WiFo-MUD：面向异构多用户解调器的无线基础模型

杨宗辉，*IEEE* 学生会员，高世健，*IEEE* 会员，蔡雪松，*IEEE* 高级会员，程翔，*IEEE* 会士，杨柳青，*IEEE* 会士。

arXiv:2601.00612v1 [eess.SP] 2 Jan 2026

摘要—多用户信号解调对无线通信至关重要，直接影响传输的可靠性和效率。然而，现有解调器在通用多用户环境中的表现不佳：传统解调器难以兼顾准确率与复杂度，而基于深度学习的方法在异构配置下缺乏适应性。尽管扩散模型已被引入解调领域，但其实用灵活性仍然有限。为解决这些问题，本文提出 WiFo-MUD，这是一种用于多用户解调的通用扩散基础模型。该模型通过定制主干网络对齐用户间的信噪比失衡，并执行条件去噪。此外，本文设计了一种通信感知一致性蒸馏方法和一种动态用户分组策略，以提升推理效率。WiFo-MUD 在大规模异构数据集上取得了最先进的结果，展现出高效的推理能力，以及在不同系统配置下的强泛化能力。

索引词—**无线基础模型、多用户解调、扩散模型、一致性蒸馏。**

I. 引言

MULTI-user 信号解调是数字通信系统设计中的一项关键任务[1–3]。准确的信号解调可确保数据传输的可靠性并降低重传率，从而提升整体频谱效率 [4, 5]。因此，它是自动驾驶、低空经济和具身智能等新兴应用的性能与可行性的基石 [6]。

由于发射符号具有离散性，解调问题被认为是NP困难的。尽管最大似然（ML）检测在理论上具有最优性能，但其复杂度会随着发射天线数量和调制阶数的增加而呈指数增长，这对实际系统而言是难以承受的。线性解调器，例如最小二乘（LS）和线性最小均方误差（LMMSE） [7]，，是一类将接收信号直接与滤波矩阵相乘以获得估计值，然后进行判决的方法。尽管线性方法因其低复杂度而受到青睐，但在大规模多输入多输出（MIMO）系统以及采用高阶调制时，会造成显著的性能下降。与线性解调器相比，非线性解调具有更好的性能，但复杂度也有所增加 [8–11]。例如，球形译码算法通过施加一定限制来缩小最大似然检测的搜索空间。基于消息传递的方法 [9, 10] 在理想高斯信道下能够实现近乎最优的性能，但对偏离这一假设的情况非常敏感。随后提出了正交近似消息传递（OAMP） [11] ，以放宽对信道特性的限制。然而，上述解调器未能在计算复杂度与解调精度之间实现权衡：线性解调器效率较高但并非最优，而非线性方法则需要付出高昂的计算代价才能接近最优性能。

深度学习的兴起推动了面向物理层传输的数据驱动设计的发展[12, 13]。神经网络能够学习从信道分布和接收信号到发送符号的复杂映射，为信号解调提供了颇具前景的工具，并实现了更好的平衡。 [14, 15] DnCNN 等模型 [16]直接采用卷积神经网络（CNN）进行信号解调，但准确率有限。其他方法，如 RE-MIMO [17] 和 OAMP-Net [18], , 则将基于模型的迭代与神经网络相结合，以提升性能。然而，它们在相关多用户信道上的泛化能力仍然不足。此外，所有这些基于学习的模型在异构系统配置和变化的数据分布之间都表现出较差的泛化能力。在天线数量、调制方式或分布外信道条件发生变化时，其性能会显著下降，这凸显了其鲁棒性方面的关键不足。

近年来，扩散模型在计算机视觉和自然语言处理等领域展现出的强大生成与建模能力，推动了其在无线通信中的应用 [19, 20]。这些模型有望刻画复杂的信道与噪声分布，从而提升信号解调精度 [21–25]。例如， [21] 开发了一种基于退火朗之万动力学（ALD）的解调算法，该算法利用退火机制促进信道相关评分网络的训练，并迭代优化信号估计值。此外， [22] 提出了一种基于扩散模型的信道与信号联合估计方法，[23] 通过预测器-校正器

Z. Yang、X. Cai 和 X. Cheng 任职于北京大学电子学院光子学与通信国家重点实验室，中国北京 100871（电子邮件：yzh22@stu.pku.edu.cn；xuesong.cai@pku.edu.cn；xiangcheng@pku.edu.cn）。S. Gao 任职于香港科技大学（广州）物联网学域，中国广州 511400（电子邮件：shijiangao@hkust-gz.edu.cn）。L. Yang 任职于香港科技大学（广州）物联网学域和智能交通学域，中国广州；同时任职于香港科技大学电子及计算机工程学系，中国香港特别行政区（电子邮件：lqyang @ust.hk）。

方案提升了解调精度。 [25] 中的研究利用扩散模型实现了端到端的鲁棒图像语义传输，但这些方法未能纳入真实的信道分布。

尽管基于扩散的方法为信号解调提供了新颖的视角，但其实际部署仍面临重大挑战。首先，这些方法忽视了真实的多用户场景，而用户间干扰（IUI）以及用户间不平衡的信噪比（SNR）可能会显著降低性能。此外，这些方法在适应动态变化方面灵活性有限，例如用户数量或信道维度的变化，也未能充分利用多用户信道相关性。因此，现有方法缺乏在异构多用户环境中的泛化能力，导致解调精度下降。近期研究已探索无线基础模型 [26–28], ，旨在提升性能，同时保持在异构系统中的适应性。然而，现有模型均无法解决这一特定的信号解调任务。

为解决上述局限性，本文提出 WiFo-MUD，这是一种面向异构系统配置下多用户信号解调的基础模型赋能框架。该框架首先通过专用对齐模块缩小多用户信噪比差距，随后采用无线扩散Transformer架构对粗略估计进行去噪，并将无线信号的结构先验融入其中。借助基于扩散的预训练策略，WiFo-MUD 能够有效捕获多样的信道和噪声分布，从而较现有方法提高解调精度。利用 Transformer骨干网络的灵活性和定制的嵌入模块，该模型还支持根据用户数量和天线数量的变化进行动态适配。尽管如此，扩散模型中的迭代反向过程可能导致较高的推理延迟。尽管此前的研究（如 [29, 30] ）引入了近似技术或动量项以降低复杂度，但仍需进行大量迭代才能达到令人满意的性能。为克服这一瓶颈，我们进一步设计了一种通信感知的一致性蒸馏方案，无需迭代细化即可大幅降低时延，满足实时需求。此外，我们还采用用户分组策略来缓解用户间干扰。在大规模异构信号解调数据集上进行预训练和蒸馏后，WiFo-MUD 在全样本和零样本场景中均展现出优异性能，从而能够在配置多样的多用户系统中实现可靠、高效的通信。

O我们的贡献可总结如下 ws:

- 我们提出了 WiFo-MUD，这是一种用于异构多用户信号解调的基础模型，融合了多用户对齐模块与无线扩散 Transformer 主干网络。通过基于扩散的预训练，该模型在不同系统配置下实现了高准确率和强泛化能力。
- 为提升推理效率，我们设计了一种通信感知一致性蒸馏技术以实现精准的单步推理，并设计了一种动态用户分组机制来缓解用户间干扰。
- 在大规模异构数据集上进行的大量实验表明，我们的模型在全样本和零样本场景下均超越了现有基准，同时实现了低时延推理并保持了鲁棒泛化能力。

本文其余部分安排如下。第二节介绍系统模型和问题表述。第三节简要介绍扩散模型的背景知识，作为预备内容。第四节阐述用于多用户信号解调的无线基础模型，第五节介绍针对模型推理的进一步优化。第六节展示仿真结果，第七节总结全文。

符号说明： a、 $\boldsymbol{a}$和 $\boldsymbol{A}$分别表示标量、向量和矩阵。$(\cdot)^{\mathrm{T}}$、$(\cdot)^{\mathrm{H}}$、Tr(·)、rank(·)、 $\|\cdot\|_2$ 和 $\|\cdot\|_{\mathrm{F}}$分别表示转置、共轭转置、迹、秩、二范数和弗罗贝尼乌斯范数。 $\mathbb{E}(\cdot)$表示期望。$\mathbb{P}(\cdot)$表示概率。 $\mathcal{CN}(m,\sigma^2)$表示均值为 m、协方差为 σ^2的复高斯分布。 $\mathbb{R}$ 和 $\mathbb{C}$分别表示实数集和复数集。 $|z|$和 $\angle z$ 分别表示复数 z 的模和相位。 $\bigcap$ 和 $\bigcup$ 表示交集和并集， $\emptyset$ 表示空集。

II. 系统模型与问题表述

A. 系统设置

我们考虑一个小区中的 U 个活跃用户，其中第 u 个用户配备 N_{t_u} 根天线，并同时向配备 N_r 根天线的基站（BS）发送信号。基站接收到的信号表示为

$$\boldsymbol{y}=\sum_{u=1}^{U}\boldsymbol{H}_u\boldsymbol{x}_u+\boldsymbol{n}, \tag{1}$$

其中， $\boldsymbol{H}_u\in\mathbb{C}^{N_r\times N_{t_u}}$ 表示从第 u 个用户到 BS 的上行信道矩阵， $\boldsymbol{n}\sim\mathcal{CN}(\boldsymbol{0},\sigma_n^2\boldsymbol{I})$ 表示方差为 σ_n^2 的复圆对称高斯噪声向量。向量 $\boldsymbol{x}_u\in\mathcal{X}^{N_{t_u}}$ 包含来自第 u个用户的发射符号，其中 $\mathcal{X}$ 是由星座点构成的有限集。可以采用各种调制方案，并将发射符号归一化，使每根天线的平均功率为单位值，即 $\mathbb{E}\left[\|\boldsymbol{x}_u\|_2^2\right]=N_{t_u}$。假设所有用户使用相同的星座集，且每个符号被用户选取的概率相同。此外，假设已知接收噪声 σ_n 的统计特性，并假设在基站处估计时隙级信道状态信息（CSI），记为 $\boldsymbol{H}_u$。信道估计的均方误差表示为 $\sigma_{H,u}^2=\mathbb{E}\left[\|\hat{}\boldsymbol{H}_u-\boldsymbol{H}_u\|_{\mathrm{F}}^2\right]$。我们使用估计信道和接收信号来恢复发射符号。

B. 异构多用户信号解调

ˆ 给定基站处针对 $u=1,\cdots,U$ 的观测信道 $\boldsymbol{H}_u$ 以及观测信号 $\boldsymbol{y}$，目标是使用

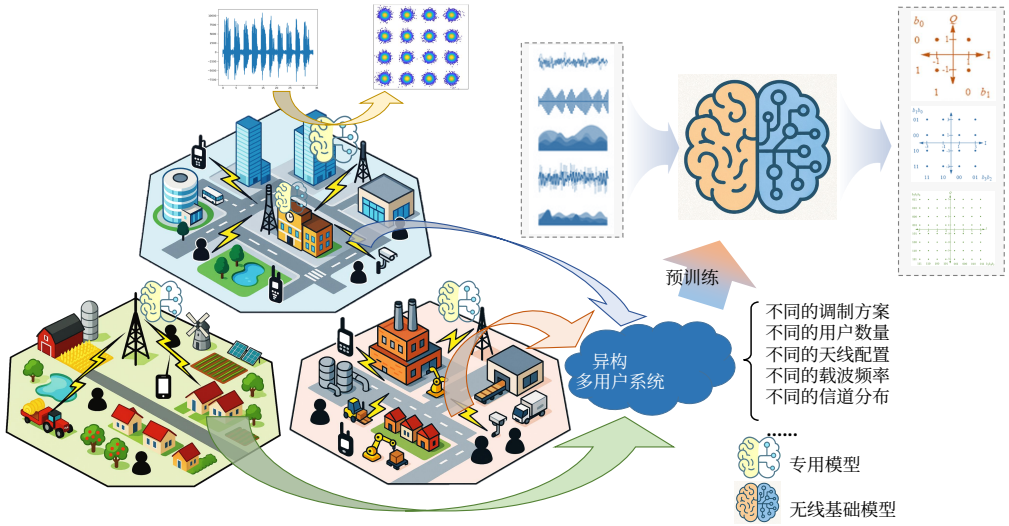


图 1：异构多用户信号解调模型范式的比较。

能够最大化后验概率的解调器 $\mathcal{D}$ 进行恢复。该 is,

$$\begin{aligned}\hat{\boldsymbol{x}}_{\mathrm{MAP}} &= \mathcal{D}(\boldsymbol{y}, \{\hat{\boldsymbol{H}}_u\}_{u=1}^U) \\ &= \underset{\{\boldsymbol{x}_u\}_{u=1}^U \in \mathcal{X}}{\operatorname{argmax}} \; p(\boldsymbol{x}|\boldsymbol{y}, \{\hat{\boldsymbol{H}}_u\}_{u=1}^U).\end{aligned} \tag{2}$$

在星座点服从均匀先验分布且存在高斯测量噪声的假设下，该最大后验（MAP）解调器可简化为最大似然解调器（ML demodulator）。具体而言，式 (2) 中的估计问题可归结为求解以下优化问题：

$$\hat{\boldsymbol{x}}_{\mathrm{ML}} = \underset{\{\boldsymbol{x}_u\}_{u=1}^U \in \mathcal{X}}{\operatorname{argmin}} \; \|\boldsymbol{y} - \sum_{u=1}^{U} \hat{\boldsymbol{H}}_u \boldsymbol{x}_u\|_2^2, \tag{3}$$

约束条件是每个 $\boldsymbol{x}_u$ 均属于有限离散星座。该问题已知为 NP 困难问题，因此在实际系统中，精确计算 $\hat{\boldsymbol{x}}_{\mathrm{ML}}$ 难以实现；其组合复杂度与严格的时延和硬件约束根本冲突，尤其是在用户数量和调制阶数增加时。现有方法试图以可控的计算复杂度解决这一问题，但难以在准确率和时延之间实现令人满意的平衡。此外，许多方法未能充分考虑含噪观测和异构配置的影响，例如不同的 U、N_{t_u}、N_r 和 $\mathcal{X}$。

III. 扩散模型背景

首先，我们简要介绍扩散模型的背景知识，以此作为构建本文框架的基础。扩散模型 [19, 20] 是一类通过逆转逐步加噪过程来学习数据分布的生成模型。通过随机微分方程（SDE）可以对这类模型提供统一的理解视角，随机微分方程描述了扩散过程的连续时间动力学。

A. 扩散过程的随机微分方程（SDE）表述

前向扩散过程可以表示为以下形式的伊藤随机微分方程：

$$d\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, t)dt + g(t)d\boldsymbol{w}, \tag{4}$$

其中，$\boldsymbol{w}$ 是标准维纳过程，$\boldsymbol{f}(\boldsymbol{x}, t)$ 是漂移系数，而 $g(t)$ 是扩散系数。这种表述概括了各种离散时间扩散过程。

与公式相对应的反向时间生成过程由反向随机微分方 (4) 程表示为

$$d\boldsymbol{x} = \left[\boldsymbol{f}(\boldsymbol{x}, t) - g(t)^2 \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})\right] dt + g(t)d\bar{\boldsymbol{w}}, \tag{5}$$

其中，$\bar{\boldsymbol{w}}$ 是一个沿时间反向运行的维纳过程，而 $\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$ 是时间 t处扰动数据分布的得分函数。

除了随机反向时间随机微分方程外，扩散模型还存在一个由常微分方程（ODE）描述的相关确定性过程，称为概率流 ODE [20]。该 ODE 通过移除反向随机微分方程中的布朗噪声项推导得到，表示为

$$d\boldsymbol{x} = \left[\boldsymbol{f}(\boldsymbol{x}, t) - \frac{1}{2} g(t)^2 \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})\right] dt. \tag{6}$$

值得注意的是，该常微分方程定义了一条确定性轨迹，其边缘概率密度 $p_t(\boldsymbol{x})_{t=0}^T$ 与前向扩散随机微分方程相同。通过自适应常微分方程求解器，这一公式既能实现精确的似然计算，也能更高效地进行采样；相关方法将在推理加速中实现。

B. 离散化与特定实例

通过为 $\boldsymbol{f}(\boldsymbol{x}, t)$ 和 $g(t)$ 选择特定形式，并对随机微分方程进行分步离散化，我们可以得到常见的扩散框架。

令 $\boldsymbol{f}(\boldsymbol{x}, t) = -\frac{1}{2}\beta(t)\boldsymbol{x}$，且 $g(t) = \sqrt{\beta(t)}$，并采用线性噪声调度 $\beta(t)$，即可得到方差保持型随机微分方程。使用 $\beta_t = \beta(t)\Delta t$ 对其进行离散化后，可得到去噪扩散概率模型（DDPM）[19]。前向过程变为 $\boldsymbol{x}_t = \sqrt{1-\beta_t}\boldsymbol{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{\epsilon}$，$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$。

或者，令 $\boldsymbol{f}(\boldsymbol{x}, t) = \mathbf{0}$ 和 $g(t) = \sqrt{\frac{d\sigma^2(t)}{dt}}$，并使 $\sigma(t)$ 递增，即可得到方差爆炸型随机微分方程，

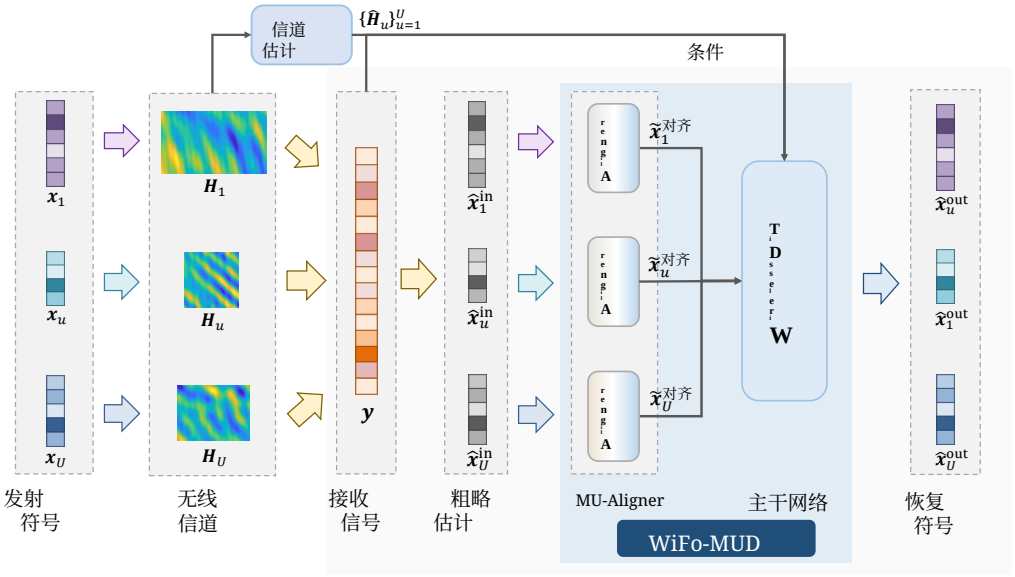


图2：所提出的 WiFo-MUD 在异构多用户系统中的工作流程。

底层的基于得分的模型 [20]。前向过程添加的噪声方差为 $\sigma^2(t)$，其中 $\boldsymbol{x}_t = \boldsymbol{x}_0 + \sigma(t)\boldsymbol{\epsilon}$。

在这两种情况下，模型都学习近似得分函数 $\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$；或者等价地，对于去噪扩散概率模型（DDPM），通过以下目标来预测噪声 $\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t,\ t)$：

$$\mathcal{L} = \mathbb{E}_{t,\boldsymbol{x}_0,\boldsymbol{\epsilon}}\left[\lambda(t)\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\|_2^2\right], \tag{7}$$

其中，$\lambda(t)$ 是一个加权函数。这种随机微分方程视角为理解和开发扩散模型提供了一个统一框架，有助于理论分析和实际算法设计。

IV. 用于多用户信号解调的无线基础模型

为克服现有解调器的瓶颈，我们提出了一种基于扩散模型的无线基础模型，用于信号解调，称为 **WiFo-MUD**。它是一种灵活且可扩展的模型，适用于异构系统配置下的多用户信号解调。WiFo-MUD结合了多用户对齐和条件去噪扩散 Transformer骨干网络，旨在实现对不同系统配置的强泛化能力，同时保持高准确率并解决多用户不平衡问题。

具体而言，首先快速计算 $\{\hat{\boldsymbol{x}}_u^{\text{in}}\}_{u=1}^U$的粗略估计，然后由**多用户对齐器（MU-Aligner）**根据各用户不同的信噪比对这些估计进行预处理。随后，将这些估计输入**无线扩散 Transformer（WirelessDiT）**进行去噪，以将估计细化为$\hat{\boldsymbol{x}}_u^{\text{out}}$。WiFo-MUD的工作流程如图2所示。

A. 多用户对齐器

将不同用户的粗略符号估计直接纳入模型，可能会因噪声水平较高且用户间信噪比严重失衡而导致性能下降。为缓解这一问题，我们首先引入一个轻量级的 **MU-Aligner** 作为预处理阶段。在该模块中，计算每个粗略估计的等效噪声功率，以确定等效去噪时间步。基于此，MU-Aligner 通过基于扩散的去噪机制执行针对特定用户的噪声抑制和信噪比对齐。

1）时间步计算：第 u 个用户的粗略符号估计值，记为 $\hat{\boldsymbol{x}}_u$，可通过低复杂度线性方法获得。例如，LMMSE估计为

$$\hat{\boldsymbol{x}}_u = (\hat{\boldsymbol{H}}^{\text{H}}\hat{\boldsymbol{H}} + \sigma_n^2 \boldsymbol{R}_u^{-1})^{-1}\hat{\boldsymbol{H}}^{\text{H}}\boldsymbol{y}, \tag{8}$$

其中 $\boldsymbol{R}_u = \mathbb{E}[\boldsymbol{x}_u\boldsymbol{x}_u^{\text{H}}] = \boldsymbol{I}_{N_{t_u}}$。将估计误差定义为$\boldsymbol{\xi}_u = \hat{\boldsymbol{x}}_u - \boldsymbol{x}_u$，给定 $\boldsymbol{H}_u$ 时的条件误差协方差为$\mathbb{E}[\boldsymbol{\xi}_u\boldsymbol{\xi}_u^{\text{H}}|\boldsymbol{H}_u] = \boldsymbol{A}_u^{-1}\left[(\sigma_n^2 + \sigma_{H,\ u}^2 N_{t_u})\boldsymbol{H}_u^{\text{H}}\boldsymbol{H}_u + \sigma_n^4\boldsymbol{I}\right]\boldsymbol{A}_u^{-1}$，其中 $\boldsymbol{A}_u = \boldsymbol{H}_u^{\text{H}}\boldsymbol{H}_u + \sigma_n^2\boldsymbol{R}_x^{-1} = \boldsymbol{H}_u^{\text{H}}\boldsymbol{H}_u + \sigma_n^2\boldsymbol{I}_{N_{t_u}}$。

要估计当前信道条件下的误差方差，需要获取信道统计信息。然而，在实际系统中，这类信息往往难以获得。因此，我们利用存储在内存缓冲区中的历史测量CSI，作为完整信道分布的采样。在拥有足够的内存样本时，这些样本集合渐近地逼近长期信道统计信息。基于这些统计信息，我们对 $\boldsymbol{H}_u$ 求边缘化，等效噪声功率可计算为

$$\bar{\sigma}_u^2 = \mathbb{E}_{\boldsymbol{H}_u}\left[\|\boldsymbol{\xi}_u\|_2^2\right] = \mathbb{E}_{\boldsymbol{H}_u}\left[\text{Tr}\left(\mathbb{E}[\boldsymbol{\xi}_u\boldsymbol{\xi}_u^{\text{H}}|\boldsymbol{H}_u]\right)\right] \tag{9}$$
$$=(\sigma_n^2+\sigma_{H,u}^2 N_{t_u})\text{Tr}\left(\mathbb{E}_{\boldsymbol{H}_u}\left[\boldsymbol{H}_u^{\text{H}}\boldsymbol{H}_u(\boldsymbol{A}_u)^{-2}\right]\right)+\sigma_n^4\text{Tr}\left(\mathbb{E}_{\boldsymbol{H}_u}\left[\boldsymbol{A}_u^{-2}\right]\right).$$

在包含 T 个时间步的扩散过程中，每个去噪时间步都对应一个特定的等效噪声功率级别。根据 [31] 中的假设，相邻两个时间步之间的噪声功率差为常数，与估计值 $\hat{\boldsymbol{x}}_u$ 对应的源时间步由以下关系确定：

$$\hat{t}_u = \frac{\bar{\sigma}_u^2(T-1)}{\bar{\alpha}_1 - \bar{\alpha}_T}. \tag{10}$$

2) 用户特定去噪：给定第 u 个用户 $\hat{}\ t_u$ 的源时间步，剩下的任务是将 $\hat{\boldsymbol{x}}_u$ 从 $\hat{t}_u$ 去噪至目标时间步 $\hat{t}^{\text{align}} = \text{argmin}_u \hat{t}_u$。MU-Aligner 的详细架构如图 3 所示。

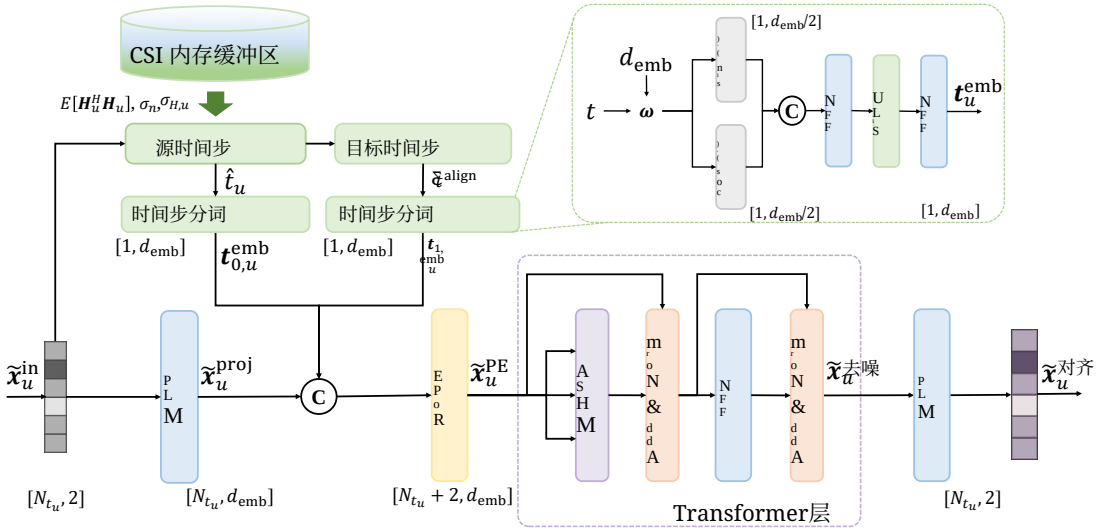


图 3：WiFo-MUD 中 MU-Aligner 模块的网络架构。

输入投影：将

$$\tilde{\boldsymbol{x}}_u^{\text{in}} = \left[\{\boldsymbol{x}_u^{\text{in}}\} \quad \{\boldsymbol{x}_u^{\text{in}}\} \right] \in \mathbb{R}^{N_{t_u} \times 2},$$

Re ˆ, Im ˆ 表示为第 - 个估计值的实值输入。首先将 $\tilde{\boldsymbol{x}}_u^{\text{in}}$ 投影到高维潜在空间，得到 $\tilde{\boldsymbol{x}}_u^{\text{proj}} = \text{FFN}(\tilde{\boldsymbol{x}}_u^{\text{in}}) \in \mathbb{R}^{N_{t_u} \times d_{\text{emb}}}$，从而获得 N_{t_u} 个信号令牌；其中，FFN(·) 表示前馈神经网络（FFN）的一层，d_{emb} 表示潜在空间的维度。

位置嵌入：为捕捉输入中的序列依赖关系，使用旋转位置嵌入（RoPE）对相对位置信息进行编码，得到 $\tilde{\boldsymbol{x}}_u^{\text{PE}} = \text{RoPE}(\tilde{\boldsymbol{x}}_u^{\text{proj}})$。该方法通过在复平面中以旋转方式嵌入位置，适应可变的输入长度；同时在保留词元间方向关系的同时，在注意力分数中引入与距离相关的衰减。

时间步分词：同时，初始步 ˆ t_u和目标时间步 ˆt^{align} 均被投影为时间步词元 $\boldsymbol{t}_{0,u}^{\text{emb}}$，$\boldsymbol{t}_{1,u}^{\text{emb}} \in \mathbb{R}^{1 \times d_{\text{emb}}}$，以控制扩散模型的去噪程度。具体而言，$\boldsymbol{t}_{i,u}^{\text{emb}} = \text{FFN}(\text{SiLU}(\text{FFN}(\psi_i)))$，其中 $i \in \{0,1\}$，SiLU(·) 表示 Sigmoid 线性单元（SiLU）激活函数，而 $\boldsymbol{\psi} = [t\boldsymbol{\omega}, t\boldsymbol{\omega}]$，$\boldsymbol{\omega}[j] = \left(-\frac{j \cdot \log(10000)}{d_{\text{emb}}/2-1}\right) = 10000^{-j/(d_{\text{emb}}/2-1)}$ sin( ) cos( ) with exp)。这种正弦编码策略可在所有时间步上提供平滑且连续的表示。

去噪 Transformer 层：一层 Transformer 编码器作为主干网络，用于捕捉词元之间的关系。添加位置嵌入的信号词元与嵌入的时间步拼接后输入该层，并预测和消除从 ˆt^{align} 到 ˆ t_u 的噪声。

$$\tilde{\boldsymbol{x}}_u^{\text{denoise}} = \tilde{\boldsymbol{x}}_u^{\text{PE}} - \text{Transformer}\left(\left[\tilde{\boldsymbol{x}}_u^{\text{PE, T}}, \boldsymbol{t}_1^{\text{emb, T}}, \boldsymbol{t}_2^{\text{emb, T}}\right]^{\text{T}}\right). \quad (11)$$

信号解码：潜在空间中的去噪信号随后被投影回原始平面，得到 $\tilde{\boldsymbol{x}}_u^{\text{align}} = \text{FFN}(\tilde{\boldsymbol{x}}_u^{\text{denoise}}) \in \mathbb{R}^{N_{t_u} \times 2}$。

MU-Aligner 模块通过测量每个用户的等效噪声功率，并据此执行时间步感知去噪，将所有用户的信噪比调整至相近水平。此过程缓解了多用户不平衡问题，从而为后续的估计优化奠定了稳健的基础。

B. 无线扩散 *Transformer*

WiFo-MUD 的核心是**无线扩散 Transformer（WirelessDiT）**，这是一种基于扩散 Transformer（DiT）的条件去噪架构，专为异构多用户系统中的无线信号解调而设计。它由四个组件构成：信号嵌入、条件融合、去噪块和信号解码。对齐后的多用户估计 $\{\tilde{\boldsymbol{x}}_u^{\text{align}}\}_{u=1}^{U}$ 被输入 WirelessDiT，其输出为符号 $\{\hat{\boldsymbol{x}}_u^{\text{out}}\}_{u=1}^{U}$的细化结果。WirelessDiT 的整体架构如图 4 所示。

1) 信号嵌入：该模块通过多层感知机（MLP）将对齐后的估计值 $\tilde{\boldsymbol{x}}_u^{\text{align}}$编码为高维潜在表示，即 $\boldsymbol{z}_u = \text{MLP}(\tilde{\boldsymbol{x}}_u^{\text{align}}) \in \mathbb{R}^{N_{t_u} \times d_{\text{emb}}}$。然后，将所有用户的潜在表示拼接为 $\boldsymbol{Z} = \left[\boldsymbol{z}_1^{\text{T}}, \cdots, \boldsymbol{z}_U^{\text{T}}\right]^{\text{T}} \in \mathbb{R}^{L \times d_{\text{emb}}}$，其中包含 $L = \sum_{u=1}^{U} N_{tu}$，随后通过 RoPE 嵌入相对位置信息，即 $\boldsymbol{Z}^{\text{PE}} = \text{RoPE}(\boldsymbol{Z})$。

2) 条件融合模块：在包括接收信号 $\boldsymbol{y}$ 和信道 $\boldsymbol{H}_u$在内的条件引导下，对潜在表示进行去噪。在条件融合模块中，首先分别对 $\{\hat{\boldsymbol{H}}_u\}_{u=1}^{U}$ 和 $\boldsymbol{y}$ 进行编码，然后将其融合以构建条件向量。

CSI嵌入模块：我们首先构建一个 CSI 嵌入（CE）模块 $\mathcal{E}_H$，用于预处理信道信息，从而生成信道条件向量

$$\boldsymbol{c}_{H,u} = \mathcal{E}_H\left(\hat{\boldsymbol{H}}_u, t, u, [N_r, N_{t_u}]\right), \quad (12)$$

其中，时间步 t、用户身份 u和信道形状$[N_r, N_{t_u}]$ 被嵌入信道相关条件中，作为辅助信息，以帮助模型适应不同的配置。具体而言，$\boldsymbol{H}_u$ 首先被划分为大小为 $p_x \times p_y$的非重叠块。这些块经过嵌入后形成 $L_{H,u} = \frac{N_r N_{t_u}}{p_x p_y}$词元，随后应用 RoPE，得到 $\boldsymbol{H}_u^{\text{emb}} = \text{RoPE}\left[\text{多层感知机}\left(\text{Conv2D}(\hat{\boldsymbol{H}}_u)\right)\right] \in \mathbb{R}^{L_{H,u} \times d_{\text{emb}}}$。与此同时，$u$、$[N_r, N_{t_u}]$和 t 的可学习嵌入被生成为

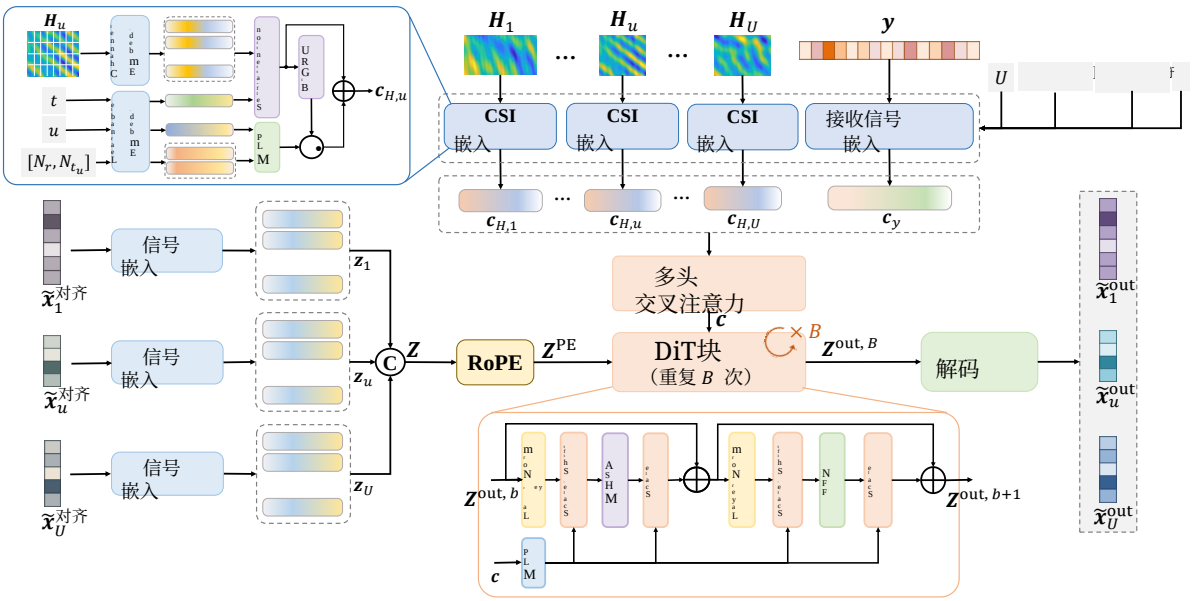


图 4：WiFo-MUD 中 WirelessDiT 的网络架构。

$\boldsymbol{u}^{\text{emb}} \in \mathbb{R}^{d_{\text{emb}}}$、$\boldsymbol{s}^{\text{emb}} \in \mathbb{R}^{2d_{\text{emb}}}$ 和$\boldsymbol{t}^{\text{emb}} \in \mathbb{R}^{d_{\text{emb}}}$。这些辅助嵌入通过门控融合机制与信道令牌相结合。第 u 个用户的完整 CSI 嵌入表示为

$$\begin{aligned}\boldsymbol{c}_{H,u} = &\text{Average}\left(\text{Self-Attention}\left([\boldsymbol{H}_u^{\text{emb,T}}, \boldsymbol{t}^{\text{emb,T}}]^{\text{T}}\right)\right) \\ &+\text{MLP}\left([\boldsymbol{u}^{\text{emb}}, \boldsymbol{s}^{\text{emb}}]\right) \odot \text{Average}\left(\text{BiGRU}(\boldsymbol{H}_u^{\text{emb}})\right),\end{aligned} \tag{13}$$

其中，Average(·)执行逐词元平均，BiGRU(·)捕获词元 [32] 之间的序列依赖关系。

接收信号嵌入模块：对接收 $\boldsymbol{y}$ 进行预处理，以获得 $\boldsymbol{c}_y = \mathcal{E}_y(\boldsymbol{y},\ t,\ U,\ N_r)$。在此模块中，系统配置参数（包括 U、N_r 和 t）通过专用的离散嵌入模块注入来自 $\boldsymbol{y}$ 的信号词元中。这些模块的运作方式与前文所述的 CSI 嵌入类似，因此为简洁起见，省略其详细设计。

交叉注意力模块：为整合与信道相关和与信号相关的条件，我们通过多头交叉注意力模块融合其特征表示，具体如下：

$$\boldsymbol{c} = \text{CrossAttention}(\{\boldsymbol{c}_{H,u}\}_{u=1}^{U}, \boldsymbol{c}_y) \in \mathbb{R}^{1\times d_{\text{emb}}}. \tag{14}$$

该模块不仅对两种条件类型之间的依赖关系进行建模，还捕获多用户信道之间的相关性，从而有助于减轻用户间干扰。随后，利用所得的统一条件向量 $\boldsymbol{c}$引导估计细化过程。

3) 去噪块：该组件在潜在空间中执行反向扩散过程，根据粗略估计重建干净信号。在每个扩散步 t，输出都会通过一系列 WirelessDiT 模块进行更新。

首先，条件向量 $\boldsymbol{c}$ 被投影，以生成每个 WirelessDiT 模块中的调制参数：

$$[\boldsymbol{\gamma}_1^b, \boldsymbol{\beta}_1^b, \boldsymbol{\gamma}_2^b, \boldsymbol{\beta}_2^b, \alpha^b] = \text{MLP}(\boldsymbol{c}), \tag{15}$$

其中，$\boldsymbol{\gamma}_1^b$、$\boldsymbol{\beta}_1^b$、$\boldsymbol{\gamma}_2^b$、$\boldsymbol{\beta}_2^b \in \mathbb{R}^{d_{\text{emb}}}$ 是第 b 个 WirelessDiT 块中的缩放和偏移参数，而 $\alpha^b \in \mathbb{R}$ 是门控参数，其操作方式与 AdaLN-Zero [33]相同。然后，输入使用由条件推导出的参数进行自适应层归一化

$$\boldsymbol{Z}^{in,b} = \boldsymbol{\gamma}_1^b \odot \frac{\text{LN}(\boldsymbol{Z}^{\text{PE}}) - \mu(\text{LN}(\boldsymbol{Z}^{\text{PE}}))}{\sigma(\text{LN}(\boldsymbol{Z}^{\text{PE}})) + \epsilon} + \boldsymbol{\beta}_1^b, \tag{16}$$

其中，LN(·) 执行层归一化，$\mu(\cdot)$ 和 $\sigma(\cdot)$ 沿特征维度计算均值和标准差，ϵ 是一个用于保证数值稳定性的小常数。接下来是多头自注意力，其表达式为

$$\boldsymbol{Z}^{\text{attn},b} = \text{Softmax}\left(\frac{\boldsymbol{Z}^{\text{in},b}\boldsymbol{W}_Q(\boldsymbol{Z}^{\text{in},b}\boldsymbol{W}_K)^{\text{T}}}{\sqrt{d_{qk}}}\right)\boldsymbol{Z}^{\text{in},b}\boldsymbol{W}_v, \tag{17}$$

其中，$\mathbf{W}_Q^b$、$\mathbf{W}_K^b$、$\mathbf{W}_V^b \in \mathbb{R}^{d_{\text{emb}}\times d_{qk}}$ 是第 b 个 WirelessDiT 模块中的查询、键和值投影矩阵，d_{qk} 是每个注意力头的维度。注意力输出通过条件生成的参数进行门控，表示为 $\boldsymbol{Z}^{\text{gate},\,b} = \alpha^b \boldsymbol{Z}^{\text{attn},\,b}$，其中 α^b 融合了注意力机制。随后，原始输入通过残差连接与处理后的特征结合，表示为 $\boldsymbol{Z}^{\text{res},\,b} = \boldsymbol{Z}^{\text{gate},\,b} + \boldsymbol{Z}^{\text{in},\,b}$。之后，这些特征经过第二次自适应归一化和前馈处理，表示为 $\boldsymbol{Z}^{\text{ffn},\,b}$。第 b 个 WirelessDiT 模块的最终输出计算为 $\boldsymbol{Z}^{\text{out},\,b} = \boldsymbol{Z}^{\text{res},\,b} + \boldsymbol{Z}^{\text{ffn},\,b}$。这些模块体现了随机微分方程所规定的逐步误差校正原理。随着过程沿轨迹演化，粗略估计会不断得到细化，从而降低整体估计误差。

4) 信号解码：最后，将优化后的潜在信号投影回符号空间，得到最终估计值 $\tilde{\boldsymbol{x}}^{\text{out}} = \text{多层感知机}(\boldsymbol{Z}^{out,\ B}) \in \mathbb{R}^{L\times 2}$，并恢复复值符号 $\hat{\boldsymbol{x}}^{out} = \tilde{\boldsymbol{x}}^{\text{out}}[:,1] + j\tilde{\boldsymbol{x}}^{\text{out}}[:,2]$。基于这些优化后的估计值，通过符号量化恢复发送比特。

C. 模型训练

在 WiFo-MUD 中，MU-Aligner 和 WirelessDiT 按顺序进行预训练。首先训练 MU-Aligner，以对齐不同用户的估计值。WirelessDiT 则使用 MU-Aligner 的输出来作为部分输入进行预训练，从而

进一步对估计值进行去噪。对于这两个模块，我们采用自监督的扩散重建预训练策略，以学习能够抵御各种信道损伤和干扰模式的稳健信号表示。训练过程遵循 DDPM 框架，包括前向扩散过程和反向去噪过程。

令 $\boldsymbol{x}^{(0)}$ 为输入符号向量，为简洁起见，令条件为 $\boldsymbol{c}$ 。前向扩散过程在一系列 T 时间步中逐渐向 $\boldsymbol{x}^{(0)}$ 添加高斯噪声，生成一系列带噪潜变量$\boldsymbol{x}^{(1)}$，$\boldsymbol{x}^{(2)}$，…，$\boldsymbol{x}^{(T)}$。从 $\boldsymbol{x}^{(t-1)}$到 $\boldsymbol{x}^{(t)}$的转移定义为

$$q(\boldsymbol{x}^{(t)}|\boldsymbol{x}^{(t-1)},\boldsymbol{c})=\mathcal{N}(\boldsymbol{x}^{(t)};\sqrt{1-\beta_t}\boldsymbol{x}^{(t-1)},\beta_t\mathbf{I}), \quad (18)$$

其中，$\{\beta_t\in(0,1)\}_{t=1}^{T}$是一个固定的方差调度，用于控制每一步添加的噪声量。该马尔可夫链的一个显著性质是，给定 $\boldsymbol{x}^{(0)}$，任意中间状态 $\boldsymbol{x}^{(t)}$ 都可以用闭式形式表示为：

$$\boldsymbol{x}^{(t)}=\sqrt{\bar{\alpha}_t}\boldsymbol{x}^{(0)}+\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, \quad (19)$$

其中，$\alpha_t=1-\beta_t$、$\bar{\alpha}_t=\prod_{s=1}^{t}\alpha_s$和 $\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$为标准高斯噪声。

为简洁起见，模型以 θ 为参数化形式，通过从带噪输入 $\boldsymbol{x}^{(t)}$预测噪声向量 $\boldsymbol{\epsilon}$来学习逆转该扩散过程。训练目标是最小化预测噪声与真实噪声之间的均方误差，即优化

$$\mathcal{L}(\theta)=\mathbb{E}_{t,\boldsymbol{x}^{(0)},\boldsymbol{\epsilon}}\left[\|\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_\theta(\mathbf{x}^{(t)}|t,\boldsymbol{c})\|_2^2\right], \quad (20)$$

其中，t从 $\{1,2,\ldots,T\}$中均匀采样，$\boldsymbol{x}^{(0)}$从训练数据集中抽取，且 $\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$。$\boldsymbol{c}$包含接收信号 $\boldsymbol{y}$和估计信道 $\{\hat{\boldsymbol{H}}_u\}_{u=1}^{U}$。训练算法总结于算法 1。

算法 1 扩散重建训练算法 输入：训练数据集 $\mathcal{D}$，其中包含符号向量 $\boldsymbol{x}^{(0)}$及对应条件 $\boldsymbol{c}$；采样时间步总数 T；训练轮数 E_{diff}输出：已学习的去噪模型 $\boldsymbol{\epsilon}_\theta$；步骤：对于 $n=1$ 到 E_{diff}：00 采样 $t\sim\text{Uniform}(1,\ldots,T)$；00 采样噪声 $\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})$；00 计算 $\boldsymbol{x}^{(t)}=\sqrt{\bar{\alpha}_t}\boldsymbol{x}^{(0)}+\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$；00 对 $\nabla_\theta|\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_\theta(\mathbf{x}^{(t)}|t,\boldsymbol{c})|^2$执行梯度步；00 根据梯度更新 $\boldsymbol{\epsilon}_\theta$。

请注意，MU-Aligner 和 WirelessDiT 的训练方式相似，但输入条件不同。MU-Aligner 以初始时间步和目标时间步为条件，而 WirelessDiT 则以各种配置信息和单个时间步为条件。在 WirelessDiT 的训练过程中，预训练MU-Aligner模块保持冻结。

V. 推理优化

训练完成后，WiFo-MUD 会根据初始的粗略输入生成更精细的比特估计。然而，其实际部署面临两项主要挑战：推理延迟与物理层帧不兼容，以及准确率因用户间干扰而下降。为应对这些问题，我们从两方面优化推理过程。首先，通信感知的一致性蒸馏技术加速了去噪过程。其次，动态用户分组方案采用串行干扰消除来减轻干扰。

A. 通信感知一致性蒸馏

为提升 WiFo-MUD 的推理速度，我们提出一种**通信感知一致性蒸馏**策略。该方法将知识从预训练的 WirelessDiT 模型迁移至效率更高的学生网络，旨在通过单步前向推理近似多步采样，同时保持解调精度。

我们将预训练教师 WirelessDiT,$\mathcal{T}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{x}}^{\text{align},(t)};t,\boldsymbol{c})$ 蒸馏到学生模型 $\mathcal{S}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{x}}^{\text{align},(t)};t,\boldsymbol{c})$ 中，该学生模型旨在将任意噪声状态$\{\tilde{\boldsymbol{x}}^{\text{align},(t)}\}_{u=1}^{U}$ 在时间步 t 直接映射为对应的干净符号 $\tilde{\boldsymbol{x}}^{(0)}$。我们使用多目标损失函数对蒸馏优化进行如下公式化：

$$\mathcal{L}(\boldsymbol{\theta})=\eta_d\mathcal{L}_{\text{CD}}(\boldsymbol{\theta})+(1-\eta_d)\mathcal{L}_{\text{MSE}}(\boldsymbol{\theta}), \quad (21)$$

结合一致性蒸馏损失与估计均方误差损失，并根据当前信道条件自适应地调整权重。

1）一致性蒸馏损失：与 [34], 类似，一致性蒸馏损失定义为

$$\mathcal{L}_{\text{CD}}(\boldsymbol{\theta})=\mathbb{E}_t[\lambda(t_n)\|\mathcal{S}_\theta(\tilde{\boldsymbol{x}}^{\text{align},(t_{n+1})};t_{n+1},\boldsymbol{c}) - \mathcal{S}_{\theta^-}(\tilde{\boldsymbol{x}}^{\text{align},(t_n)};t_n,\boldsymbol{c})\|_{\text{F}}], \quad (22)$$

其中，$\lambda(t)$ 是依赖于时间步的权重，用于优先处理不同的噪声水平。$\tilde{\boldsymbol{x}}^{\text{align},(t_n)}$ 是使用教师模型执行一步去噪后得到的结果。$\boldsymbol{\theta}^-$ 是学生模型参数$\boldsymbol{\theta}$ 的指数移动平均（EMA），用于稳定训练。该损失项促使处于同一概率轨迹上的样本产生一致的模型输出。

2）通信感知损失设计：为进一步提高蒸馏过程中的数据效率并改善解调稳定性，我们通过改进的损失函数融入领域知识。将学生模型估计的符号与真实值进行比较，以计算均方误差损失：

$$\mathcal{L}_{\text{MSE}}(\boldsymbol{\theta})=\mathbb{E}\|\mathcal{S}_\theta(\tilde{\boldsymbol{x}}^{\text{align},(t_{n+1})};t_{n+1},\boldsymbol{c})-\tilde{\boldsymbol{x}}^{(0)}\|_2^2. \quad (23)$$

为实现不同损失分量之间的最佳平衡，我们根据信道条件设计了一个自适应权重：

$$\eta_d=\frac{1}{1+\exp\{-k_0(\frac{\sum_{u=1}^{U}\|\hat{\boldsymbol{H}}_u\|_{\text{F}}^2}{U\sigma_n^2}-s_0)\}}, \quad (24)$$

其中 k_0 和 s_0 为正数常量。当信道条件恶化时，会更加注重最小化均方误差损失；而在较高信噪比场景下，则优先优化一致性蒸馏损失。

学生模型 $\mathcal{S}_\theta$ 与教师模型采用相同的架构，但参数彼此独立。蒸馏后的 WirelessDiT$\mathcal{S}_\theta$ 在准确率上可媲美迭代教师模型，同时将推理时间从 N 步缩短至一步，因此适用于时延预算较低的实时系统。算法2概述了完整的通信感知一致性蒸馏流程。

算法2 通信感知一致性蒸馏**输入**：数据集 $\mathcal{D}$；预训练教师 WirelessDiT 模型 $\mathcal{T}_{\boldsymbol{\phi}}$；$\{\lambda(t_n)\}_{n=1}^{T}$；指数移动平均更新权重 μ；蒸馏轮次 E_{dis}**输出**：蒸馏后的 WirelessDiT 模型 $\mathcal{S}_{\boldsymbol{\theta}}$**步骤：** 初始化学生模型参数 $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}_0$；初始化指数移动平均参数 $\boldsymbol{\theta}^{-} \leftarrow \boldsymbol{\theta}$；

重述

Sample $\{\tilde{\boldsymbol{x}}^{\text{align},(0)}, \mathbf{c}\}$ from $\mathcal{D}$;
Sample $n \sim \mathcal{U}[1, T]$;
Generate $\tilde{\boldsymbol{x}}^{\text{align},(t_{n+1})} \sim \mathcal{N}(\tilde{\boldsymbol{x}}^{\text{align},(0)}; t_{n+1}^2 \boldsymbol{I})$;
Compute teacher outputs: $\tilde{\boldsymbol{x}}^{\text{align},(t_n)} \leftarrow \tilde{\boldsymbol{x}}^{\text{align},(t_{n+1})} - (t_n - t_{n+1}) t_{n+1} \mathcal{T}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{x}}^{\text{align},(t_{n+1})}, t_{n+1}, \boldsymbol{c})$;
Compute the combined loss $\mathcal{L}(\boldsymbol{\theta})$ as Eq. (40);
Update $\boldsymbol{\theta}$ using Adam optimizer: $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta \nabla_{\theta} \mathcal{L}_{\text{CD}}(\boldsymbol{\theta})$;
Update EMA: $\boldsymbol{\theta}^{-} \leftarrow \text{stopgrad}(\mu \boldsymbol{\theta}^{-} + (1-\mu)\boldsymbol{\theta})$;
until convergence **or** E_{dis} is reached

B. 动态用户分组与连续推理策略

为有效缓解用户间干扰，我们引入了**动态用户分组**和连续推理消除策略。

首先，根据所有用户的信道强度对其进行排序。信道强度通过 $h_u = \frac{\|\boldsymbol{H}_u\|_{\text{F}}^2}{N_r N_{t_u}}$ 测量。用户按降序排列，形成如下有序序列：

$$\Pi = \text{argsort}\left(\{h_1, h_2, \cdots, h_U\}\right). \tag{25}$$

用户被划分为若干组，并优先处理信道条件较好的用户，以提高解调可靠性，并促进后续阶段更准确地进行干扰消除。

为避免低秩信道条件，数据流总数不应超过阈值 N_{thres}，其中$N_{\text{thres}} \leqslant N_r$。因此，采用贪心方式递归划分并选择各组。用户索引的第 i 组定义为 $\mathcal{G}^{(i)}$，该组的用户数可表示为

$$|\mathcal{G}^{(i)}| = \underset{\sum_{l=1+\sum_{j=1}^{i-1}|\mathcal{G}^{(j)}|}^{k+\sum_{j=1}^{i-1}|\mathcal{G}^{(j)}|} N_{t_{\Pi[l]}} \leqslant N_{\text{thres}}}{\text{argmax}} k. \tag{26}$$

换言之，从剩余集合 Π 中按降序将用户选入第 i 组，同时确保数据流总数不超过阈值。组 i 对应的信道子集和传输符号矩阵 $\hat{}$ 分别为 $\{\boldsymbol{H}_j\}_{j\in\mathcal{G}^{(i)}}$ 和 $\{\hat{}\boldsymbol{x}_j\}_{j\in\mathcal{G}^{(i)}}$。

各组之间依次进行解调。对于第 i 组，接收信号计算如下：

$$\boldsymbol{y}^{(i)} = \boldsymbol{y}^{(0)} - \sum_{j=1}^{i-1} \sum_{u\in\mathcal{G}^{(j)}} \hat{\boldsymbol{H}}_u \hat{\boldsymbol{x}}_u^{\text{out}}, \tag{27}$$

其中，$\boldsymbol{y}^{(0)} = \boldsymbol{y}$ 和$\hat{}\boldsymbol{x}_u^{\text{out}}$是 WiFo-MUD 对前面 $i-1$组的精化输出。假设后续各组的贡献构成干扰，WiFo-MUD 通过单步推理估计第 i组的符号，结果如下：

$$\{\hat{\boldsymbol{x}}_u^{\text{out}}\}_{u\in\mathcal{G}^{(i)}} = \text{WiFo-MUD}(\hat{\boldsymbol{x}}_u, \boldsymbol{c}), \tag{28}$$

其中，$\boldsymbol{c}$ 包括 $\{\hat{}\boldsymbol{H}_j\}_{j\in\mathcal{G}^{(i)}}$、$\boldsymbol{y}^{(i)}$ 及其他配置信息。估计第 i组的符号和信道后，也会从当前接收信号中移除该组的贡献（类似于式 (27)），以减少对后续各组的干扰。此循环持续进行，直至所有用户均完成分组和处理。

VI. 实验

本节对所提出的 WiFo-MUD 框架进行全面评估。我们首先介绍用于模型训练和测试的数据集、用于比较的基线方法，并定义评估指标。随后，我们从多个角度评估所提出的 WiFo-MUD 方法的性能。

A. 数据集构建

为全面评估所提出的模型，我们构建了一个大规模异构多用户信号解调数据集，融合了仿真数据和实际测量数据。该数据集涵盖广泛的传输场景、系统配置和调制方案，以确保多样性和保真度。信道数据按照 [27] 中概述的方法生成，包含以下组成部分：

- **统计信道建模**：该子集使用*QuaDriGa* [35] 按照 3GPP 38.901 标准生成，涵盖城市宏蜂窝（UMa）、城市微蜂窝（UMi）、农村宏蜂窝（RMa）以及室内场景，并包含视距和非视距条件。
- **射线追踪**：从*DeepMIMO* 数据集 [36] 和通过 Wireless InSight 生成的*SynthSoM*数据集 [37], 中引入高保真信道数据，用于对多种场景进行详细的射线追踪仿真。
- **实际测量**：实际测得的信道数据取自*Argos* 数据集 [38],，该数据集采集于一所大学校园的室内和室外环境；还取自*Dichasus* 数据集 [39] ，涵盖教室、工厂和室外区域。

我们的数据集包含均衡分布的调制方案（四相相移键控、八相相移键控、16 阶正交幅度调制、64 阶正交幅度调制），每种场景的样本数量相等。每个实例包括发射符号 $\{\boldsymbol{x}_u\}_{u=1}^{U}$、接收信号 $\boldsymbol{y}$、信道矩阵 $\{\boldsymbol{H}_u\}_{u=1}^{U}$和噪声方差 σ_n^2。我们数据集的命名约定和信道相关参数配置也与 [27] 保持一致。

表 I：模型训练超参数 g

MU-Aligner		WirelessDiT		模型蒸馏	
训练轮数	100	训练轮数	100	训练轮数	80
批次大小	64	批次大小	64	批次大小	64
L_r	0.002	L_r	0.001	L_r	0.0008
优化器	AdamW	优化器	AdamW	优化器	Adam
权重衰减	0.0001	时间步	1000	权重 μ	0.2

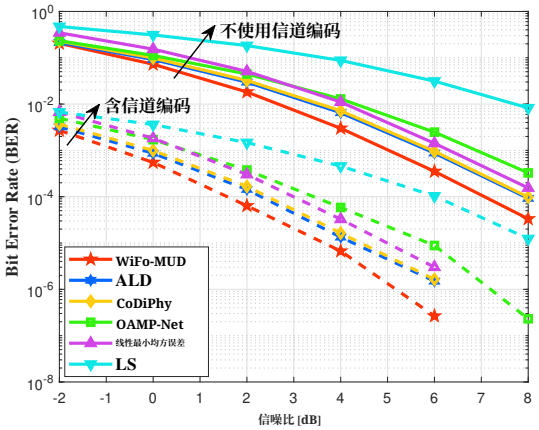


图 5：不同解调器在不同信噪比下的误码率比较。

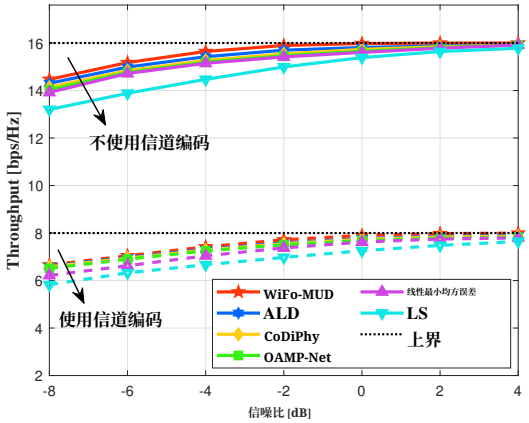


图 6：不同解调器在不同信噪比下的吞吐量比较。

B. 比较方法

在我们的实验中，我们开发了所提出的 WiFo-MUD 模型的三种架构变体，分别称为 **WiFo-MUD-Large**、**WiFo-MUD-Base** 和 **WiFo-MUD-Small**，它们在网络层数量和潜在空间维度方面有所不同。WiFo-MUD 模型采用的训练超参数列于表 I 中，而详细的架构规格总结于表II 中。我们将 WiFo-MUD 与以下代表性基线方法进行比较：

- 线性解调器：最小二乘和线性最小均方误差估计器。
- *OAMP-Net* [18]：一种深度学习解调器，将正交近似消息传递算法展开为可训练网络，并通过端到端训练优化可学习的阻尼参数和散度参数。
- *ALD* [21]：一种采用退火朗之万动力学的基于扩散的解调器，通过使用预训练的信道特定得分网络，在后验得分中进行退火采样来估计符号。
- *CoDiPhy* [24]：一种基于条件 DDPM 的信号解调框架。由于对异构场景的适应性有限，该框架需要针对每种系统配置分别进行独立训练。

所有基于学习的基准模型均采用其官方推荐的配置进行训练。

C. 性能指标

通信可靠性：为了评估不同解调器的准确率，我们将每个样本的误比特率（BER）作为主要指标，其定义为

$$\mathrm{BER} = \frac{\sum_{u=1}^{U}\sum_{n=1}^{N_{t_u}} |\hat{\boldsymbol{b}}(u,n) - \boldsymbol{b}(u,n)|}{|\mathcal{X}| U \sum_{u=1}^{U} N_{t_u}}, \tag{29}$$

其中，$\boldsymbol{b}(u,n)$ 是第 n 个符号中来自第 u 个用户的发送比特序列，$\hat{\boldsymbol{b}}(u,n)$ 是解调器对其的估计。

通信效率：我们以吞吐量作为评估这些解调器所实现传输数据速率的指标。吞吐量定义为单位时间内成功传输的比特数，可计算为

$$\mathrm{Throughput} = \frac{\sum_{u=1}^{U}\sum_{n=1}^{N_{t_u}} \left(|\mathcal{X}| - |\hat{\boldsymbol{b}}(u,n) - \boldsymbol{b}(u,n)|\right)}{T_{\mathrm{total}}}, \tag{30}$$

其中，L 是每个符号的比特数，T_{total} 是传输所考虑符号的总时长。

计算复杂度：除了通信可靠性和效率之外，我们还通过浮点运算次数（FLOPs）和推理时延（毫秒/样本）评估计算复杂度。

D. 比较研究

我们首先评估 WiFo-MUD-Base 在信噪比为 0 dB 时的全样本解调精度。WiFo-MUD 在所有训练数据集上完成预训练后，会在具有相同分布和配置的验证集上进行测试。相比之下，我们针对维度相同的信道数据分别训练并评估了 ALD 模型，并针对每个不同的配置数据集独立训练了 CoDiPhy 和 OAMP-Net。为了评估其在实际系统中的性能，我们还模拟了一个包含信道编码和解码的传输系统，具体采用码率为 0.5 的 LDPC码。表 III 总结了各方法在有无信道编码两种情况下的全样本准确率。结果表明

表II：不同规模 WiFo-MUD 的网络架构参数量。

模型	对齐器深度	对齐器宽度	对齐器头数	WirelessDiT深度	WirelessDiT宽度	WirelessDiT头数	参数量
WiFo-MUD-Small	1	64	4	3	64	8	1.29M
WiFo-MUD-Base	1	64	4	6	128	8	3.92M
WiFo-MUD-Large	2	64	4	8	256	8	5.83M

表 III：**全样本** 不同解调器之间的误码率比较。 浅红色阴影表示最高准 蓝色表示次优性能。

数据集	Ours		ALD		CoDiPhy		OAMP-Net		线性最小均方误差		LS	
	未编码	已编码	未编码	已编码	未编码	已编码	未编码	已编码	未编码	已编码	未编码	已编码
Q1	**0.0782**	**0.0005**	0.0803	0.0006	0.0835	0.0007	0.0832	0.0007	0.1212	0.0013	0.1366	0.0014
Q2	**0.0809**	**0.0005**	0.0832	0.0006	0.0856	0.0008	0.0872	0.0009	0.1202	0.0012	0.1347	0.0013
Q3	**0.0898**	**0.0006**	0.0913	0.0007	0.0941	0.0009	0.0940	0.0009	0.1230	0.0013	0.1387	0.0015
Q4	**0.0781**	**0.0004**	0.0798	0.0006	0.0803	0.0007	0.0832	0.0008	0.1092	0.0011	0.1181	0.0013
Q5	**0.0790**	**0.0005**	0.0812	0.0007	0.0819	0.0007	0.0831	0.0008	0.1216	0.0012	0.1306	0.0013
Q6	**0.0782**	**0.0005**	0.0801	**0.0005**	0.0818	0.0007	0.0822	0.0008	0.1210	0.0012	0.1287	0.0014
Q7	**0.0773**	**0.0004**	0.0811	0.0006	0.0835	0.0008	0.0840	0.0009	0.1241	0.0013	0.1311	0.0014
Q8	**0.0819**	**0.0005**	0.0860	0.0007	0.0854	0.0006	0.0892	0.0007	0.1160	0.0011	0.1252	0.0014
D1	**0.0881**	**0.0006**	0.0898	0.0008	0.0908	0.0008	0.0952	0.0009	0.1186	0.0013	0.1365	0.0015
D2	**0.0870**	**0.0006**	0.0901	0.0008	0.0896	0.0007	0.0975	0.0009	0.1196	0.0012	0.1329	0.0014
Ave.	**0.0819**	**0.0005**	0.0843	0.0007	0.0857	0.0007	0.0879	0.0008	0.1195	0.0012	0.1313	0.0014

表 IV：**零样本** 不同解调器之间的误码率比较。 浅红色阴影表示最高准 蓝色突出显示次优性能。

数据集	Ours		ALD		CoDiPhy		OAMP-Net		线性最小均方误差		LS	
	未编码	已编码	未编码	已编码	未编码	已编码	未编码	已编码	未编码	已编码	未编码	已编码
Q9	**0.0891**	**0.0007**	0.0907	0.0008	0.0923	0.0008	0.1542	0.0016	0.1218	0.0013	0.1320	0.0015
Q10	**0.0896**	**0.0007**	0.0899	0.0008	0.0908	0.0008	0.1239	0.0014	0.1227	0.0014	0.1357	0.0015
Q11	**0.0903**	**0.0008**	0.0903	0.0009	0.0907	0.0009	0.1315	0.0016	0.1272	0.0015	0.1408	0.0016
Q12	**0.0910**	**0.0008**	0.0914	0.0009	0.0955	0.0010	0.1402	0.0015	0.1197	0.0013	0.1337	0.0015
Q13	**0.0862**	**0.0008**	0.0877	0.0010	0.0895	0.0010	0.1104	0.0012	0.1178	0.0013	0.1289	0.0015
Q14	**0.0859**	**0.0007**	0.0861	**0.0007**	0.0865	0.0009	0.0998	0.0011	0.1089	0.0012	0.1295	0.0014
Q15	0.0924	**0.0009**	**0.0921**	**0.0009**	0.0984	0.0010	0.1087	0.0011	0.1106	0.0013	0.1298	0.0014
Q16	**0.0906**	**0.0008**	0.0909	0.0009	0.0918	0.0010	0.1154	0.0011	0.1160	0.0012	0.1307	0.0015
D3	**0.0934**	**0.0009**	0.0946	0.0011	0.0968	0.0012	0.1187	0.0014	0.1204	0.0014	0.1358	0.0016
D4	**0.0894**	**0.0008**	0.0895	0.0009	0.0928	0.0010	0.0974	0.0011	0.1185	0.0013	0.1289	0.0015
S1	**0.0948**	**0.0008**	0.0963	0.0010	0.1005	0.0010	0.1201	0.0014	0.1196	0.0013	0.1335	0.0016
Ave.	**0.0902**	**0.0008**	0.0909	0.0009	0.0932	0.0010	0.1200	0.0013	0.1184	0.0013	0.1327	0.0015

所提出的 WiFo-MUD 在两种情况下均在所有数据集上实现了最低的误码率，展现出其优越性。基于扩散的 ALD 方法在大多数数据集上排名第二，显著优于线性解调器和 OAMP-Net，证实了迭代采样在复杂信道分布下的优势。

我们进一步比较了各方法的零样本性能，即预训练模型在分布外（OOD）数据上的性能，如表 IV 所示。测试涵盖了多种场景，包括仿真和实际测量。结果表明，所有基于学习的模型在 OOD 数据上的准确率均有所下降，其中 OAMP-Net 的降幅最大，凸显了分布偏移对深度学习模型泛化能力带来的挑战。相比之下，WiFo-MUD 在大多数 OOD 数据集上实现了最先进的准确率。这种强大的泛化能力归因于其在预训练期间对噪声分布的精确建模，以及在模型推理过程中对统计特性的利用。值得注意的是，在 OOD 场景下，WiFo-MUD 与 ALD 之间的性能差距有所缩小。这一现象可以通过 ALD 类方法的设计来解释：这类方法只需泛化到信道维度，同时保持其他计算与模型无关。这一特性使 ALD 在应对某些类型的分布偏移时具有更强的适应性。

在图 5 中，我们比较了系统配置为 $N_t = 2, U = 4, N_r = 16$ 且采用 QPSK 调制时，不同解调器的平均误比特率（BER）随信噪比变化的情况。在低信噪比区域，基于学习的解调器通常比 LMMSE 解调器实现更低的误码率。然而，其中一些解调器（如 OAMP-Net）在高信噪比下的表现不及 LMMSE。相比之下，WiFo-MUD 在所有信噪比水平下始终实现最低的误码率，比 ALD 提高近 1 dB。引入信道编码后，这一准确率优势依然保持，展现了 WiFo-MUD 在通信中的竞争力与可靠性。

图 6进一步展示了采用和未采用LDPC编码时不同解调方案的吞吐量性能。所提出的WiFo-MUD表现出显著的

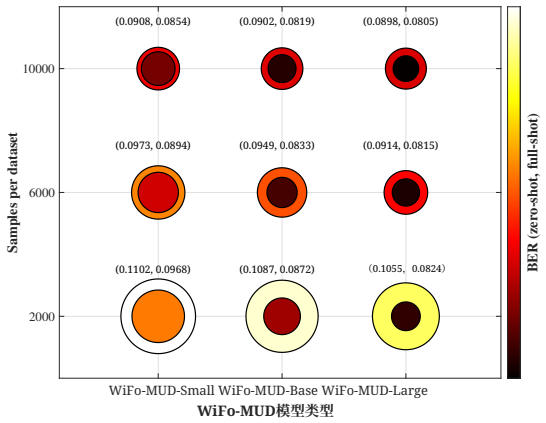


图7：不同 WiFo-MUD 模型规模和数据集规模下的全样本与零样本误码率。

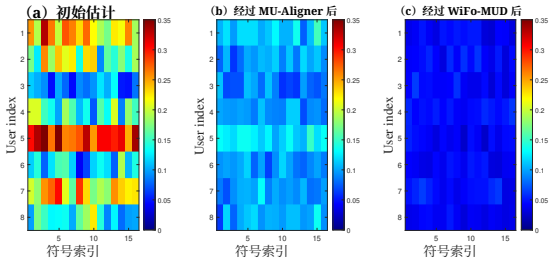


图8：WiFo-MUD 不同处理阶段的估计误差水平可视化。

吞吐量增益。在未采用编码时，传统线性解调器由于干扰抑制能力较弱，吞吐量增长有限；而WiFo-MUD通过利用多用户对齐和条件去噪，有效缓解了用户间干扰，并接近理论吞吐量上限。采用LDPC编码后，所有方法均受益于编码增益，但WiFo-MUD仍优于其他方法，尤其是在低信噪比区域保持了更高的吞吐量。这些结果证实了WiFo-MUD方法在各种场景中的有效性。

现在，我们对所提出的方法进行超参数分析，重点关注补丁大小的敏感性，以及模型规模和数据集规模变化时的性能。通过改变用于分割条件嵌入中信道条件的补丁大小，我们在数据集Q1上对模型进行预训练和测试，并在表 V 中比较计算复杂度。结果表明，补丁过小虽然能使模型捕捉信道条件中的细粒度相关性，却会妨碍其提取大规模依赖关系，导致准确率下降。相反，补丁过大则会导致特征提取过于粗略，同样会降低性能。因此，选择适当的补丁大小对于有效提取特征至关重要。此外，增大补丁大小会减少输入 WirelessDiT 模块的词元数量，从而降低总体计算成本。

我们进一步比较了在不同规模数据集上预训练的三种 WiFo-MUD 模型规模的性能，结果见

表 V：不同补丁大小下的性能

补丁大小	Q1 上的误码率	浮点运算次数 (M)
(2, 2)	0.0812	38.45
(2, 4)	0.0805	35.92
(4, 2)	0.0794	35.92
(4, 4)	**0.0782**	32.86
(4, 8)	0.0812	30.67
(8, 4)	0.0820	30.67
(8, 8)	0.0861	**29.73**

表 VI：MU-Aligner 模块的消融实验

模型	使用 MU-Aligner	不使用 MU-Aligner
全样本比特错误率	**0.0782**	0.0881
零样本误码率	**0.0891**	0.0974
推理时间 [毫秒]	0.57	0.53

图7。当在大小相同的数据集上训练时，随着模型参数量增加，全样本准确率和零样本准确率均有所提升。全样本准确率的提升更为显著，而零样本准确率的提升则相对平缓。对于固定的模型规模，增加预训练数据量能够持续提升全样本准确率和零样本准确率，其中后者受益更为明显。这表明，更大的训练数据集能够有效提升模型对分布外数据的泛化能力。这些观察结果初步表明，与大语言模型类似，规模效应同样存在于面向无线通信的基础模型中。

我们进一步通过图8和表 VI 验证 MU-Aligner 在缓解多用户不平衡问题方面的有效性。图8展示了来自数据集Q1的8个用户的信号估计值的噪声水平，并比较了初始的线性最小均方误差粗略估计、应用 MU-Aligner 后的估计值，以及经过完整 WiFo-MUD 处理后的最终估计值。比较结果表明，MU-Aligner 通过时间步感知去噪，有效地均衡了各用户之间的信噪比。随后，WirelessDiT 进一步降低了总体估计误差。表 VI 给出了针对 MU-Aligner 模块的消融研究。结果表明，该模块仅带来时延的 marginal 增加，便显著提升了最终解调准确率。

我们进一步进行了消融实验，以评估所提出的通信感知一致性蒸馏机制的有效性。对于预训练和蒸馏后的 WiFo-MUD-Base 模型，我们保持用户分组阈值不变，同时改变推理步数。详细的推理过程见 [34]。如表VII所示，随着推理步数增加，全样本和零样本准确率均有所提升，这得益于去噪能力的增强。与原始的 1000 步推理相比，单步推理的准确率仅略有下降，证实了我们的蒸馏方法能够有效保留噪声分布知识。同时，总推理时间随步数近似线性增长。因此，通过适当选择

表VII：蒸馏方案的消融实验

推理步数	1	2	5	10	100	1000
全样本比特错误率	0.0819	0.0816	0.0814	0.0812	0.0810	**0.0809**
零样本误码率	0.0902	0.0899	0.0897	0.0894	0.0893	**0.0892**
推理时间 [毫秒]	**0.59**	1.08	2.46	5.49	58.96	602.14

表 VIII：蒸馏损失设计的消融实验

一致性损失权重	0	1	0.5	自适应
全样本比特错误率	0.0826	0.0823	0.0820	**0.0819**
零样本误码率	0.0941	0.0907	0.0906	**0.0902**

推理步数，模型便能保持准确率优势，而不会增加额外时延。

我们还在表 VIII 中研究了蒸馏过程中不同损失函数的影响。损失加权的变化反映了一致性损失与均方误差损失之间的权衡。可以观察到，仅使用均方误差损失会妨碍从教师模型有效迁移知识，导致零样本性能严重下降。相反，仅依赖一致性损失会导致准确率相比教师模型出现不可忽视的下降。尽管采用固定权重的组合损失比单一损失配置有所改善，但零样本准确率仍会下降。相比之下，我们提出的自适应加权策略在全样本和零样本场景下均取得了最佳性能。

我们还验证了动态用户分组策略的有效性。表 IX 比较了 WiFo-MUD 在数据流数量上限不同时的解调精度。理论最大值由 N_r 决定。上限越低，每组处理的用户越少，因此分组数量越多。结果表明，随着组规模减小，解调精度会提高，这说明更精细的分组能够缓解推理过程中的用户间干扰，并降低整体误码率。不过，这会导致推理运行次数增加，计算负载也更高。这种行为与非线性判决反馈解调器的机制类似。

在图 9 中，我们分析了不同方案在用户接入密度各异的场景下的准确率。在 $N_r = 32$ 且每个用户配备 2 根天线时，用户数量从 4 增至 20。在这种条件下，由于信噪比差异和用户间干扰加剧，所有基线解调器的误码率都会上升。其中，ALD 解调器会受到评分网络中的误差累积影响

在用户密度较高时，这种现象会进一步加剧，并在迭代采样过程中放大，导致其准确率低于 CoDiPhy。相比之下，WiFo-MUD 的动态用户分组策略仅带来微小的复杂度开销，便能有效缓解用户间干扰，使其在用户数量增加时仍保持较高的准确率。

最后，我们比较了各方法的计算开销。尽管 WiFo-MUD 的参数量多于其他基于深度学习的模型，但经过一致性蒸馏后，它只需进行单步推理，因此显著减

表 IX：用户分组方案消融实验

分组阈值	N_r	$N_r/2$	$N_r/4$	ALD
全样本比特错误率	0.0833	0.0819	**0.0815**	0.0843
零样本误码率	0.0913	0.0902	**0.0898**	0.0909
推理时间 [毫秒]	**0.33**	0.59	1.26	40.46

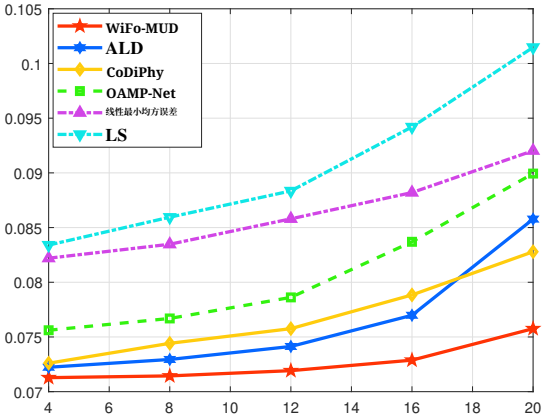


图 9：用户数量变化时的误码率比较。

少了浮点运算次数，远低于迭代式扩散基准方法。MU-Aligner 采用精简的层设计并支持并行推理，带来的时延可以忽略不计。如表 X 所示，WiFo-MUD 的推理时间与小模型方法相当，证实了其适用于实时系统。

VII. 结论

本文介绍了 WiFo-MUD，这是一种面向异构配置下多用户信号解调的基础模型。该架构由 MU-Aligner 和 WirelessDiT 模块组成，针对无线信号特性和多用户解调任务进行了设计。通过基于扩散的自监督预训练，该模型能够从粗略且不均衡的估计中恢复发送信号。为提升推理效率，本文采用通信感知的一致性蒸馏方案，实现单步采样，并引入动态用户分组策略。在大规模异构数据集上的评估结果表明，WiFo-MUD 在全样本和零样本设置下均以单一模型实现了最先进性能，展现出在物理层通信系统中的巨大潜力。

参考文献

[1] S. Yang 和 L. Hanzo，《MIMO检测五十年：迈向大规模MIMO之路》，*IEEE*通信综述与教程，第17卷，第4期，第1941–1988页，2015年第四季度。[2] M. A. Albreem、M. Juntti 和 S. Shahabuddin，《大规模MIMO检测技术：综述》，*IEEE*通信综述与教程，第21卷，第4期，第3109–3132页，2019年第四季度。[3] X. Lin，《3GPP Release 18中5G Advanced演进概述》，*IEEE*通信标准杂志，第6卷，第3期，第77–83页，2022年9月。[4] Z. Yang、S. Gao、X. Cheng 和 L. Yang，《叠加式IM-OFDM（S-IM-OFDM）：一种用于一体化感知与通信的增强型OFDM》，*IEEE*车辆技术汇刊，第73卷，第10期，第15 832–15 836页，2024年10月。[5] B. Liu、S. Gao、Z. Yang、X. Cheng 和 L. Yang，《嵌入波束方向图调制的一体化感知与通信混合收发器优化》，*IEEE*无线通信汇刊，第1–1页，2025年。

表 X：不同方案之间的性能和复杂度比较

方法	WiFo-MUD-Small	WiFo-MUD-Base	WiFo-MUD-Large	ALD	CoDiPhy	OAMP-Net
全样本SER	0.0853	0.0819	**0.0803**	0.0843	0.0857	0.0879
零样本SER	0.0906	0.0902	**0.0895**	0.0909	0.0932	0.1200
参数量 (M)	1.29	3.92	5.83	1.34	1.19	**0.56**
浮点运算次数 (M)	15.47	32.86	59.35	15.80*70	5.16*500	**12.64**
推理时间（毫秒）	0.36	0.59	0.73	40.46	34.61	**0.34**

[6] X. Cheng 等，《智能多模态感知通信一体化：机器联觉》，IEEE通信综述与教程，第26卷，第1期，第258–301页，2023年第一季度。[7] A. Patel 和 B. Kosko，《Neyman–Pearson准则及不等式约束统计信号检测中的最优噪声增益》，IEEE信号处理汇刊，第57卷，第5期，第1655–1669页，2009年5月。[8] Z. Guo 和 P. Nilsson，《用于MIMO检测的K最佳球形译码算法与实现》，IEEE通信领域精选期刊，第24卷，第3期，第491–503页，2006年3月。[9] H. He、X. Yu、J. Zhang、S. Song 和 K. B. Letaief，《消息传递与图神经网络相遇：大规模MIMO系统的新范式》，IEEE无线通信汇刊，第23卷，第5期，第4709–4723页，2024年5月。[10] S. Wu、L. Kuang、Z. Ni、J. Lu、D. Huang 和 Q. Guo，《使用近似消息传递的大规模多用户MIMO-OFDM系统低复杂度迭代检测》，IEEE信号处理专题期刊，第8卷，第5期，第902–915页，2014年10月。[11] J. Ma 和 L. Ping，《正交AMP》，IEEE Access，第5卷，第2020–2033页，2017年1月。[12] Z. Yang、S. Gao、X. Cheng 和 L. Yang，《双重动态车联网中机器联觉（SoM）增强的ISAC预编码》，IEEE通信汇刊，2025年，提前在线发表。[13] ——，《空地网络中机器联觉（SoM）增强的亚太赫兹ISAC传输》，2025年，arXiv:2506.12831。[14] H. He、C.-K. Wen、S. Jin 和 G. Y. Li，《用于MIMO检测的模型驱动深度学习》，IEEE信号处理汇刊，第68卷，第1702–1715页，2020年2月。[15] Y. Zhang、J. Sun、J. Xue 和 Z. Xu，《利用超网络进行大规模MIMO检测的跨信道模型驱动学习》，IEEE无线通信汇刊，第24卷，第3期，第1964–1977页，2025年3月。[16] S. Khan、K. S. Khan 和 S. Y. Shin，《利用深度卷积神经网络残差学习对高阶M-QAM进行符号去噪》，载于2019年第16届IEEE年度消费通信与联网会议（CCNC），2019年，第1–6页。[17] K. Pratik、B. D. Rao 和 M. Welling，《RE-MIMO：循环式置换等变神经MIMO检测》，IEEE信号处理汇刊，第69卷，第459–473页，2021年1月。[18] H. He、C.-K. Wen、S. Jin 和 G. Y. Li，《用于MIMO检测的模型驱动深度学习网络》，载于2018年IEEE全球信号与信息处理会议（GlobalSIP），2018年，第584–588页。[19] J. Ho、A. Jain 和 P. Abbeel，《去噪扩散概率模型》，载于《神经信息处理系统进展》，H. Larochelle、M. Ranzato、R. Hadsell、M. Balcan 和 H. Lin 编，第33卷，Curran Associates, Inc.，2020年，第6840–6851页。[20] Y. Song、J. Sohl-dickstein、D. P. Kingma、A. Kumar、S. Ermon 和 B. Poole，《通过随机微分方程实现基于得分的生成建模》，2021年。 [在线]。可访问：https://openreview.net/forum?id=PxTIG12RRHS [21] N. Zilberstein、C. Dick、R. Doost-Mohammady、A. Sabharwal 和 S. Segarra，《用于大规模MIMO检测的退火朗之万动力学》，IEEE无线通信汇刊，第22卷，第6期，第3762–3776页，2023年6月。[22] N. Zilberstein、A. Swami 和 S. Segarra，《基于扩散模型的大规模MIMO系统联合信道估计与数据检测》，载于ICASSP 2024——2024年IEEE声学、语音与信号处理国际会议，2024年，第13 291–13 295页。[23] K. Ying、Z. Gao、S. Chen、T. Q. S. Quek 和 H. V. Poor，《生成式扩散模型驱动的大规模MIMO系统大规模随机接入》，2025年，arXiv:2505.12382。[24] N. Peyman 和 M. Jian，《CoDiPhy：将去噪扩散模型应用于无线通信系统物理层的通用框架》，2025年，arXiv:2503.10297。[25] T. Wu、Z. Chen、D. He、L. Qian、Y. Xu、M. Tao 和 W. Zhang，《CDDM：用于无线语义通信的信道去噪扩散模型》，IEEE无线通信汇刊，第23卷，第9期，第11 168–11 183页，2024年。[26] B. Liu、S. Gao、X. Liu、X. Cheng 和 L. Yang，《WiFo：用于信道预测的无线基础模型》，《中国科学：信息科学》，第68卷，第8期，第162302页，2025年5月。[27] X. Liu、S. Gao、B. Liu、X. Cheng 和 L. Yang，《WiFo-CF：用于CSI反馈的无线基础模型》，2025年，arXiv:2508.04068。[28] X. Cheng、B. Liu、X. Liu、E. Liu 和 Z. Huang，《基础模型赋能的机器联觉（SoM）：AI原生智能多模态感知通信一体化》，IEEE网络科学与工程汇刊，第1–21页，2025年，提前在线发表。

[29] L. He、Z. Wang 和 Y. Huang，《基于去噪扩散模型的大规模多输入多输出采样检测策略》，发表于*2024*年国际无线通信与移动计算会议（*IWCMC*），2024年，第1443–1448页。[30] N. Zilberstein、C. Dick、R. Doost-Mohammady、A. Sabharwal 和 S. Segarra，《基于退火欠阻尼朗之万动力学的加速大规模多输入多输出检测器》，发表于*ICASSP 2023——2023*年*IEEE*声学、语音与信号处理国际会议（*ICASSP*），2023年，第1–5页。[31] Y. Li 等，《去噪与增强：扩散模型在增强信道状态信息恢复中的双重用途》，2025年，*arXiv:2510.02744*。[32] V. Vukoti、C. Raymond 和 G. Gravier，《借助对话感知双向GRU网络，在局部观测基础上进一步实现口语理解》，发表于*Interspeech 2016*，2016年。[33] W. Peebles 和 S. Xie，《基于变换器的可扩展扩散模型》，2023年。

[34] [ ] 2212 09748 在线。可访问：https://arxiv.org/abs/。Y. Song、P. Dhariwal、M. Chen 和 I. Sutskever，《一致性模型》，*arXiv*预印本 *arXiv:2303.01469*，2023年。[35] S. Jaeckel、L. Raschkowski、K. B¨orner 和 L. Thiele，《QuaDRiGa：一种具有时间演进特性的三维多小区信道模型，可用于虚拟现场试验》，*IEEE*天线与传播汇刊，第62卷，第6期，第3242–3256页，2014年。[36] A. Alkhateeb，《DeepMIMO：适用于毫米波和大规模多输入多输出应用的通用深度学习数据集》，2019年，*arXiv:1902.06435*。[37] X. Cheng 等，《SynthSoM：用于机器联觉（SoM）的合成智能多模态感知通信数据集》，*Scientific Data*，第12卷，第819号，2025年5月。[38] C. Shepard、J. Ding、R. E. Guerra 和 L. Zhong，《理解真实的多天线MU-MIMO信道》，发表于*2016*年第*50*届阿西洛马会议：信号、系统与计算机，2016年，第461–467页。[39] F. Euchner、M. Gauger、S. Doerner 和 S. ten Brink，《一种由“大型信道状态信息数据”驱动机器学习的分布式大规模多输入多输出信道探测仪》，发表于*WSA 2021*；第*25*届国际*ITG*智能天线研讨会，2021年，第1–6页。