# 一种结合基于人工神经网络和传统解映射的混合方法，实现通信中的高效现场可编程门阵列实现

Jonas Ney[†]、Bilal Hammoud[*]、Norbert Wehn[‡] 凯撒斯劳滕
工业大学 德国凯撒斯劳滕 电子邮箱：[†]ney@eit.uni-kl.de、[*]
hammoud@eit.uni-kl.de、[‡]wehn@eit.uni-kl.de

arXiv:2304.05042v1 [eess.SP] 11 Apr 2023

**摘要—在通信系统中，自编码器（AE）是指通过人工神经网络（ANN）替代发射机和接收机的部分组件，使系统能够在信道模型上进行端到端训练的概念。这种方法旨在提升通信性能，尤其适用于信道条件不断变化的场景，但代价是训练和推理过程具有较高的计算复杂度。已有研究表明，现场可编程门阵列（FPGA）是实现高能效人工神经网络的合适平台。然而，人工神经网络的运算数量庞大且模型规模较大，限制了其在资源受限设备上的性能，这对于低延迟、高吞吐量的通信系统至关重要。为应对这一挑战，我们提出了一种基于人工神经网络的现场可编程门阵列高效解映射新方法，将自编码器的适应性与传统解映射算法的效率相结合。在适应信道条件后，提取由人工神经网络隐式学习到的信道特征，从而能够在推理过程中使用经过优化的传统解映射算法。我们通过提供现场可编程门阵列实现结果，并将通信性能与传统系统进行比较，验证了该方法的硬件效率。我们的工作为基于人工神经网络的通信算法在现场可编程门阵列上的实际应用打开了大门。**

## I. 引言

为了满足嵌入式设备日益增长的通信需求，并推动自动驾驶等领域的进一步发展，高效的通信系统硬件实现至关重要。在5G和6G时代，嵌入式设备上的通信算法设计面临三项重大挑战：**低延迟** 对于对延迟敏感的通信至关重要，**低功耗** 是满足嵌入式应用能量约束的必备条件，而由于通信实体之间存在变化的信道条件，近实时的**适应性** 也具有高度重要性。

在通信领域，自编码器（AE）方法应对着设计端到端系统的挑战，使其能够近乎实时地适应信道波动。该方法首次在 [1], 提出：用人工神经网络（ANNs）替代经典发射机和接收机的部分组件，从而联合训练所有组件，优化全局系统的性能。自编码器用于通信符号映射和解映射的方法，首次在 [2] 通过使用空中测量在真实世界信道中得到验证，使用软件定义无线电（SDR）。为了解决信道模型的不足，尤其是适应变化的信道条件，在针对加性高斯白噪声（AWGN）等抽象信道模型进行初始端到端训练后，又通过真实信道对解映射器人工神经网络进行了再训练。

尽管人工神经网络能够提升传统通信系统的性能 [3, 4],但在高效硬件实现方面仍面临挑战。对于基于人工神经网络的方法，其较高的内存需求和计算需求会导致硬件实现产生较大的延迟、功耗和能耗。这些挑战对于资源受限通信系统而言尤为显著。此外，传统算法已针对高效实现进行了高度优化。由于这些原因，迄今为止，人工神经网络解决方案尚未对基带信号处理的实际实现产生根本性影响。本文聚焦于解映射，这是基带信号处理中的一项重要任务，并且会在很大程度上影响整体通信性能。

现场可编程门阵列（FPGA）已被证明非常适合从功耗、延迟和可重构性方面实现高效的人工神经网络 [5, 6]。然而，大多数基于FPGA的人工神经网络架构仅针对推理，通过在软件中训练网络，并将训练后的参数导出至优化推理引擎 [7, 8]。对于在我们的场景中应用自编码器而言，这种方法并不充分，因为可训练系统的主要优势在于能够实时适应不断变化的信道条件。因此，能够实现实时训练的人工神经网络FPGA实现对于我们的应用而言是必不可少的。

由于现场可编程门阵列的可重构性，在执行再训练之前，可以充分利用可用资源进行推理，随后利用所有资源进行高效训练。尽管人工神经网络训练算法的复杂度高于推理，但推理的执行频率要高得多。因此，优化推理对于降低系统的功耗、能量和延迟具有重要意义。

因此，本文的创新之处在于

- 一种新的混合方法，通过提取训练后的解映射器所学习到的信道特征，将人工神经网络的适应性与传统算法相结合。

©2022 IEEE。允许个人使用本材料。任何其他用途均须获得 IEEE 许可，无论当前或未来媒体形式，包括为广告或促销目的重印/再版本材料、创作新的合集作品、转售或重新分发至服务器或列表，或在其他作品中重复使用本作品的任何受版权保护部分。2022 IEEE 国际并行与分布式处理研讨会论文集（IPDPSW）978-1-6654-9747-3/22/$31.00 ©2022 IEEE DOI：10.1109/IPDPSW55747.2022.00024

通过这种方式，我们将低复杂度传统解映射器的效率与适应不断变化的信道条件的优势相结合。

- 我们展示了在 Xilinx ZCU3EG 现场可编程门阵列上的新方法的详细实现结果，并将其与传统解映射算法进行比较。

## II. 混合方法

我们提出了一种用于高效实现基于自编码器的解映射器的新方法，该方法既能利用自编码器概念来适应不断变化的信道条件，又能在推理时使用高度复杂的传统解映射算法，如图1所示。

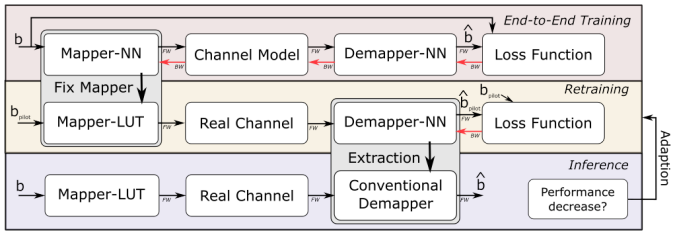


图1. 混合方法

我们的方法可分为多个步骤：
1) **端到端训练：** 在软件中针对抽象信道模型，对包含发射机和接收机人工神经网络的端到端系统进行训练。
2) **再训练：** 固定映射器星座，并在硬件中对解映射器人工神经网络进行再训练，以补偿信道模型的缺陷，从而使系统适应真实信道条件。3) **推理：** 通过从解映射器人工神经网络中提取判决区域，实现使用传统解映射算法进行高效推理。如果信道条件变化导致通信性能下降，则触发再训练。

下面将更详细地介绍我们方法的各个步骤。

### *A.* 端到端训练

在传统通信系统中，发射机和接收机由多个功能模块组成（例如编码器、解码器、映射器和解映射器），用于克服信道引入的信号失真和噪声。在端到端训练步骤中，我们着重用人工神经网络取代传统的映射和解映射模块，以学习一种能够抵抗抽象信道模型中噪声的复杂符号表示。映射器是一个功能模块，它接收由 $m$ 个比特组成的输入向量，并将其映射为复数星座符号，随后通过信道传输。映射器的目标是在给定系统设置下，找到能够传递最多信息的符号星座。随后，解映射器接收受信道噪声影响的相应复数符号，其目标是尽可能可靠地提供以 $m$ 输出概率的形式估计最初传输的比特。对于我们的方法，映射器和解映射器的人工神经网络拓扑结构被限制为不含记忆的层（例如循环神经网络（RNN）），因为解映射器的输入空间必须是二维的。除此之外，网络结构可以任意选择。在端到端训练 步骤中，这些网络经过特定数量的训练轮次，通过最小化二元交叉熵损失来提高逐比特互信息（MI）。如 [4]所示，采用这种方法，人工神经网络能够学习非均匀星座；与针对底层信道模型的传统正交振幅调制（QAM）星座相比，这些星座能够提高逐比特互信息。

### *B.* 再训练

由于用于端到端训练的抽象信道模型无法完美模拟真实信道的所有特性，而真实信道还可能随时间变化，因此有必要在运行时适应不断变化的信道条件。因此，我们提出了一种用于通信系统解映射器的可训练人工神经网络的现场可编程门阵列实现。为降低带宽并消除反馈信道的通信开销，我们在端到端训练后固定发射机人工神经网络的星座。这样，接收机便可适应信道的波动，如 [2] 所示，其中接收机的微调使空中传输性能提升了最高达 1dB。因此，我们的方法基于一种用于解映射的可训练人工神经网络的低功耗现场可编程门阵列实现，该实现可针对不断变化的信道条件重新训练。我们使用Vivado高级综合（高级综合）和Vivado Design Suite2019.2，将人工神经网络的前向路径和反向路径实现为流水线架构。为降低延迟并提高吞吐量，我们为每个人工神经网络层实例化独立的硬件模块。这些层基于XilinxFINN库 [7] ，并支持灵活调整并行度（并行度），从而能够在延迟与功耗之间进行权衡。

### 推理

在完成再训练并且网络已适应信道后，我们按如下方式从人工神经网络中提取所学习的解映射特性：首先，我们对解映射器人工神经网络的二维输入空间进行采样，以获取每个复数输入样本（人工神经网络输入）对应的学习符号（人工神经网络输出）。由此，我们得到每个符号的判决区域（判决区域），如后文图3所述。由于该判决区域图可以解释为Voronoi图，因此我们可以为对应于符号 $s_i$ 的每个Voronoi单元找到一个质心$c_i$ ，使得 $c_i$ 表示所有满足其到 $c_i$ 的距离小于其到所有其他质心的距离的点 $X$ 。质心根据每个 Voronoi 单元的顶点计算得出，并可供传统解映射算法用于高效推理。需要注意的是，这些质心不一定会复现映射器的星座图，但会隐式包含人工神经网络所学习的信息，以补偿信道

损伤，例如在图3中，它们包括信道的相移。

为检测信道条件的变化，可以定期评估系统性能：定期发送导频符号；如果误码率（BER）达到阈值，则触发解映射器再训练；也可以如 [9] 所示，使用外层纠错码（ECC）。ECC 所纠正的比特翻转数量可作为通信系统的性能指标，从而检测环境变化并启动再训练。

## III. 结果

### *A.* 设置

为评估我们方法的效率，我们针对基于 16-QAM 调制的通信系统开展了一项案例研究。我们在 PyTorch 中通过加性高斯白噪声信道，针对不同信噪比（SNR）联合训练映射器和解映射器人工神经网络，以获得每种信噪比下的优化星座。映射器由一个包含 16 个输入和两个输出的可训练嵌入层，以及一个平均功耗归一化层组成，用于将每个输入符号映射为一个复数星座点。解映射器人工神经网络有两个输入，由三个各含 16 个神经元的全连接层组成，后接一个整流线性单元（ReLU）层和一个最终的 sigmoid 层，用于接收四个比特各自的输出概率。完成初始的 Python 端到端训练后，我们在搭载 Xilinx ZU3EG 现场可编程门阵列的 Avnet Ultra96-V2 上将解映射器实现为可训练人工神经网络。此外，我们还在该开发板上实现了从训练后的解映射器提取质心的算法，以及一种传统软解映射算法。为降低解映射的复杂度，我们实现了 [10] 中提出的次优软解映射算法，该算法以简化计算取代指数和对数函数，如下所示：

$$llr(b_k|s_r) = \frac{1}{2\sigma^2}\cdot\left\{\min_i(s_r - c_{i,k=0})^2 - \min_i(s_r - c_{i,k=1})^2\right\}$$

其中，$b_k$ 是第 $k$个比特，$s_r$ 是接收符号，而$c_{i,\ k=0}$和$c_{i,\ k=1}$ 分别表示标签为 $i \in \{1:M\}$ 的区域的质心，其第 $k$个比特分别为“0”和“1”。

为验证系统的适应性并展示环境条件的变化，我们选择一种相对于端到端训练所用加性高斯白噪声信道引入固定相位偏移的信道。

### *B.* 通信性能比较

为评估我们方法的通信性能，我们在加性高斯白噪声信道上以端到端方式训练自编码器，使其适用于不同的信噪比，从而针对每种信噪比得到不同的星座图。随后，将解映射器人工神经网络在不同信噪比下学习到的参数映射到我们的硬件架构中。接着，我们应用该算法，从训练好的人工神经网络中提取学习得到的质心，并基于这些质心进行次优软解映射。图2展示了该方法的误码率，并将其与传统软解映射算法的误码率以及使用自编码器进行推理时的误码率进行了比较。

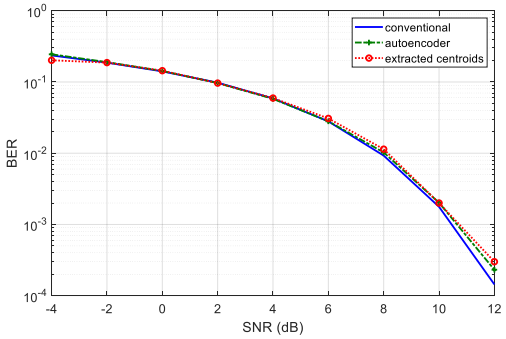


图2. 不同解映射算法的误码率（BER）。

可以看出，在信噪比不超过10 dB时，自编码器和提取的质心的性能与传统解映射器相当；在12 dB时，性能仅略有下降。这些结果验证了我们的提取算法在通信性能方面的可行性。我们将在第III-D节考察该方法相应的硬件效率。

### *C.* 适应性示意

为展示我们的方法对信道波动的适应性，我们使用相移为0的加性高斯白噪声信道进行端到端训练。随后，将人工神经网络映射到现场可编程门阵列，并在板上针对相移为 $\frac{\pi}{4}$ 的信道对解映射器人工神经网络进行再训练。之后，提取再训练后的解映射器的质心，并将其应用于次优软解映射算法。图3展示了信噪比分别为-2和8 dB时，再训练前后解映射器的判决区域及相应质心。

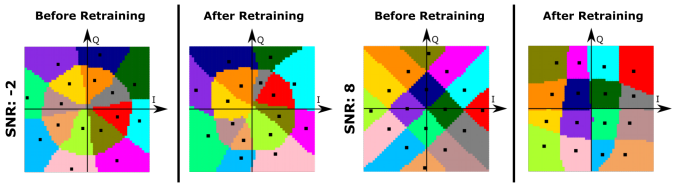


图3. 判决区域和质心示意图

可以看出，对于两种信噪比，再训练后判决区域均旋转了$\frac{\pi}{4}$。这使自编码器推理以及应用于提取的质心的传统算法的通信性能得到提升，如表1所示。

表1 自编码器及应用于提取的质心的传统算法的相位偏移适应

| SNR | 基线 | 再训练前 | | 再训练后 | |
|---|---|---|---|---|---|
| | | AE | 中心提取 | AE | 质心提取 |
| | BER | BER | BER | BER | BER |
| -2 | 0.19 | 0.318 | 0.319 | 0.199 | 0.2005 |
| 8 | 0.0103 | 0.316 | 0.323 | 0.0127 | 0.0143 |

基线误码率给出了下界，对应于不存在任何相位偏移的信道。上界则由

表2 基于自编码器的推理与传统软解映射的比较

| | 延迟 [s] | 吞吐量 [符号/秒] | BRAM | DSP | FF | LUT | 功耗 [W] | 能量 [J/符号] |
|---|---|---|---|---|---|---|---|---|
| 采用学习得到的质心的软解映射器 | $5.33{\cdot}10^{-8}$ | $7.50{\cdot}10^{7}$ | 0 | 1 | 1042 | 1107 | $5.5{\cdot}10^{-2}$ | $7.33{\cdot}10^{-10}$ |
| 自编码器推理 | $8.10{\cdot}10^{-8}$ | $1.23{\cdot}10^{7}$ | 18.5 | 352 | 10895 | 11343 | $4.53{\cdot}10^{-1}$ | $3.67{\cdot}10^{-8}$ |
| 自编码器训练 | $2.67{\cdot}10^{-7}$ | $3.75{\cdot}10^{6}$ | 89 | 343 | 19013 | 19793 | $5.47{\cdot}10^{-1}$ | $1.46{\cdot}10^{-7}$ |

再训练前的误码率给出，对应于未针对信道波动进行任何自适应的传统算法。可以观察到，再训练后的误码率几乎达到基线误码率，因此相移几乎得到了完全补偿。对于自编码器推理和基于提取的质心的解映射，情况均是如此；这表明，就所评估的信道波动而言，与自编码器推理相比，使用提取的质心并不会对通信性能造成损失。

### *D.* 现场可编程门阵列结果

解映射器人工神经网络的推理旨在最大限度地利用 Xilinx ZCU3EG 上的资源，但受到可用数字信号处理器数量的限制。为进行公平比较，我们实现了一种软解映射算法，其延迟和吞吐量与人工神经网络推理处于同一数量级。此外，如前文各节所示，两种设计实现了相同的通信性能。表 2 从硬件性能（延迟、吞吐量）、资源利用率和效率（功耗、能量）方面对两种实现进行了比较。与自编码器相比，软解映射器在资源利用率、功耗和能效方面均具有显著优势。它的查找表用量降低了 10×，数字信号处理器用量减少了 352×，从而降低了 10× 功耗并提高了 50× 能效。通过实例化多个软解映射器模块，即可并行执行解映射，使吞吐量接近 吉比特每秒 量级；由于资源数量有限，自编码器推理无法实现这一点。这些结果验证了我们提出的新方法：在保持自编码器通信性能的同时，大幅提升硬件效率，从而在资源消耗降低与固定资源下吞吐量提高之间实现取舍。

在表 2 中，我们还展示了人工神经网络训练模块的现场可编程门阵列实现结果。可以看出，由于复杂度增加，其资源利用率远高于推理。由于推理的执行频率高得多，这会导致专用集成电路（ASIC）上的训练模块出现大量空闲时间。相比之下，现场可编程门阵列可以重新配置，以执行训练或推理，从而更高效地利用资源。这正是现场可编程门阵列非常适合此应用和我们的混合方法的原因。

## IV. 结论

我们提出了一种混合方法，用于在现场可编程门阵列上高效实现基于自编码器的解映射算法。这一新颖的方法该方法基于可训练人工神经网络现场可编程门阵列架构，可在运行时用于针对信道波动进行再训练。随后，将学习到的信息以沃罗诺伊质心的形式提取出来，从而能够应用经过优化的传统解映射算法进行推理。该方法在保留传统算法效率的同时，实现了可适应自编码器概念的应用。我们通过展示相应的硬件实现结果，评估了该方法的通信性能和硬件特性。因此，我们对提取的质心应用了次优软解映射，并验证了通过对人工神经网络进行再训练，可以补偿 $\frac{\pi}{4}$ 的相位偏移。这表明，我们的混合方法将自编码器在通信性能方面的优势与高度优化的传统解映射算法的效率相结合。


## 资助

本研究部分由德国联邦教育与研究部（德国联邦教育与研究部）资助，资助项目编号为 16KIS1185（FunKI）和 16KISK004（Open6GHuB）。


## 参考文献


[1] T. O'Shea 和 J. Hoydis，“物理层深度学习简介”，*IEEE*认知通信与网络汇刊，第3卷，第4期，第563–575页，2017年。[2] S. D¨orner 等，“基于深度学习的空口通信”，*IEEE*信号处理精选主题期刊，第12卷，第1期，第132–143页，2018年。[3] B. Karanov 等，“光纤通信的端到端深度学习”，光波技术期刊，第36卷，第20期，第4843–4855页，2018年10月。 [在线]。可访问：http://dx.doi.org/10.1109/JLT.2018.2865109[4] S. Cammerer 等，“可训练通信系统：概念与原型”，*IEEE*通信汇刊，第68卷，第9期，第5489–5503页，2020年。[5] N. B. Gaikwad 等，“用于实时人体活动分类的多层感知机高效现场可编程门阵列实现”，*IEEE Access*，第7卷，第26696–26706页，2019年。[6] S. I. Venieris 和 C.-S. Bouganis，“面向延迟的基于现场可编程门阵列的卷积神经网络设计”，载于*2017*年第*27*届现场可编程逻辑与应用国际会议（*FPL*），2017年，第1–8页。[7] Y. Umuroglu 等，“FINN：用于快速、可扩展二值化神经网络推理的框架”，载于*2017*年*ACM/SIGDA*现场可编程门阵列国际研讨会论文集，系列：FPGA '17。美国纽约州纽约：计算机协会，2017年，第

65 74 [ ] 10 1145 3020078 3021744 –

[8]

页。在线。可访问：https://doi.org/。/。X. Zhang 等，“DNNBuilder：用于构建面向现场可编程门阵列的高性能深度神经网络硬件加速器的自动化工具”，载于*2018*年*IEEE/ACM*计算机辅助设计国际会议（*ICCAD*），2018年，第1–8页。[9] S. Schibisch 等，“通过纠错码实现基于深度学习的通信在线标签恢复”，2018年。[10] P. Robertson 等，“在对数域中运行的最优与次优最大后验概率解码算法比较”，载于*IEEE*国际通信会议论文集 *ICC '95*，第2卷，1995年，第1009–1013页，第2卷。