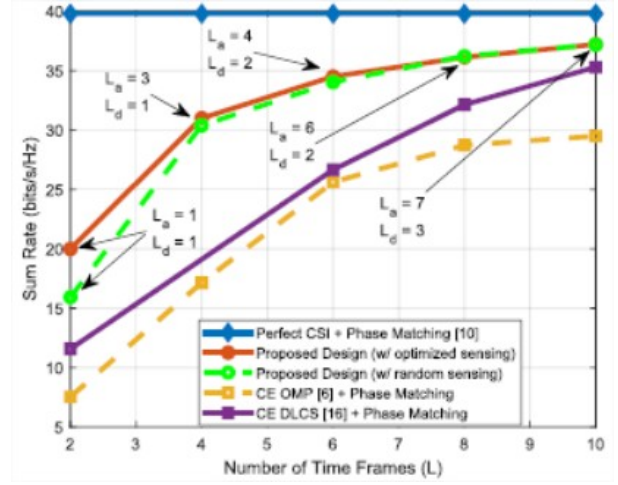


图 3: 当 $K = N_{\text{RF}} = 4$ 时, 针对不同的 L 值, 本文提出的方法与现有方案的性能比较。

2) 使用正交匹配追踪 (OMP) 的具有不完整 CSI 的混合设计 [6]: 在该设计中, 使用 OMP 算法估计每个用户的高维信道的参数, 然后使用上述 [10] 中的方案从恢复的信道中确定预编码矩阵。

3) 使用深度学习压缩感知 (DLCS) 的具有不完整 CSI 的混合设计 [16]: 为了克服 OMP 解析信道路径的复杂方式, [16] 的作者开发了一种受 OMP 启发的 DNN 架构来恢复信道参数。我们在这里使用相同的 DNN 架构来估计信道, 并根据 [10] 中的设计方案从恢复的信道中设计预编码矩阵。

我们使用 TensorFlow 实现了图 2 中提出的 DNN 和 DLCS。对于拟议的 DNN, 我们将中间层的数量设置为 $R = 3$, 其中密集层的宽度分别为 2048、1024 和 512。为了更快地收敛, 每个密集层之前都有批量规范化层。另外, 对于第一层, 我们首先按照区间为 $[0, 2\pi]$ 的均匀分布来随机设置模拟预编码矩阵的相位。对于 DLCS, 我们考虑一个宽度为 1024、512、256 和 128 的 4 层密集网络。对这两个网络的训练是分批进行的, 并且需要几个周期。每个批次包含 500 个样本, 每个周期有 20 个批次。我们在上行链路训练阶段处理信道和噪声分布, 因此我们可以生成训练 DNN 所需的任意数量的数据样本。

在图 3 中, 对 $N_{\text{RF}} = K = 4$, 我们针对不同的导频帧 L 的数量绘制了总速率。对于所提出的方法, 用箭头表示 L_a 和 L_d 。我们可以看到, 本文提出的方案明显优于 OMP 和 DLCS。例如, 与 $L = 10$ 的 DLCS 相比, 本文所提出的方法在 $L=8$ 时可实现整个 CSI 系统总和速率的 90% 以上, 从而表明与传统方案相比, 导频开销节省了 20% 以上。这支持了我们的主要观点, 即在训练开销有限的混合系统中, 信道估计和预编码设计的分离并不是最佳的。但是我们在这里注意到, 在长导频长度的限制下, 传统的分离方案可以做得很好。最后, 我们观察到, 仅在导频长度非常短时,

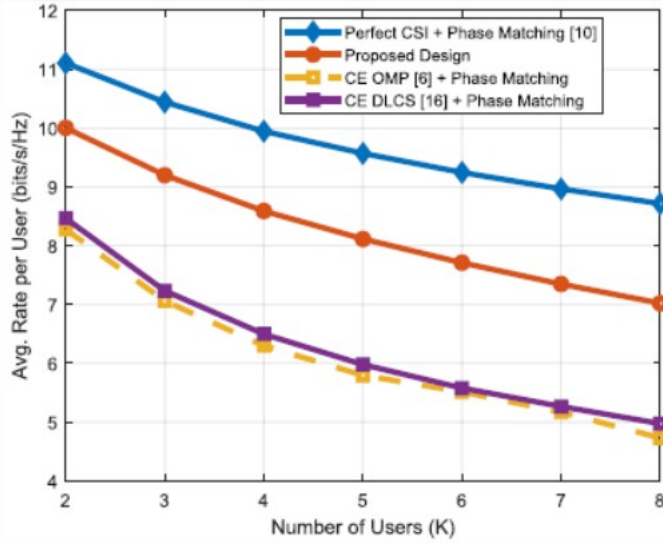


图 4: 该方案对用户数量的普遍性。这里, 我们将 N_{RF} 设为 8, K 从 2 个用户变到 8 个用户。

优化信道感知矩阵的效果才显著。

接下来, 我们研究关于用户数量的普遍性。为此, 我们设 $L=3$, $N_{\text{RF}}=8$, 并且 K 从 2 变到 8。对于本文的方案, 我们设 $L_a=2$, $L_d=1$, 并设计了一个用于 $K=8$ 情况的单用户 DNN。如果 $K<8$, 我们做出简化选择, 仅使用 K 个 RF 链进行下行链路预编码。注意, 还必须做出简化选择, 以使常规方案能够正常工作。在图 4 中, 我们绘制了所有四种方案的用户平均速率与用户数量之间的关系。我们观察到, 与基于信道恢复的方法相比, 我们的方法能够支持有关用户数量的通用操作, 同时保持明显更好的性能。

V. 总结

本文介绍了毫米波 TDD 大规模 MIMO 系统中模拟预编码和数字预编码矩阵的设计。本文所提出的基于学习的预编码方案可用于任何数量的用户, 并且通过直接从接收到的导频构造预编码矩阵而无需进行估计高维信道的中间步骤克服了现有方案的局限性。数值评估表明, 与基于信道恢复的方案相比, 频谱效率有了显著提高。 [1]–[18]

参考文献

- [1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5G cellular: It will work!,” *IEEE Access*, vol. 1, p. 335–349, May 2013.
- [2] T. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, p. 3590–3600, Nov. 2010.
- [3] X. Zhang, A. Molisch, and S.-Y. Kung, “Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection,” *IEEE Trans. Signal Process.*, vol. 53, no. 11, p. 4091–4103, Oct. 2005.

- [4] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, p. 1499–1513, Mar. 2014.
- [5] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, p. 831–846, Oct. 2014.
- [6] J. Lee, G. Gil, and Y. H. Lee, “Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications,” *IEEE Trans. Commun.*, vol. 64, no. 6, p. 2370–2386, June 2016.
- [7] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, “Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, p. 8954–8958, Sept. 2018.
- [8] F. Bellili, F. Sofrabi, and W. Yu, “Generalized approximate message passing for massive MIMO mmWave channel estimation with Laplacian prior,” *IEEE Trans. Commun.*, vol. 67, no. 5, p. 3205–3219, May 2019.
- [9] F. Sofrabi and W. Yu, “Hybrid digital and analog beamforming design for large-scale antenna arrays,” *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, p. 501–513, Jan. 2016.
- [10] L. Liang, W. Xu, and X. Dong, “Low-complexity hybrid precoding in massive multiuser MIMO systems,” *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, p. 653–656, Dec. 2014.
- [11] A. Li and C. Masouros, “Hybrid precoding and combining design for millimeter-wave multi-user MIMO based on SVD,” *IEEE Int. Conf. Commun. (ICC), Paris, France*, p. 1–6, May 2017.
- [12] M. Borgerding, P. Schniter, and S. Rangan, “AMP-inspired deep networks for sparse linear inverse problems,” *IEEE Trans. Signal Process.*, vol. 65, no. 16, p. 4293–4308, Aug. 2017.
- [13] T. Lin and Y. Zhu, “Beamforming design for large-scale antenna arrays using deep learning,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, p. 103–107, Jan. 2020.
- [14] A. M. Elbir, “CNN-based precoder and combiner design in mmWave MIMO systems,” *IEEE Commun. Lett.*, vol. 23, no. 7, p. 1240–1243, July 2019.
- [15] A. M. Elbir and A. K. Papazafeiropoulos, “Hybrid precoding for multiuser millimeter wave massive MIMO systems: A deep learning approach,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, p. 552–563, Jan. 2020.
- [16] W. Ma, C. Qi, Z. Zhang, and J. Cheng, “Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO,” *IEEE Trans. Commun.*, vol. 68, no. 5, p. 2838–2849, Jan. 2020.
- [17] G. S. Smith, “A direct derivation of a single-antenna reciprocity relation for the time domain,” *IEEE Trans. Antennas Propag.*, vol. 52, no. 6, p. 1568–1577, June 2004.
- [18] M. Abadi et al., “TensorFlow: Large-scale machine learning on heterogeneous distributed systems,” Mar. 2016. [Online]. Available: <https://arxiv.org/abs/1603.04467>.