生成式解压缩：分布失配下的最优有损解码

Saeed R. Khosravirad，*IEEE* 高级会员，Ahmed Alkhateeb，*IEEE* 会士，
以及 Ingrid van de Voorde，*IEEE* 会员

arXiv:2602.03505v1 [cs.IT] 3 Feb 2026

摘要—**本文探讨有损压缩中的最优解码策略：压缩器设计所依据的分布与信源的实际（真实）分布不匹配。这一问题与标准化通信系统直接相关，因为解码器可能获得关于真实分布的侧信息或先验信息，而固定编码器无法获取这些信息。我们正式定义了失配量化问题，并证明最优重构规则（即生成式解压缩）与经典贝叶斯估计相一致：在给定量化索引的条件下，根据真实分布计算条件期望，并使其适应固定编码器约束。该策略实质上是在解码器端进行生成式贝叶斯校正，且性能严格优于传统的质心规则。我们将这一框架扩展到噪声信道传输，推导出一种稳健的软解码规则，用以量化失配情况下标准化模块式信源—信道分离架构的低效性。此外，我们将该方法推广至面向任务的解码，表明最优策略会从条件均值估计转变为最大后验概率（MAP）检测。针对高斯源和基于深度学习的语义分类开展的实验结果表明，生成式解压缩弥合了与理想联合优化基准之间的大部分性能差距，从而无需修改编码器即可实现自适应、高保真度重构。**

索引词—**量化、失配、失真最小化、联合信源信道编码、生成式解码。**

一、引言

有损数据压缩是媒体和遥测通信与存储中的基础构件。简而言之，它将连续值随机变量映射到有限个离散级别，以实现高效表示[1]。通常，压缩器的设计目标是在假定的源分布下最小化失真，例如均方误差（MSE）。然而，在实际场景中，由于估计误差、非平稳性或模型不准确性，真实源分布可能偏离设计假设 [2]。

当接收端已知此类分布失配时，解码器便可利用这一侧信息调整重构规则，从而降低失真。这一概念根植于经典信息论，具体而言，是关于解码器端存在侧信息时进行源编码的 Wyner-Ziv 定理 [3]，但近年来它在更多领域重新受到关注，尤其是在生成式通信和语义通信方面的研究与开发工作中。

Saeed R.Khosravirad 任职于美国新泽西州穆雷希尔 07974 的 Nokia Bell Laboratories（电子邮箱：saeed.khosravirad@nokia-bell-labs.com）。
Ahmed Alkhateeb 任职于亚利桑那州立大学无线智能实验室及电气、计算机与能源工程学院，地址为美国亚利桑那州坦佩 85281（电子邮箱：alkhateeb@asu.edu）。
Ingrid van de Voorde 任职于比利时安特卫普的 Nokia Bell Laboratories（电子邮箱：ingrid.van de voorde@nokia-bell-labs.com）

A. 应用场景

a) 上下文感知语义解码：在语义压缩中，解码器的目标可能会随时间变化。虽然消息编码时以信号保真度为目标，但接收端可能更关注与下游任务（例如分类）相关的特定语义特征 [4]。任务感知解码器可以调整重构结果，以最大限度地降低语义误差 [5]。

*b) CSI*压缩：在无线通信中，信道状态信息（CSI）通常由标准化编码器进行压缩，该编码器假设采用通用信道模型。然而，基站接收机会随着时间推移学习用户特定的信道统计特性。通过更新解码器以反映这些已学习的统计特性，可以显著降低CSI反馈的重构误差 [6]。

c) 联邦学习与广播：考虑一种带有接收端边信息的广播场景。中央服务器（编码器）使用针对用户平均统计特性优化的全局码本，对模型或信号进行压缩并广播。然而，每个接收端都具有各自不同的本地统计特性。每个解码器都可以利用其本地*true*分布重新解读收到的量化索引，从而实现个性化解压缩，而无需更改比特流。

B. 文献综述

统计失配条件下的最优编码问题源于经典信息论。Lapidoth [7] 刻画了信道解码的失配容量，而 Gray 和 Linder [2] 将这一结论扩展至高码率矢量量化，并证明，在失配分布上设计量化器所造成的惩罚，由设计分布与真实分布之间的相对熵（Kullback-Leibler 散度）决定。

这一框架与带侧信息的信源编码密切相关。开创性的 Slepian-Wolf 定理 [8] 及 Wyner 和 Ziv [3] 对其作出的有损扩展确立了这样一点：对于相关源，即使侧信息仅在解码器处可用，压缩效率也能接近编码器掌握侧信息时的水平。这种将复杂度转移到接收端的理论可能性，是现代分布式信源编码策略的基础 [9], [10]。

在深度学习时代，处理分布失配的方式已从解析式点密度优化转向数据驱动的生成式建模。Blau 和 Michaeli [11] 正式

定义了感知-失真权衡，并证明仅最小化失真会导致统计上不太可能出现的重建结果。为跨越这一最优前沿，Tschannen 等人 [12] 提出了分布保持有损压缩（DPLC），其解码器受到约束，只能输出来自真实数据流形的样本。对抗训练框架 [13]–[14], 进一步支持了这一生成式方法，确保在低比特率下实现高感知保真度。这些方法与神经离散表示学习 [15] 和端到端优化压缩[16], 的进展相契合，后者能够学习可在分布偏移下泛化的鲁棒编码。

这些原理正日益应用于语义通信与面向任务的通信 [17]–[19]。在此，解码器以下游任务效用而非信号保真度为优化目标，并常常借助语义韧性减轻信道或源失配的影响。该范式的一项关键应用是大规模多输入多输出（大规模多输入多输出）中的CSI反馈。尽管像CsiNet这样基于深度学习的压缩 [20] 已展现出前景，近期工作[6] 表明，基站侧的生成式解码——利用站点特定的数字孪生数据作为边信息——能够校正标准化用户设备编码器的量化伪影。最后，Zhou 等人 [21] 的有限码长分析以及近期深度联合信源信道编码（深度联合信源信道编码）方面的进展 [22] 表明，在分布失配和有限码长条件下，严格的模块化信源信道架构可能招致不可忽略的性能损失，从而促使采用联合信源信道解码策略，尽管香农分离定理在其标准联合设计假设下仍然渐近最优。

C. 贡献

尽管取得了这些进展，将这些原理应用于采用严格标准化编码器的系统仍存在一个关键缺口。大多数神经压缩框架都要求对编码器和解码器进行联合训练，因此无法兼容旧式硬件或第五代（5G）CSI反馈等固定协议。相比之下，经典失配理论虽然描述了次优编码的基本极限，却几乎没有针对严重分布偏移下实用高维重建的指导。本文通过将解码表述为一个贝叶斯推断问题来弥补这一缺口，其中编码器保持不变。我们提出生成式解压缩，以证明仅在解码器端进行调整——利用真实分布的生成先验——即可弥补因分布失配造成的显著性能损失，而无需修改发射机。在本文中，我们使用“生成式解压缩”一词，指利用信源上的生成先验进行贝叶斯重建，而非如生成对抗网络或扩散模型那样进行随机样本生成。本文的主要贡献如下：

- 我们形式化了失配量化问题：量化器固定为设计分布 $P_{(d)X}$，而解码器已知真实分布 $P_{(t)X}$。
  我们证明，最优重构规则——称为生成式解压缩——与经典的最小均方误差（MMSE）估计器一致，具体而言，就是基于真实分布的条件期望。只要存在失配，这种简单的贝叶斯校正就严格优于传统的质心规则。虽然所得估计器在数学上等同于经典的 MMSE 解，但其新意在于将其作为纯粹由解码器端实施的补丁，应用于标准化且不可变的量化协议。
- 我们将该框架扩展到噪声信道，并推导出最优软解码规则。该规则将真实源先验与信道可靠性统计不可分割地耦合起来。我们量化了在分布失配下，采用硬判决接口的标准模块化信源信道分离架构所产生的超额失真，并展示了联合软解码规则如何消除这一惩罚。
- 我们给出了高斯和拉普拉斯失配情形的闭式表达式与高码率渐近理论。我们识别出两种不同的情形——中心失配（增益 modest 且趋于消失）和尾部失配（增益显著且随码率持续存在）——并证明生成式解压能够消除过载偏差造成的失真下限，在许多实际情形下恢复理想的 $O(2^{-2b})$衰减。
- 我们表明，同一条件期望原理可推广到任意任务损失，从而得到任务最优重构，而非均方误差重构。我们在采用加权均方误差波束成形损失的5G CSI反馈，以及用于语义分类的深层向量量化变分自编码器（VQ-VAE）模型上验证了该框架，显示仅改进解码器在解析分析和高维学习场景中均能带来显著增益。

D. 论文组织结构

本文其余部分安排如下。第二节对失配量化问题进行形式化定义，介绍相关符号，并详细阐述无噪声情形下的生成式解压缩优化。第三节将理论具体应用于标量量化，推导高斯和拉普拉斯信源的显式失真表达式，并结合数值验证开展高率渐近分析。第四节将框架扩展至噪声信道上的传输，并严格量化在编码器固定时，失配条件下模块化信源信道分离的低效性。第五节将这一方法推广至任意任务特定失真度量，并针对 CSI波束成形反馈给出解析见解，同时通过大规模实验验证类别偏移和域偏移下的语义分类效果。最后，第六节总结全文并讨论更广泛的影响。

II. 问题表述

A. 信源模型与符号说明

设 X 为一个实值随机变量。 X的真实分布记为 $P_X^{(t)}$，其概率密度函数为 $f_t(x)$。编码器设计

在可能失配的设计分布 $P_X^{(d)}$ 下的量化器，其概率密度函数为 $f_d(x)$。全文中，我们使用如下简写：

$$\mathbb{E}^{(t)}[\cdot] \triangleq \mathbb{E}_{X\sim P_X^{(t)}}[\cdot], \quad \mathbb{P}^{(t)}(\cdot) \triangleq \mathbb{P}_{X\sim P_X^{(t)}}(\cdot),$$

对于设计分布也采用类似处理，并使用 $\mathbb{P}$作为概率测度。

在实践中，解码器无法通过解析方式访问 $P_X^{(t)}$，但可以从长期观测、侧信息或特定站点的数字孪生中推断它。因此，我们将 $P_X^{(t)}$ 视为解码器可用、但固定编码器不可用的统计模型；固定编码器在 $P_X^{(d)}$ 下设计一次后即保持不变。

B. 在失配模型下设计的量化器

编码器在 $P_X^{(d)}$ 下设计一个 b 位标量量化器。令 $N = 2^b$，并令 Q_d： $\mathbb{R} \to \{1, \dots, N\}$ 为该量化器，其诱导出划分 $\{\mathcal{R}_i\}_{i=1}^N$。在 $P_X^{(d)}$ 下的最优重构值为各质心：

$$a_i^d = \mathbb{E}_d[X \mid X \in \mathcal{R}_i]. \tag{1}$$

C. 生成式解压缩优化

信源服从 $f_t(x)$，从而产生索引 $I = Q_d(X)$。在给定编码器确定的固定划分边界的情况下，解码器寻求一个映射 $g_t(i) = a_i^t$，以最小化真实分布下的期望失真：

$$D_t(\mathbf{a}^t) \triangleq \mathbb{E}_t\big[(X - a_I^t)^2\big]. \tag{2}$$

命题 1. 使$D_t(\mathbf{a}^t)$ 最小的重构点对应于真实分布下的 *MMSE* 估计器：

$$a_i^{t\star} = \mathbb{E}_t[X \mid X \in \mathcal{R}_i]. \tag{3}$$

最小失真为

$$D_{\min} = \sum_{i=1}^{N} \mathbb{P}_t(I = i)\,\mathrm{Var}_t[X \mid I = i]. \tag{4}$$

证明： 这一结果直接应用了最小均方估计的正交原理。对于任意平方误差代价函数，最优估计器都是给定观测条件下随机变量的条件期望。此处，观测为事件 $\{X \in \mathcal{R}_i\}$。对每个量化区间分别取条件期望，即可得到这一结果。 ∎

最优重构规则实质上是在真实分布下，而非设计分布下应用经典的 Lloyd-Max 质心条件。超额失真为 $\Delta D \triangleq D_t(\mathbf{a}^d) - D_{\min} \geq 0$。

III. 案例研究：标量量化

A. 一比特量化（高斯真实信源）

我们将该表述专门化到高斯源。设设计分布为 $X \sim \mathcal{N}(\mu_0,\ \sigma_0^2)$，真实分布为 $X \sim\mathcal{N}(\mu_1,\ \sigma_1^2)$。

设计分布的最优 1 位量化器在均值 $\tau = \mu_0$ 处设有阈值，重构点为：

$$a_{1,2}^d = \mu_0 \mp \sigma_0\sqrt{\frac{2}{\pi}}.$$

1）未调整解码器失真：将此量化器应用于真实源 $X \sim \mathcal{N}(\mu_1,\ \sigma_1^2)$时，归一化阈值为 $\alpha \triangleq (\mu_0 - \mu_1)/\sigma_1$。使用未调整的设计质心所产生的失真为：

$$D_t(\mathbf{a}^d) = \sigma_1^2\mathbb{E}_Z\left[(Z - Q'(Z))^2\right], \tag{5}$$

其中， $Q'(Z)$ 将标准化变量 Z 映射为归一化的设计质心。利用截断正态矩，可将其展开为：

$$D_t(\mathbf{a}^d) = \sigma_1^2\Big(\Phi(\alpha)[1 - \alpha\lambda_L + 2a_1'\lambda_L + (a_1')^2] + (1 - \Phi(\alpha))[1 + \alpha\lambda_R - 2a_2'\lambda_R + (a_2')^2]\Big), \tag{6}$$

其中， $\phi(\cdot)$ 和 $\Phi(\cdot)$ 分别表示标准正态概率密度函数和累积分布函数，而 $\lambda_L = {}^{\phi}{}_{(\alpha)}\ {}_{\Phi}{}_{(\alpha)}$ 以及 ~~$\lambda_R = {}^{\phi}{}_{(\alpha)}\ {}_{1-\Phi}{}_{(\alpha)}$~~ 是逆米尔斯比率。

2）最优解码器调整：根据命题 1，最优解码器会将重构点调整为真实条件均值：

$$a_1^{t\star} = \mathbb{E}[X \mid X < \mu_0] = \mu_1 - \sigma_1\lambda_L, \tag{7}$$

$$a_2^{t\star} = \mathbb{E}[X \mid X \geq \mu_0] = \mu_1 + \sigma_1\lambda_R. \tag{8}$$

最小可达失真 $D_{\min}$ 由条件方差的加权和给出：

$$D_{\min} = \sigma_1^2\Big(\Phi(\alpha)[1 - \alpha\lambda_L - \lambda_L^2] + (1 - \Phi(\alpha))[1 + \alpha\lambda_R - \lambda_R^2]\Big). \tag{9}$$

3）方差失配情形：考虑均值匹配（$\mu_0 = \mu_1 = 0$）但方差不同（$\sigma_1 \neq\sigma_0$）的情形。此时 $\alpha = 0$且 $\lambda_L = \lambda_R = \sqrt{2/\pi}$。未调整失真为 $D_t(\mathbf{a}^d) = \sigma_1^2(1 + {}^{2}{}_{\pi\sigma_1^2})$。最优失真为 $D_{\min} = \sigma_1^2(1 - {}^{2}{}_{\pi})$。解码器自适应带来的相对增益 is:

$$\text{Gain} = 1 - \frac{D_{\min}}{D_t(\mathbf{a}^d)} = 1 - \frac{1 - 2/\pi}{1 + 2/(\pi\sigma_1^2)}. \tag{10}$$

当 $\sigma_1 \to \infty$ （编码器严重低估方差）时，增益趋近于 $2/\pi \approx 64\%$。这一结果证明，在高方差失配情形下，简单的解码器端重缩放即可弥补大部分编码损失。本节后续给出的数值实验表明，无论量化区间的数量是多少， $2/\pi$ 极限都适用于高斯方差失配的情况。

B. 多比特量化与高率分析

对于 $b > 1$ 比特，量化将实数轴划分为有界的内部单元和半无限的外部单元（过载区域）。设编码器支撑区域为 $S_d = [\tau_0, \tau_{N-1}]$，过载区域为 $\mathcal{O}_d = \mathbb{R} \setminus S_d$。为分析较大 N时的失真，我们将总失真分解为颗粒（内部）和过载（外部）两部分：

$$D \approx \underbrace{\frac{1}{12N^2}\int_{S_d}\frac{f_t(x)}{\lambda_d(x)^2}dx}_{D_{\text{granular}}} + \underbrace{\sum_{i\in\{1,N\}}\int_{\mathcal{R}_i\cap\mathcal{O}_d}(x - a_i)^2 f_t(x)dx}_{D_{\text{overload}}}. \tag{11}$$

颗粒项 D_{granular} 是 Bennett 积分 [23]，其中$\lambda_d(x)$ 表示固定编码器的点密度。只要密度函数光滑，无论采用何种译码规则，D_{granular}都会以 $O(N^{-2})$ 的速度衰减。因此，渐近性能由 D_{overload}决定。我们来分析最外侧区间 $\mathcal{R}_N = (\tau_{N-1}, \infty)$。设 $\mu^{(t)}_{\text{尾部}} = \mathbb{E}_t[X|X > \tau_{N-1}]$ 为真实分布下尾部的条件均值。设 $p_i^{(t)} \triangleq \mathbb{P}_t(X \in \mathcal{R}_i)$ 表示信源 X 落入第 i 个区间 $\mathcal{R}_i$的真实概率。根据均方误差的偏差-方差分解 [24], ，该区间内采用任意重构值 a_N 时的失真为:

$$\begin{aligned} D^{(N)}_{\text{overload}} &= \mathbb{P}_t(\mathcal{R}_N) \cdot \mathbb{E}_t[(X - a_N)^2 \mid X \in \mathcal{R}_N] \\ &= p_N^{(t)} \left(\text{Var}_t[X \mid X \in \mathcal{R}_N] + (\mu^{(t)}_{\text{tail}} - a_N)^2 \right). \quad (12) \end{aligned}$$

方差项是不可约的，而平方偏差项 $(\mu^{(t)}_{\text{尾部}} - a_N)^2$ 取决于解码器。

1）未调整解码器：固定编码器设定 $a_N = a_N^d$，即设计分布的质心。如果 f_t 的尾部比 f_d更重（例如拉普拉斯分布与高斯分布的比较，或 $\sigma_1 \gg \sigma_0$），真实概率质量会显著向外偏移，使得 $\mu^{(t)}_{\text{尾部}} \gg a_N^d$。这会产生一个较大的、严格为正的偏差项 $(\mu^{(t)}_{\text{尾部}} - a_N^d)^2$；随着码率提高，量化误差逐渐趋于零，该偏差项将主导失真。

2）生成式解码器：所提出的解码器设定 $a_N = a_N^{t\star} = \mu^{(t)}_{\text{尾部}}$。这一选择使偏差项为零。生成式解码器的剩余失真严格等于条件方差。

$$D_{\text{overload}} = p_N^{(t)} \text{Var}_t[X \mid X > \tau_{N-1}]. \quad (13)$$

这导致生成式解压缩呈现出两种不同的渐近状态:

- **中心失配：** 如果尾部相匹配（即尾部中的概率质量以相同速率衰减），则未调整失真中的偏差项可以忽略。主导误差是 $D_{\text{量化}}$误差，因此对于较大的 b，适应性的相对增益会随着 $O(N^{-2})$ 而减小。
- **尾部失配：** 如果相对于 f_d 而言，f_t 具有厚尾特性，则未调整解码器产生的偏差项不会消失（或以比 N^{-2} 更慢的速度衰减），从而实际上形成误差平台。生成式解压缩消除了这一平台，使解码器能够在实际上扩展量化器的动态范围，从而产生随 N 持续存在或增长的增益。

为量化上述预测的失真降低，我们使用针对标准高斯 $\mathcal{N}(0, 1)$ 设计的固定编码器进行了实验，该编码器采用 Lloyd–Max算法（迭代 60 次，3×10^5 个训练样本）。我们测试了不同的比特深度$b \in \{1, 2, 3, 4\}$（$N \in \{2, 4, 8, 16\}$）。性能报告如下:

$$\text{Relative Gain} = \left(1 - \frac{D_{\min}}{D_{\text{std}}}\right) \times 100, \quad (14)$$

其中 D_{std} 使用原始设计质心，$D_{\min}$ 使用最优真实条件均值。

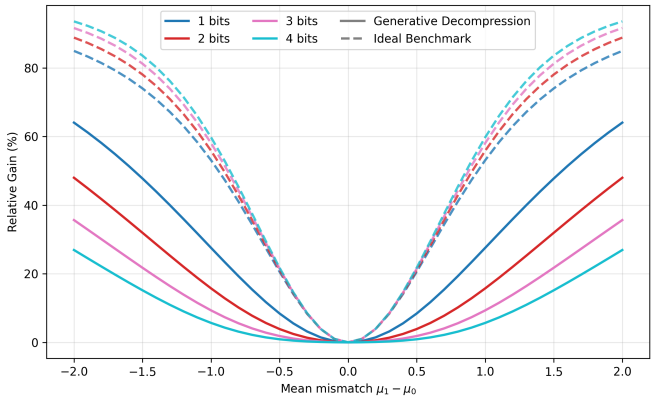

图 1。高斯源的译码端增益与均值失配的关系。针对 $\mathcal{N}(0, 1)$ 设计的量化器，在 $\mathcal{N}(\mu_1, 1)$ 上进行评估。

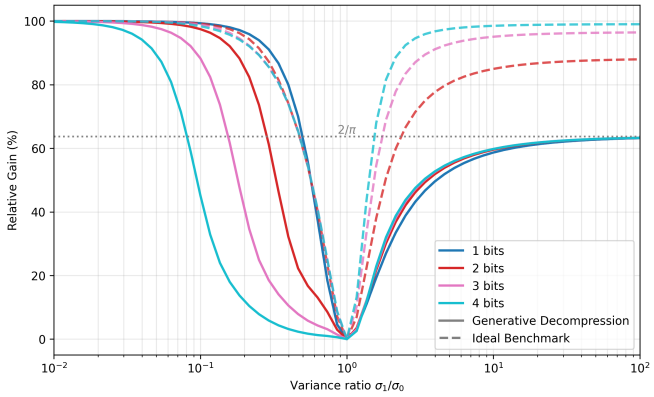

图 2。高斯源的译码端增益与方差比的关系。针对 $\mathcal{N}(0, 1)$ 设计的量化器，在 $\mathcal{N}(0, \sigma_1^2)$ 上进行评估。

a) 均值漂移：图1展示了当真实源为 $\mathcal{N}(\mu_1, 1)$，$\mu_1 \in [-2, 2]$时的相对增益。增益在 $\mu_1 = 0$ 附近可忽略不计，但在 $|\mu_1| = 2$处对称上升至近60%。这证实，当信源发生漂移时，解码器自适应能够有效地使重构重新居中。理想基准情形也得到了展示（见第III-C节）。

b) 方差缩放：图 2 显示了增益与方差比 $r = \sigma_1/\sigma_0$（对数尺度）的关系。图中呈现出典型的 U 形：在过度分散区间（$\sigma_1 \gg \sigma_0$），更高的位深会带来更大的增益，因为外侧分箱的质心可以向外大幅移动，从而有效解除尾部的截断——这与前文推导的尾部失配区间一致。在分散不足区间，解码器会将重建级别向零收缩，以减少不必要的颗粒噪声。对于 1 比特的情况，理想基准与生成式解码器的性能相当。

表 I 真实拉普拉斯分布下的生成式解码。

Bits	1	2	3	4
生成式解码器相对增益	1.65%	6.09%	11.01%	16.74%
理想基准相对增益	1.65%	7.84%	24.78%	41.87%

c) 拉普拉斯分布：拉普拉斯分布的尾部按 $\propto e^{-|x|}$ 衰减，而高斯分布的尾部衰减得更快，按 $\propto e^{-x^2}$ 衰减。如表 I 所示，对于零均值、单位方差的拉普拉斯信源，生成式解压缩带来的增益随码率单调增加。在更高码率下，较重的拉普拉斯尾部会被专门的外侧分箱捕获，此时生成式解码器可以积极纠正高斯设计对尾部概率质量的严重低估——这正是产生持续高码率增益的尾部失配情形。

这些实验从数值上证实了理论预测：在纯粹的中心失配下，解码器自适应带来的增益有限；但只要真实分布的尾部质量超过设计分布的预期，增益就会显著提升，并且不会随码率提高而消失（甚至还会增加）。

C. 与理想匹配量化进行基准比较

要严格评估生成式解压缩的效果，我们不仅要将其与未调整解码器进行比较，还要与理想基准进行比较：在该系统中，编码器和解码器都与真实分布 $f_t(x)$ 完美匹配。

设固定编码器的点密度函数为$\lambda_d(x) \propto f_d(x)^{1/3}$。我们定义三个失真值：

1) 理想基准（D_{ideal}）：编码器和解码器均针对 $f_t(x)$ 重新训练。根据 Panter-Dite 渐近关系 [1], [25]:

$$D_{\text{ideal}} \approx \frac{1}{12N^2}\|f_t\|_{1/3}^3, \tag{15}$$

其中 $\|f\|_\alpha \triangleq \left(\int f^\alpha dx\right)^{1/\alpha}$。

2) 未调整解码器（D_{fix}）：解码器使用针对 f_d设计的质心。这会在每个区间内引入系统偏差；如果 f_t 的支持集超出 f_d，那么随着 $N \to \infty$，这种偏差不会消失。

3) 生成式解码器（D_{gen}）：编码器固定为 λ_d，但重构是最优的。失真由失配 Bennett 积分 [23]给出：

$$D_{\text{gen}} \approx \frac{1}{12N^2}\int \frac{f_t(x)}{\lambda_d(x)^2}dx. \tag{16}$$

定义 1 (失配惩罚因子)。 令 $\mathcal{L}$ 为生成式失真与理想失真的渐近比值：

$$\mathcal{L} \triangleq \lim_{N\to\infty}\frac{D_{gen}}{D_{ideal}} = \frac{\int f_t(x)\lambda_d(x)^{-2}dx}{\|f_t\|_{1/3}^3}. \tag{17}$$

该因子使我们能够区分生成式解码器可恢复和不可恢复的情形。在温和的正则性条件下，式（17）中的积分是有限的——例如，当设计密度在尾部的衰减速度不快于真实密度时——因此失配惩罚因子有明确的定义。

我们首先考察二元量化的基本极限（$N = 1$）。对于尺度族分布（例如方差不同的高斯分布或拉普拉斯分布），这一情形具有强最优性结论。

定理 1 (在 $N = 1$ 处达到完美最优性)。 设设计分布与真实分布关于同一均值对称且单峰，二者仅在方差（尺度）上不同。对于 *1* 位标量量化器，生成式解码器能够精确达到理想失真：

$$D_{gen} = D_{ideal} < D_{fix}. \tag{18}$$

在这种情形下，失配惩罚因子趋于零（$\mathcal{L} = 1$）。这一结论适用于所考虑的标量、*1* 比特设置。

证明：对于任意以 μ为中心的对称单峰分布，最优1比特量化阈值是唯一的，且等于 μ。由于设计分布与真实分布共享 μ，固定编码器的划分 $\mathcal{R}_1 = (-\infty, \mu)$ $\mathcal{R}_2 = [\mu, \infty)$与针对真实源优化的理想编码器的划分相同。由于这些划分相同，$D_{\text{gen}} = D_{\text{ideal}}$。然而，未调整解码器使用质心 $a^d \neq a^{t\star}$，由于III-A节中推导的平方偏差项，会产生严格更高的失真。 ∎

定理2（失真层级关系）。 对于任意码块长度 N以及设计划分不同于最优真实划分的任意失配情形，失真满足如下严格层级关系：

$$D_{ideal} < D_{gen} < D_{fix}. \tag{19}$$

等式 $D_{gen} = D_{fix}$ 仅当设计质心恰好与真实条件均值一致时成立。等式$D_{ideal} = D_{gen}$ 仅当设计划分对真实源最优时成立（例如，定理中的对称 $N = 1$ 情形 *1)*.

证明：不等式 $D_{\text{gen}} \leq D_{\text{fix}}$ 由命题1中生成式解码器的定义可得。对于固定的划分 $\mathcal{P}$，函数 $g(\mathbf{a}) = \mathbb{E}_t[(X - a_I)^2]$ 严格凸，且唯一在条件均值 $\mathbf{a}^{t\star}$处取最小值。由于未调整解码器使用质心 $\mathbf{a}^d \neq \mathbf{a}^{t\star}$（在非平凡失配的情况下），因此严格有 $D_{\text{gen}} < D_{\text{fix}}$。

不等式 $D_{\text{ideal}} \leq D_{\text{gen}}$ 由优化空间的约束可得。令 $\mathbb{Q}_N$ 为所有可能的 N级量化器（划分与码本组成的配对）的集合。理想失真是全局最小值：$D_{\text{ideal}} = \min_{(P,C)\in\mathbb{Q}_N} \mathbb{E}_t[(X - Q(X))^2]$。生成式失真是一个受约束的最小值，其中划分 P 固定为设计划分 P_d：$D_{\text{gen}} = \min_C \mathbb{E}_t[(X - Q_{P_d,C}(X))^2]$。由于受约束的最小值不可能严格小于全局最小值，$D_{\text{ideal}} \leq D_{\text{gen}}$。当且仅当 P_d不是 f_t的最优沃罗诺伊划分时，该不等式为严格不等式；在方差失配下，$N > 1$ 确实如此，因为最优单元宽度取决于信源密度。 ∎

在更高码率（$N \to \infty$）下，编码器的划分密度成为主导因素。此时，固定编码器不再是最优的，但生成式解码器能够对收敛速率进行关键修正。

定理 3 (渐近码率恢复)。 假设真实分布 $P_X^{(t)}$ 是绝对连续的，且其尾部比设计分布 $P_{(d)X}$ 更重（例如，$f_t(x)/f_d(x)$ 在 $|x|$ 上最终单调不减，并且在过载区域内有正下界）。在存在这种分布

失配时，未调整解码器失真会饱和于一个常数下限，而生成式解码器则恢复理想的渐近衰减率 $O(N^{-2})$。

$$\lim_{N\to\infty} D_{fix} > 0, \tag{20}$$

$$\lim_{N\to\infty} N^2 D_{gen} = constant. \tag{21}$$

证明：未调整解码器的误差可分解为方差（量化误差）+偏差的平方。当 $N \to \infty$ 时，量化误差趋于零，但由于过载区域存在失配，偏差 $\mathbb{E}_t[(X - a_I^q)^2 \mid \mathcal{O}_d]$ 仍严格大于零（见第 III-B 节）。生成式解码器通过将质心设为条件均值 $a_I^{t\star}$ 消除了这一偏差，因此只剩下条件方差，其随单元宽度的平方缩放$\Delta_i^2 \propto N^{-2}$。 ∎

虽然定理 3 证明生成式解压缩恢复了正确的码率 收敛速度，但与理想基准相比，仍存在一个常数惩罚因子。

引理 1（高斯失配的不对称性）。设真实分布和设计分布分别为方差为σ_1^2 和 σ_0^2的零均值高斯分布。令 $r = \sigma_1/\sigma_0$ 为失配比。惩罚因子 $\mathcal{L}$ 对 r的依赖具有不对称性：

- 分辨率损失，$r \to 0$：如果信源比设计分布更窄（$\sigma_1 \ll \sigma_0$），惩罚将发散：$\mathcal{L} \to \infty$。
- 尾部失配，$r > 1$：如果信源比设计范围更宽（$\sigma_1 > \sigma_0$），则惩罚项 $\mathcal{L}$增长缓慢，前提是渐近积分收敛。

证明：将高斯点密度 $\lambda_d(x) \propto \exp(-x^2/3\sigma_0^2)$代入（17）中的 $\mathcal{L}$ 定义，积分核的中心项决定了量化误差系数。当$r \to 0$时，$\lambda_d(x)^{-2}$ 这一比值引入了过大的权重，因为设计点密度分布得过于稀疏。这种粗粒度量化误差主导了失真，任何生成式校正都无法恢复丢失的分辨率，因此有 $\mathcal{L} \to \infty$。当 $r > 1$时，固定编码器的分辨率高于信源核心区域所需的分辨率。尽管这效率不高，但惩罚项 $\mathcal{L}$受 σ_1/σ_0 比值控制，其增长速度远低于 $1/r^2$。至关重要的是，固定解码器中不会消失的偏差项会被生成式解码器消除，从而使系统只受到次优分箱宽度的惩罚，并实现对理想增益的大幅恢复。 ∎

引理 1 刻画了恢复的效率 。在分辨率损失区间（$r < 1$）中，性能差距主要由固定编码器的粗粒度所决定——这是一种不可逆的信息损失，限制了任何解码器接近理想基准的程度。相比之下，在尾部失配区间（$r > 1$）中，主要误差来源是对外侧分箱尺度的判定。由于这更像是一个校准问题，生成式解码器能够成功弥合这一差距，恢复理想增益中的很大一部分。

IV. 噪声信道上的鲁棒生成式解压缩

在实际工程系统中，例如无线系统中的 CSI反馈，量化索引会通过有噪信道传输。我们将该框架扩展到解码器观测到编码器输出的有噪版本的场景。

A. 软解码规则的推导

设 $I = Q_d(X)$ 为传输索引，$\hat{}I$ 为经过具有转移概率 $P(I|I)$ 的离散无记忆 $\hat{}$ 信道后接收到的索引。$\hat{}$ 解码器需要找到一个估计器 $g(I)$，使其在真实源分布下的均方误差最小：

$$\min_g \mathbb{E}_t\left[(X - g(\hat{I}))^2\right]. \tag{22}$$

解是给定接收 符号后的后验均值。根据全期望定律，可将其展开为由命题1推导出的生成质心的软加权组合：

$$\begin{aligned} a_{\hat{i}}^{t,\text{noisy}} &= \mathbb{E}_t[X \mid \hat{I} = \hat{i}] \\ &= \sum_{i=1}^{N} P(I = i \mid \hat{I} = \hat{i}) \cdot \underbrace{\mathbb{E}_t[X \mid I = i]}_{a_i^{t\star}}. \end{aligned} \tag{23}$$

根据贝叶斯定理，加权项同时取决于信道统计特性和真实信源先验 $p_i^{(t)} = \mathbb{P}_t(X \in \mathcal{R}_i)$：

$$P(I = i \mid \hat{I} = \hat{i}) = \frac{P(\hat{I} = \hat{i} \mid I = i) p_i^{(t)}}{\sum_{k=1}^{N} P(\hat{I} = \hat{i} \mid I = k) p_k^{(t)}}. \tag{24}$$

噪声信道下的最优重构规则 $a^{t,\text{noisy}}{}_{\hat{i}}$ 是一个基于后验分布 $P(X \mid \hat{}I = \hat{}i)$ 进行运算的最小均方误差软估计器。该过程可分解为两个依次进行的不同校正：一个用于处理源失配（质心的生成偏移），另一个用于处理信道误差。

为获得质心的最优生成偏移，必须首先将基础重构值设为真实条件均值 $a_i^{t\star}$，该均值由真实源概率密度函数 $f_t(x)$ 和固定分箱边界 $\mathcal{R}_i$ 得出。此步骤确保在考虑信道不确定性之前，解码器的码本在统计上是准确的。

$$a_i^{t\star} = \frac{\int_{\mathcal{R}_i} x f_t(x) dx}{\int_{\mathcal{R}_i} f_t(x) dx} = \mathbb{E}_t[X \mid X \in \mathcal{R}_i] \tag{25}$$

然后，为了处理信道误差，我们使用接收到的含噪索引 $\hat{}i$ 来计算后验概率 $P(I = i \mid \hat{}I = \hat{}i)$。随后利用该后验概率对式 (25) 中计算出的生成质心进行线性组合，从而得到总体最小均方误差估计

$$a_{\hat{i}}^{t,\text{noisy}} = \sum_{i=1}^{N} P(I = i \mid \hat{I} = \hat{i}) \cdot a_i^{t\star}, \tag{26}$$

这等价于式 (23)。

这种解耦的两阶段校正表明，即使通信链路并不完美，生成式解压缩的优势依然存在。最优解码器将校正后的信源先验（$a_i^{t\star}$）融入信道的软解码过程，从而提升鲁棒性。这对于语义通信应用场景（ [26] ）尤为有益，因为联合信源信道编码（JSCC）的上下文感知解码器能够降低失真。

B. 示例案例：二元对称信道（二元对称信道）

考虑一个示例：针对 $\mathcal{N}(0,\ \sigma_0^2)$ 设计的 1 位量化器作用于信源 $X \sim \mathcal{N}(0,\ \sigma_1^2)$，并通过交叉概率为 ϵ 的二元对称信道（BSC）进行传输。

真实分布下的生成质心为 $a_{0,\ 1}^{t\star} = \mp\sigma_1\sqrt{2/\pi}$。由于对称性，$p_0^{(t)} = \overline{p_1^{(t)}} = 0.5$。如果解码器接收到 $I = 1$，后验概率为 $\hat{}\hat{}P(I = 1|I = 1) = 1 - \epsilon$ 和 $P(I = 0|I = 1) = \epsilon$。

解码器可以采用以下策略之一来确定最终估计值 $\hat{}x$。

a) 标准分离（未调整）：对 $\hat{}I$ 作硬判决，并使用设计质心（基于 σ_0）进行重构：

$$\hat{x}_{\text{std}} = \text{sgn}(\hat{I}) \cdot \sigma_0\sqrt{\frac{2}{\pi}}, \tag{27}$$

这会同时受到偏差（$\sigma_1 - \sigma_0$）和硬判决损失的影响。

b) 调整后的分离（硬生成式）：对 I 作硬判决，但使用真实的生成质心：

$$\hat{x}_{\text{hard}} = \text{sgn}(\hat{I}) \cdot \sigma_1\sqrt{\frac{2}{\pi}}, \tag{28}$$

这消除了偏差，但对于噪声索引仍然过度自信。

c) 最优联合检测与解码（软生成式）：最小均方误差估计器利用信道后验概率对真实质心进行软组合：

$$\hat{x}_{\text{opt}} = (1 - 2\epsilon) \cdot \text{sgn}(\hat{I}) \cdot \sigma_1\sqrt{\frac{2}{\pi}}. \tag{29}$$

因子 $(1-2\epsilon)$ 是一种向零收缩的基于可靠性的收缩。

量化分离惩罚：由于 $\hat{}x_{\text{opt}} = \mathbb{E}[X|\hat{}I]$ 是条件均值，因此任何其他估计器 $\hat{}x$ 的均方误差可分解为：

$$\mathbb{E}[(X - \hat{x})^2] = \underbrace{\mathbb{E}[(X - \hat{x}_{\text{opt}})^2]}_{D_{\text{opt}}} + \underbrace{\mathbb{E}[(\hat{x} - \hat{x}_{\text{opt}})^2]}_{\text{Excess Penalty}}. \tag{30}$$

这使我们只需测量分离策略与最优软重构之间的偏差，即可量化其失真惩罚。

调整后的分离策略正确识别了真实的质心幅度（σ_1），但未考虑信道可靠性，实际上强制设定了 $\epsilon = 0$。偏差为：

$$\begin{aligned}\Delta_{\text{hard}} &= |\hat{x}_{\text{hard}} - \hat{x}_{\text{opt}}| \\ &= \left|1 \cdot \sigma_1\sqrt{\frac{2}{\pi}} - (1 - 2\epsilon)\sigma_1\sqrt{\frac{2}{\pi}}\right| = 2\epsilon\sigma_1\sqrt{\frac{2}{\pi}}.\end{aligned} \tag{31}$$

将其平方即可得到硬判决接口引入的精确惩罚：

$$D_{\text{hard}} - D_{\text{opt}} = 4\epsilon^2\sigma_1^2\frac{2}{\pi}. \tag{32}$$

标准分离策略会产生额外的平方偏差，因为它使用的是设计方差 σ_0 而非真实方差 σ_1。利用估计量定义 $\hat{}x_{\text{std}} = \sigma_0 K$ 和 $\hat{}x_{\text{hard}} = \sigma_1 K$（其中 $K = \text{sgn}(\hat{}I)\sqrt{2/\pi}$），我们可以明确比较它们相对于最优 $\hat{}x_{\text{opt}} = (1 - 2\epsilon)\sigma_1 K$ 的超额失真：

$$D_{\text{std}} - D_{\text{opt}} \propto (\sigma_0 - (1 - 2\epsilon)\sigma_1)^2, \tag{33}$$

$$D_{\text{hard}} - D_{\text{opt}} \propto (2\epsilon\sigma_1)^2. \tag{34}$$

在低噪声极限（$\epsilon \to 0$）下，标准惩罚会简化为源失配偏差（$\sigma_0 - \sigma_1$），而硬判决惩罚则趋于零。因此，对于任意非平凡的失配，只要 σ_0 并非恰好充当收缩因子，失真层级关系就严格为：

$$D_{\text{std}} > D_{\text{hard}} > D_{\text{opt}}. \tag{35}$$

差值 $D_{\text{std}} - D_{\text{hard}}$ 表示源偏差惩罚（通过生成式解压缩解决），而 $D_{\text{hard}} - D_{\text{opt}}$ 表示分离惩罚（通过软聚合解决）。

这一闭式分析对信源-信道分离定理有直接启示。该定理指出，当模型完全已知时，分别优化信源编码和信道编码在渐近意义下是最优的 [27]。然而，在我们的设定中，编码器固定且信源模型存在失配，因此，严格模块化的硬判决分离架构在语义失配 [28] 或编码失配 [21]下，其实际性能会受到额外惩罚。我们的结果精确量化了这一架构惩罚，并得出两点关键见解：

1）信道可靠性直接缩放幅度：最优估计器中的因子 $(1 - 2\epsilon)$ 代表接收比特的有效可靠性。当 $\epsilon = 0$ 时，我们恢复完整的生成质心 $\pm\sigma_1\sqrt{2/\pi}$。当 $\epsilon = 0.5$ ——（纯噪声）时，重构归零。当信道不可靠时，最优解码器会有效地收缩估计值，以最小化方差。

2）分离架构的次优性：严格分离的设计首先执行硬信道解码，然后应用固定的重建表。由于接口是在错误的方差（σ_0）假设下固定的，并且该架构强制采用硬判决，这种解码器无法生成因子 $\sigma_1(1 - 2\epsilon)$。如（32）所示，分离式解码器注入了与 ϵ^2 成正比的不必要功率，证明最优重构必须将真实源方差 σ_1 与信道可靠性不可分割地结合起来。

总之，即使在这种标量设置下，分布失配也暴露出带有硬判决接口的严格模块化分离架构的低效性。尽管分布感知质心（硬生成式策略）能够消除偏差，但要实现最优性能，仍需要联合信源-信道感知——即使香农分离定理在满足其建模和联合设计假设时仍然成立，这一特性也无法在采用冻结编码器的严格模块化设计中实现。

d) 与理想联合信源信道编码进行基准测试：终极基准是一个系统，其中编码器也针对真实源方差 σ_1 和信道噪声ϵ重新优化。对于二元对称信道上的对称单峰信源，无论

噪声水平或信源方差 [29]如何，最优量化阈值都是均值（）。由于固定编码器已经采用了这一阈值，我们提出的软生成式解码器达到了联合信源信道编码的全局下界：

$$D_{\text{opt}} = D_{\text{ideal}}^{\text{JSCC}}. \tag{36}$$

这意味着，对于二进制信令，固定编码器不会带来任何损失——所有增益都可以在解码器端恢复。然而，对于 $N > 1$，理想的联合信源信道编码编码器会优化比特映射（索引分配），以最大限度地减少汉明距离错误——这是固定编码器所不具备的自由度——这意味着在更高码率下，D_{opt} 最终会偏离 $D_{\text{ideal}}^{\text{JSCC}}$。

这一最优性结果依赖于这样一个事实：1 比特系统的拓扑结构很简单，发生比特翻转时总会转移到唯一的另一个质心。对于多比特系统（$N > 1$），生成式解码器与理想基准之间的关系取决于编码器的效率。对于 $N > 1$，正向差距 $D_{\text{gen}} > D_{\text{ideal}}$ 会再次出现，其大小取决于固定编码器的索引分配相对于信道噪声分布的低效程度。因此，固定编码器不会带来任何损失这一结论仅适用于这种标量二进制信令场景。

V. 失配条件下的任务感知解码

前几节提出的框架聚焦于均方误差失真。然而，在许多现代应用中，解码器的评判依据并非仅仅是重建保真度，而是取决于接收信息的下游任务。在这类场景中，设计所依据的分布与真实信源分布之间的失配依然存在，但相关的失真度量与均方误差 [11]有所不同，甚至相悖。

本节将失配量化的表述推广至一般的任务感知失真度量。我们将说明，最优解码器原则仍然成立——解码器应针对真实分布进行优化——但最优重建规则会从条件均值变为特定于任务的估计器。

A. 通用任务感知解码器优化

设 Q_d 表示在假定分布 $f_X(\cdot;\,\theta_d)$下设计的固定编码器，其生成索引 $I = Q_d(X)$。设 $d_{\mathsf{task}}(x,\ \hat{}x)$为任意特定于任务的失真度量。解码器需要寻找映射 g：$\{1,\dots,N\} \to \mathbb{R}$，使真实分布下的期望任务损失最小：

$$D_{\mathsf{task}}(g) = \mathbb{E}_{f_X(\cdot;\theta_t)}\big[d_{\mathsf{task}}(X, g(I))\big]. \tag{37}$$

给定索引 $I = i$，最优解码器需要求解 N 个相互独立的标量最小化问题：

$$g^*(i) = \arg\min_{a\in\mathbb{R}} \mathbb{E}_{\theta_t}\big[d_{\mathsf{task}}(X, a) \mid X \in \mathcal{R}_i\big]. \tag{38}$$

B. 用于波束成形的 *CSI*压缩与针对特定站点的解码器适配

在本小节中，我们将分析扩展到复值信道系数，以更准确地建模无线衰落信道。

5G新空口 (NR) CSI反馈采用固定的标准化线性组合码本，旨在适用于各种多径及空间相关信道分布。然而，在实际部署中，信道具有很强的地点特异性。gNB只能通过长期测量获知这些偏差，而UE编码器并不知情。尽管对空间相关多径信道上的目标导向失真进行数学建模通常难以实现，但使用易于处理的标量代理，推导生成式解码器可能带来的增益分析见解，仍具有重要价值。

近期研究 [6] 表明，在冻结标准化UE编码器、仅使用特定地点数据重新训练gNB侧CSI细化网络的情况下，可以在和速率相同的条件下显著降低反馈开销。这为生成式解压缩提供了大规模的实证验证。

下文将通过对CSI反馈中的关键量——主导波束的复系数——进行建模，分析驱动这些增益的主要机制。

1）任务损失模型：针对不完美 CSI 导致的下行波束成形增益（进而影响和速率）损失，一种简化但有效的代理指标是加权均方误差：

$$d_{\mathsf{CSI}}(x, \hat{x}) = |x|^2\,|x - \hat{x}|^2. \tag{39}$$

当信道系数幅度 $|x|$ 较大时，此指标会对重构误差施加更重的惩罚——也就是说，在大规模 MIMO[30]，[31]中，该情形对接收信噪比的贡献最大。需要注意的是，这种标量分析主要衡量增益估计（功率分配）的准确性，而非预编码器的空间方向（相位对齐）；后者依赖高维特征向量量化。然而，由于主导波束系数决定了有效信道增益，最小化加权幅度误差是最大化和速率的必要条件。

2）具有瑞利主导波束系数的特定站点模型：我们将主导波束系数建模为复数莱斯分布。一般性不失，旋转相位，使视距（LOS）分量位于实轴上。系数 X 为

$$X \sim \mathcal{CN}\left(\sqrt{\frac{K}{K+1}},\ \frac{1}{K+1}\right), \tag{40}$$

单位平均功率为 $\mathbb{E}[|X|^2] = 1$。

设计量化器针对码本设计中有效 $K_d \approx$ 1–3 较低到中等的典型场景进行了优化。实际的特定站点信道具有 K_t，可能偏离设计值，例如，在城市宏小区场景中可能出现的 $K_t > K_d$ （即更强的视距），或者在丰富散射场景中预期的 $K_t < K_d$。

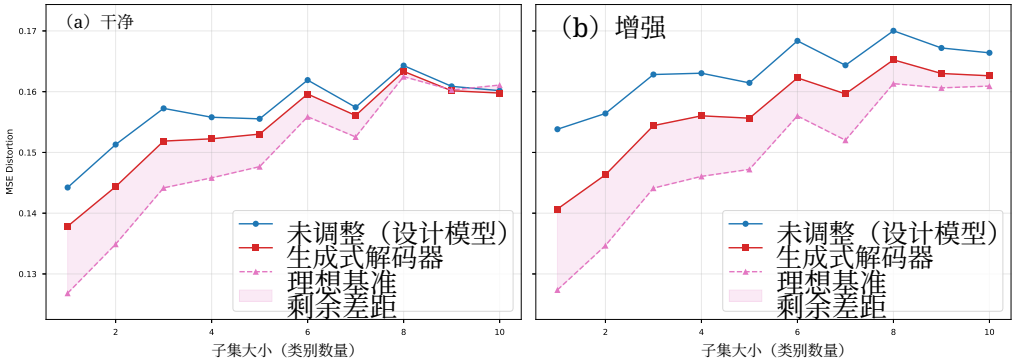

图 3.5 比特（$b = 5$）下的均方误差失真，对比干净场景和增强场景。尽管编码器被冻结，生成式解码器仍显著缩小了与理想基准之间的差距。

3) 最优与固定解码器分析：为了量化理论增益，我们定义固定（标准化）策略和自适应（生成式）策略的重构函数。

令 $\mathcal{M}_n(K) = \mathbb{E}[X^n \mid \mathrm{Re}(X) > 0; K]$ 表示参数为 K 的瑞利分布的第 n 阶条件矩。任务最优重构质心关于 K 的函数为：

$$\phi(K) \triangleq \frac{\mathcal{M}_3(K)}{\mathcal{M}_2(K)}. \tag{41}$$

我们分析一种系统失配：标准解码器使用针对设计假设 K_d 优化的固定质心 $a_{\text{fix}} = \phi(K_d)$，而生成式解码器使用真实的最优质心 $a^\star = \phi(K_t)$。

相对性能增益定义为：

$$\eta(K_t) = 1 - \frac{\mathbb{E}_{K_t}[|X|^2 |X - \phi(K_t)|^2]}{\mathbb{E}_{K_t}[|X|^2 |X - \phi(K_d)|^2]}. \tag{42}$$

我们考察真实信道 K_t 的三种情形：

a) 富散射失配（$K_t \to 0$）：此处，站点处于富散射环境（瑞利），但码本是针对主导波束（$K_d > 0$）设计的。归一化后的真实分布（$K_t = 0$）的矩为 $\mathcal{M}_2(K_t) = 1$，$\mathcal{M}_3(K_t) = \kappa = 2\sqrt{2/\pi}$ 和 $\mathcal{M}_4(K_t) = 3$。加权失真函数可简化为抛物线：

$$D(a) = 3 - 2\kappa a + a^2. \tag{43}$$

自适应解码器选择该抛物线的最小值，$a^\star = \kappa$；而固定解码器则锁定于设计分布的质心 $a_{\text{fix}} = \phi(K_d) < \kappa$。请注意，$\phi(K)$ 是一个递减函数。固定解码器造成的超额失真恰好为

$$D(a_{\text{fix}}) - D(a^\star) = (\kappa - \phi(K_d))^2, \tag{44}$$

这给出的相对增益为：

$$\eta(K_d) = \frac{(\kappa - \phi(K_d))^2}{3 - \kappa^2 + (\kappa - \phi(K_d))^2}. \tag{45}$$

增益完全源于固定解码器的僵化。对于 $K_d = 3$ 的典型设计，我们有 $\phi(3) \approx 1.3$，因此 $\eta(K_d) \approx 18\%$。自适应解码器会将重构向外偏移，以捕捉瑞利分布的重尾，从而获得一个恒定的相对增益因子，该因子由莱斯尾部积分与瑞利尾部积分之比推导得出。

b) 轻度失配 *($K_t \approx K_d$)*：考虑一种中等程度的偏差：真实位置具有 $K_t = 6$（更强的视距），而设计则假设为 $K_d = 3$。与丰富散射情形不同，在这一区域内，最优质心 $\phi(K)$ 相对稳定，其中 $\eta(6)$ 得出的相对失真降低约为 8%，尽管这种修正并非没有代价，而且在许多子带上会不断累积。

c) 视距失配（$K_t \to \infty$）：这是用户拥有纯净视距路径的情况。当 K_t 增大时，分布 X 收敛为位于 $\mu = 1$（单位功率）处的狄拉克δ函数。最优重构收敛于信号本身，

$$\lim_{K_t \to \infty} a^\star = \lim_{K_t \to \infty} \phi(K_t) = 1, \tag{46}$$

并且，自适应失真趋近于零：$D_{\text{opt}} \to |1|^2 |1 - 1|^2 = 0$。固定解码器仍锚定于 K_d，并重构 $a_{\text{fix}} = \phi(K_d) > 1$。因此，$\lim_{K_t \to \infty} \eta(K_t)$ =100%。

最优生成式解码器会自动适应特定站点的视距强度：在富散射/低-K 情况下，它会夸大较强的系数，以最大化波束成形增益；而当 K_t 较高时，则会自信地收敛到视距值。传统保真度导向解码器无法表现出这种双重行为——由于针对平均情况进行校准，它始终较为保守。因此，这一简单的解析模型从理论上解释了为何纯粹在解码器侧进行特定站点的优化（无论采用闭式形式还是学习方法）在实践中都能带来如此显著的反馈节省。

C. 上下文感知语义解码

现在，我们考察接收端的适配问题：当接收端除了关注信号保真度外，还关注下游语义效用（分类准确率）时，应如何调整解码器。

1) 语义决策规则：虽然第二节已证明，条件均值是最小化均方误差的最优选择，但要最小化语义误分类，则需要采用不同的决策策略。

令 $Y \in \{1, \ldots, M\}$ 为信源 X的语义标签。解码器观测到索引 $I = Q_d(X)$，并寻求决策 $\hat{y} = g(I)$，以最小化错误概率 $P(Y \neq \hat{Y})$。决策理论中的一个标准结论是，最优规则为最大后验概率（MAP）估计器。在

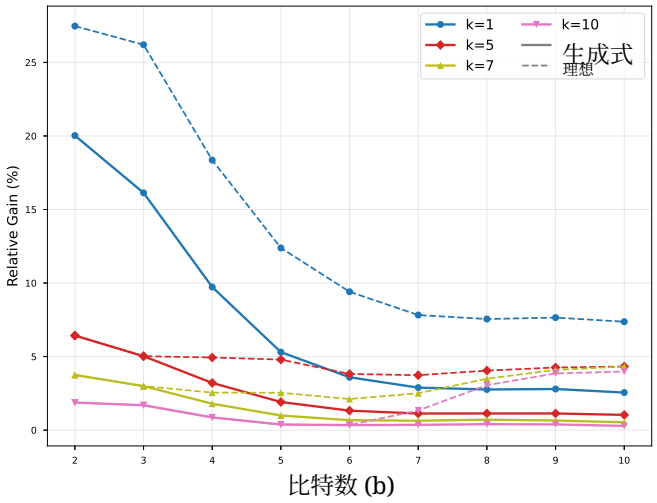

图4。不同目标子集大小 k 下，相对重构增益与码本大小（比特）的关系。严重失配（ k 较小）且比特率较低时，适应增益最高。

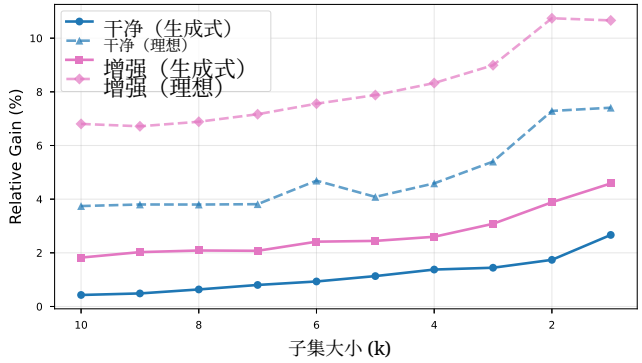

图5。8比特下干净场景与增强场景的重构增益比较。生成式自适应的相对优势在域偏移（增强）下依然稳健。

生成式解压缩中，这要求解码器根据真实分布 P_t计算后验分布：

$$\hat{y}^*(i) = \arg\max_{y\in\{1,\dots,M\}} \mathbb{P}_t(Y = y \mid X \in \mathcal{R}_i). \qquad (47)$$

这凸显了一种根本性的操作转变：保真度导向解码（均方误差）计算的是加权平均值（命题1），而任务导向解码则根据量化区间内的主导类别概率质量进行硬选择 $\mathcal{R}_i$。因此，语义场景中的生成式自适应并不严格意味着调整质心，而是要在每个固定区间内重新估计类先验 $\mathbb{P}_t(Y|I)$ 这实际上可以通过在目标分布上重新训练分类器头来实现。

2）实验设置：为在高维场景下验证生成式解压原理，我们在 Fashion-MNIST 数据集 [32] 上采用 VQ-VAE 架构。我们模拟语义鲁棒性场景：编码器在全局混合分布上训练，而解码器面对的是发射机未知的偏移分布。在本节中，我们区分理想基准与生成式解码器两种情况：前者中，编码器（包括码本）和解码器均在目标分布上端到端重新训练，而后者生成式解码器情况下，编码器被冻结（预先在源分布上训练），仅调整解码器参数。

架构与失配：固定的 VQ-VAE 将图像X 映射为潜向量 $z \in \mathbb{R}^{64}$，并量化到 $N = 2^b$ 个码本条目（$b \in \{2,\dots,10\}$）。失配通过以下方式引入：

- 类别限制： 测试时仅出现 10 个类别中的 k 个（ $k \in \{1,\dots,10\}$），以模拟专业化的部署环境。
- 域偏移： 图像会经过编码器训练期间未见过的几何和光度增强（噪声、旋转、缩放）。

自适应模式：我们比较三种策略：
1)未调整解码器： 在完整混合数据上训练的设计模型，直接在目标分布上进行评估。 2) 生成式解码器：编码器使用冻结的设计码本 $\mathcal{C}_{\text{enc}}$ 生成索引（固定划分$\{\mathcal{R}_i\}$）。解码器使用单独且解耦的码本 $\mathcal{C}_{\text{dec}}$。在自适应过程中，$\mathcal{C}_{\text{dec}}$ 及后续头部网络通过在目标分布上进行反向传播微调，以最小化均方误差或分类误差，同时 $\mathcal{C}_{\text{enc}}$ 保持冻结。 3) 理想基准：对整个处理流程（编码器和解码器）进行联合重新优化，从而提供性能上界。

指标：对于子集大小 k 和位深 b，我们测量均方误差失真 D 和分类准确率 A。所提出适应方法的相对有效性通过其恢复的理想增益占比进行量化：

$$\text{Semantic Recovery} = \frac{A_{\text{gen}} - A_{\text{fix}}}{A_{\text{ideal}} - A_{\text{fix}}}. \qquad (48)$$

3) 实验结果：

a) 重建保真度： 我们首先分析生成式解码器的信号保真度能力。图3展示了三种情形下的原始均方误差失真。未调整解码器存在系统偏差，且该偏差不会随码率增加而消失，这与理论上关于失配条件下存在失真下限的预测一致。生成式解码器降低了这一失真下限，在不修改编码器的情况下弥合了与理想基准之间的大部分差距。

图4详细说明了码本容量与适应增益之间的关系。相对增益随着位深 b的增加而单调下降。这一趋势表明，随着 b减小、失真因而增加，生成式解码器有更大的空间重新利用受限类别分布未使用的码字，从而有效提高活跃类别的分辨率。理想基准在更高码率下表现更好，因为在编码器侧调整码字能够带来额外增益。

图5展示了该方法在域偏移下的鲁棒性。虽然增强图像自然会产生更高的绝对失真，但相对恢复与干净场景保持一致。这证实了贝叶斯校正原理（第二节）对复杂的非高斯偏移具有鲁棒性；即使当信源受到噪声

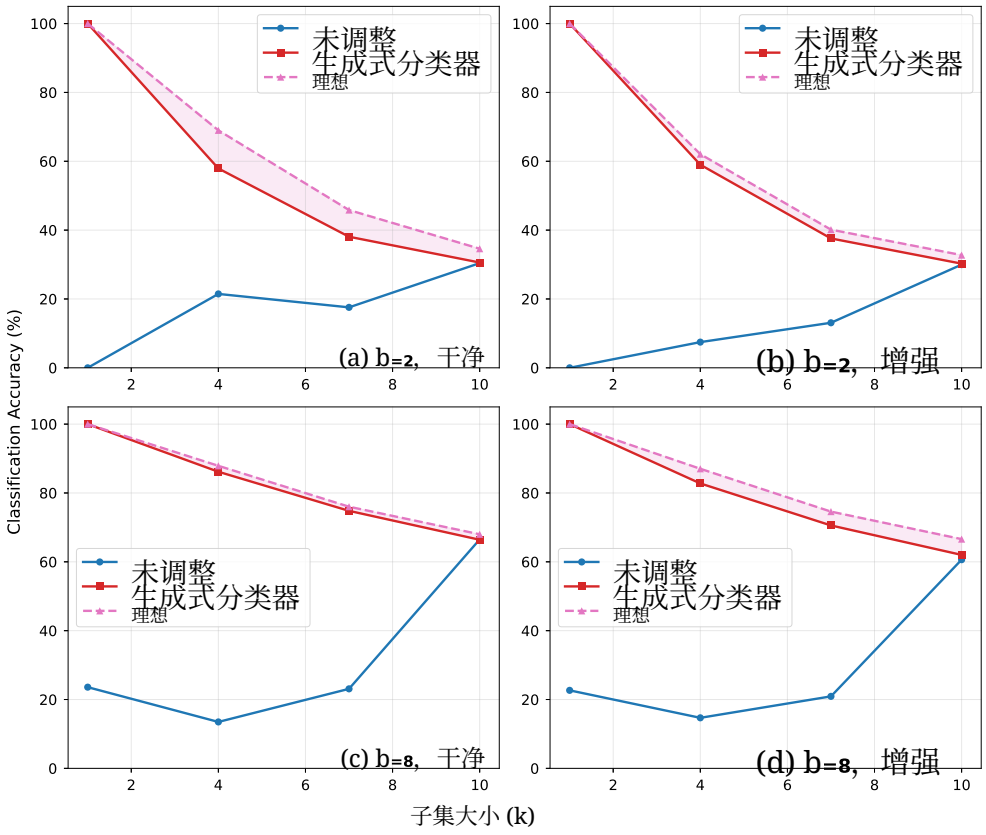

图 6. 不同位深下的分类准确率与子集大小 k 的关系。对于较小的 k，未调整基线的准确率会骤降，而生成式分类器则能显著提高准确率。

或旋转扰动时，解码端自适应仍能有效调整量化规则。

b) 语义推断：我们现在考察解码器最大化下游效用函数（准确率）而非保真度时的性能。

图6给出了绝对分类准确率随子集大小 k变化的关系。随着子集大小减小，未调整解码器（蓝色曲线）出现明显的性能崩溃。这是预期之中的，因为固定分类器在均衡10类混合的先验假设下运行，从而导致目标 k子集中不存在的类别出现假阳性。生成式分类器（红色曲线）通过在每个量化区间内有效地重新估计类先验 $P(Y|I)$，缓解了这一问题。尽管与理想基准（品红色曲线）相比仍存在差距，生成式自适应恢复了可用准确率中的关键部分。

为量化这种恢复，图7绘制了式(48)中定义的语义恢复指标。语义恢复对位深和子集大小均表现出强烈依赖。在低码率（$b = 2$）下，恢复效果明显较差。这代表一种结构性瓶颈：由冻结编码器$\mathcal{C}_{\text{enc}}$ 确定的划分过于粗糙，无法分离各类别，将来自不同标签的样本融合到同一分箱中。在此情形下，任何解码器侧的重新加权都无法解开各类别。然而，随着位深增加（$b \geq 4$），编码器的划分变得足够精细，使得解码器更新后的先验能够有效区分活跃类别，从而实现较高的恢复率。

不出所料，在 $k = 10$处恢复率几乎为0%，因为在各类别服从均匀分布的情况下，生成式解码器没有余地改进未调整的分类器。

VI. 结论

本工作为生成式解压缩建立了理论基础，这一范式将自适应负担从资源受限或标准受限的发射端转移到智能接收端。我们证明，当编码器固定或失配时——这是标准化蜂窝协议中的常见约束——最优解码策略是一种贝叶斯校正，使重构与真实源分布相一致。

我们的分析为未来通信系统揭示了三项关键见解。首先，我们证明，分布失配会形成失真下限，而只有改进解码端先验才能降低这一下限。其次，我们严格量化了固定编码器在失配条件下标准模块化信源信道分离架构的低效性，并推导出一种稳健的软解码规则。该规则将经典量化与现代联合信源信道编码（JSCC）联系起来，同时在通常假设下与 Shannon 的分离定理保持一致。第三，该框架在语义分类和 CSI反馈压缩等高维任务中的有效性表明，生成式解压缩是实现上下文感知

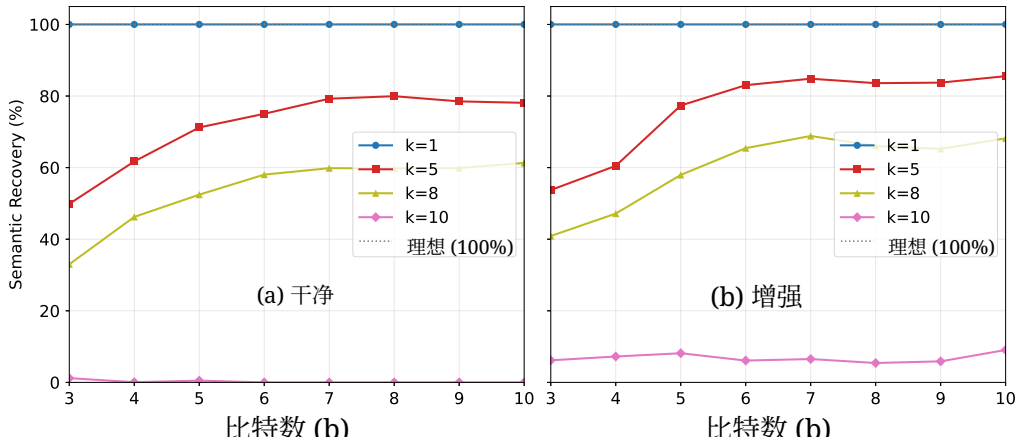

图 7. 语义恢复（%）与码本大小（b）的关系。由于冻结编码器的结构性瓶颈，低码率（$b = 2$）下的恢复效果较差，但随着码本容量增加，恢复效果显著提升。

通信的关键推动因素。随着 6G系统迈向 AI 原生接口，该框架为在接收端部署强大的生成模型以补偿粗粒度的标准化传输提供了分析依据。

参考文献

[1] R. M. Gray 和 D. L. Neuhoff，《量化》，*IEEE*信息论汇刊，第44卷第6期，页码2325–2383，1998年。[2] R. M. Gray 和 T. Linder，《熵约束矢量量化中的失配》，*IEEE*信息论汇刊，第49卷第5期，页码1204–1217，2003年。[3] A. Wyner 和 J. Ziv，《解码器端带边信息的信源编码率失真函数》，*IEEE*信息论汇刊，第22卷第1期，页码1–10，1976年。[4] S. Guler 和 A. Yener，《用于信源编码的语义信息及其在多媒体处理中的应用》，*IEEE*多媒体汇刊，第16卷第3期，页码777–792，2014年。[5] H. Li、G. Li、L. Lin、H. Yu 和 Y. Yu，《上下文感知语义修复》，*IEEE*控制论汇刊，第49卷第12期，页码4398–4411，2019年。[6] H. Luo、S. R. Khosravirad 和 A. Alkhateeb，《用于MIMO预编码设计的压缩CSI生成式解码》，*arXiv*预印本 *arXiv:2511.07783*，2025年。[7] A. Lapidoth，《失配在率失真理论中的作用》，*IEEE*信息论汇刊，第43卷第1期，页码38–47，1997年。[8] D. Slepian 和 J. Wolf，《相关信息源的无噪声编码》，*IEEE*信息论汇刊，第19卷第4期，页码471–480，1973年。[9] Z. Xiong、A. D. Liveris 和 S. Cheng，《传感器网络的分布式信源编码》，*IEEE*信号处理杂志，第21卷第5期，页码80–94，2004年。[10] S. S. Pradhan 和 K. Ramchandran，《使用综合的分布式信源编码（DISCUS）：设计与构造》，*IEEE*信息论汇刊，第49卷第3期，页码626–643，2003年。[11] Y. Blau 和 T. Michaeli，《感知-失真权衡》，载于*IEEE*计算机视觉与模式识别会议论文集，2018年，页码6228–6237。[12] M. Tschannen、E. Agustsson 和 M. Lucic，《用于保分布有损压缩的深度生成模型》，载于神经信息处理系统进展，第31卷，2018年。[13] E. Agustsson、M. Tschannen、F. Mentzer、R. Timofte 和 L. Van Gool，《用于极限学习图像压缩的生成对抗网络》，载于*IEEE/CVF*国际计算机视觉会议论文集，2019年，页码221–231。[14] F. Mentzer、G. D. Toderici、M. Tschannen 和 E. Agustsson，《高保真生成式图像压缩》，载于神经信息处理系统进展，第33卷，2020年，页码11 913–11 924。[15] A. Van Den Oord、O. Vinyals 和 K. Kavukcuoglu，《神经离散表示学习》，载于神经信息处理系统进展，第30卷，2017年。

[16] J. Ball’e、V. Laparra 和 E. P. Simoncelli，《端到端优化图像压缩》，载于国际学习表征会议，2017年。[17] D. Gunduz、Z. Qin、I. E. Aguerri、H. S. Dhillon、Z. Yang、A. Yener、K. K. Wong 和 C.-B. Chae，《超越比特传输：上下文、语义与面向任务的通信》，*IEEE*通信选定领域期刊，第41卷第1期，页码5–41，2022年。[18] Z. Qin、X. Tao、J. Lu、W. Tong 和 G. Y. Li，《语义通信：原理与挑战》，*arXiv*预印本 *arXiv:2201.01389*，2021年。[19] E. C. Strinati 和 S. Barbarossa，《6G网络：超越Shannon，迈向语义通信与面向目标的通信》，计算机网络，第190卷，文章编号107930，2021年。[20] C.-K. Wen、W.-T. Shih 和 S. Jin，《面向大规模MIMO CSI反馈的深度学习》，*IEEE*无线通信快报，第7卷第5期，页码748–751，2018年。[21] L. Zhou、V. Y. Tan 和 M. Motani，《任意信源与加性信道下失配联合信源信道编码的色散》，*IEEE*信息论汇刊，第65卷，页码2259，2019年第4期，页码2234。[22] E. Bourtsoulatze、D. B. Kurka 和 D. G”und”uz，《用于无线图像传输的深度联合信源信道编码》，*IEEE*认知通信与网络汇刊，第5卷，页码579，2019年第3期，页码567。[23] S. Na 和 D. Neuhoff，《矢量量化器的Bennett积分》，*IEEE*信息论汇刊，第41卷第4期，页码886–900，1995年。[24] O. R¨ais¨a 和 A. Honkela，《多个合成数据集上集成的偏差-方差分解》，2025年。 [在线]。可访问：https://arxiv.org/abs/ 2402 03985 。[25] P. Panter 和 W. Dite，《非均匀电平间隔脉冲计数调制中的量化失真》，《*IRE*会议录》，第39卷第1期，页码44–48，1951年。[26] S. R. Khosravirad，《无速率联合信源信道编码与6G语义通信系统设计蓝图》，2025年。 [ ] 2502 06095 在线。可访问：https://arxiv.org/abs/。[27] I. Csiszar 和 J. Korner，《离散无记忆系统的信息论：编码定理》。学术出版社，1981年。[28] D. G¨und¨uz、M. A. Wigger、T.-Y. Tung、P. Zhang 和 Y. Xiao，《联合信源信道编码：基础与实用设计的最新进展》，《*IEEE*会议录》，页码1–32，2024年。[29] A. Kurtenbach 和 P. Wintz，《面向噪声信道的量化》，*IEEE*通信技术汇刊，第17卷第2期，页码291–302，1969年。[30] N. Jindal，《有限速率反馈下的MIMO广播信道》，*IEEE*信息论汇刊，第52卷，页码5060，2006年第11期，页码5045。[31] E. Bj¨ornson、J. Hoydis、L. Sanguinetti等，《大规模MIMO网络：频谱效率、能源效率与硬件效率》，《信号处理基础与趋势》，第11卷第3–4期，页码154–655，2017年。[32] H. Xiao、K. Rasul 和 R. Vollgraf，《Fashion-MNIST：用于机器学习算法基准测试的新型图像数据集》，2017年。