研究与开发

融合Gumbel-Softmax与CNN的毫米波大规模MIMO混合预编码研究

刘庆利[1,2]，张兆庆[1,2]
（1. 大连大学信息工程学院，辽宁 大连 116622；
2. 大连大学通信与网络重点实验室，辽宁 大连 116622）

摘　要：在毫米波大规模多输入多输出（multiple-input multiple-output，MIMO）系统中，自适应全连接结构存在二值约束、恒模约束以及信道信息利用不充分问题，导致频谱效率和能量效率性能受限。为此，提出一种融合Gumbel-Softmax与卷积神经网络（convolutional neural network，CNN）的混合预编码方法。该方法设计了两个卷积神经网络子网——发送端开关网络（transmit switching network，TsNet）和发送端相移网络（transmit phase-shift network，TpsNet），分别用于优化开关预编码矩阵和相移预编码矩阵。在TsNet中，创新性地引入Gumbel-Softmax方法，将离散二值约束嵌入网络；TpsNet则通过相位层将输出限制在移相器有效相位区间，并借助C2层满足恒模约束。TsNet和TpsNet以并联方式构建预编码联合网络（precoding coordinated network，PCNet），通过残差网络提取毫米波信道特征。两子网并行训练，共享残差网络参数以增强特征一致性，使生成的预编码矩阵接近最优。仿真结果表明，PCNet相较于其他对比算法，频谱效率和能量效率均有显著提升。

关键词：混合预编码；自适应全连接结构；卷积神经网络；残差网络；Gumbel-Softmax
中图分类号：TN929.5
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2026019

Research on hybrid precoding for millimeter-wave massive MIMO by integrating Gumbel-Softmax and CNN

Liu Qingli[1,2], Zhang Zhaoqing[1,2]
1. School of Information Engineering, Dalian University, Dalian 116622, China
2. Key Laboratory of Communication and Network, Dalian University, Dalian 116622, China

Abstract: In millimeter-wave massive multiple-input multiple-output (MIMO) systems, adaptive fully connected architectures suffered from binary constraints, constant modulus constraints, and insufficient utilization of channel information, resulting in limited spectral and energy efficiency. To address this issue, a hybrid precoding method integrat-

收稿日期：2025−09−24；修回日期：2025−10−24
通信作者：张兆庆，15242564954@163.com
基金项目：国家自然科学基金资助项目（No.61931004）
Foundation Item: The National Natural Science Foundation of China (No.61931004)

ing Gumbel-Softmax and convolutional neural network (CNN) was proposed. Two CNN subnetworks, TsNet and TpsNet, were designed to optimize the switch precoding matrix and the phase shift precoding matrix, respectively. In TsNet, the Gumbel-Softmax method was innovatively introduced to embed discrete binary constraints. TpsNet used a phase layer to constrain the output to the effective phase range of the phase shifter and utilized the C2 layer to satisfy the constant modulus constraint. TsNet and TpsNet were combined in parallel to form a joint network, PCNet, which extracted millimeter-wave channel features using a residual network. The two subnetworks were trained in parallel, sharing residual network parameters to enhance feature consistency, resulting in a near-optimal precoding matrix. Simulation results show that PCNet achieves improved spectral and energy efficiency compared to other competing algorithms. This method significantly enhances system spectral and energy efficiency.

Key words: hybrid precoding, adaptive fully-connected architecture, convolutional neural network, residual network, Gumbel-Softmax

0 引言

大规模多输入多输出（multiple-input multiple-output，MIMO）和毫米波技术是现代通信系统中的关键技术。大规模MIMO利用多天线提升系统容量与可靠性，毫米波技术则凭借其丰富的频谱资源提供大带宽，以满足高速通信需求。然而，大规模MIMO系统在采用全数字预编码时的成本与能耗剧增，毫米波信号传播受大气吸收和障碍物影响，衰减严重且传播距离受限。混合预编码技术融合模拟与数字预编码，在射频端使用移相器通过波束成形增强信号，以抵消衰减并延长传播距离，同时减少射频链路，降低成本能耗；在基带端采用数字预编码进一步优化信号处理，因此成为当前研究热点。

混合预编码可分为全连接（fully connected，FC）结构和部分连接（partially connected，PC）结构[1]。在FC结构中，文献[2]将混合预编码问题构建为矩阵分解问题并采用交替最小化框架，提出了相位提取交替最小化（phase extraction alternating minimization，PE-AltMin）算法，文献[3]将优化目标从间接的矩阵近似转换为直接最小化均方误差，提出基于流行优化和广义特征值分解的算法，并将算法扩展至宽带系统。上述算法虽能有效求解非凸约束下的多元混合预编码问题，但均存在迭代耗时问题。为此，文献[4]提出交替优化混合收发机设计方法，将频谱效率最大化问题转化为均方误差问题，用元素坐标下降算法解决恒模约束，引入离散傅里叶变换码本降低复杂度。文献[5]在此基础上，针对此前方法迭代耗时、未兼顾量化误差与非理想信道信息的问题，提出MCUR-TS两阶段算法，分别设置模拟/数字预编码，降低复杂度，且兼容误差与信道信息，性能优于奇异值分解（singular value decomposition，SVD）等方法。文献[6]聚焦选频毫米波MIMO这一特定场景，创新性地考虑每射频链或天线的功率约束，所设计的混合预编码方案在低分辨率移相器下仅产生适度性能损失。该研究虽未涉及多用户场景，却补全了选频场景下FC结构混合预编码的功率约束设计环节，与文献[4]的宽频带多用户适配、文献[5]的低复杂度与非理想信道信息兼容形成场景与功能上的互补。

近年来，深度学习技术被广泛应用于通信领域[7-10]，文献[11]使用深度神经网络（deep neural network，DNN）进行特征提取，文献[12]采用卷积神经网络（convolutional neural network，CNN）以模型驱动方式实现SVD和预编码设计。针对毫米波多用户MIMO混合预编码的开销与性能平衡问题，文献[13]提出深度强化学习驱动的预编码框架，结合奖励机制优化能效与频谱效率，有效

减少射频链和波束训练开销，使频谱效率提升39%，但存在总功耗增加45%的短板，且性能依赖环境角度信息的准确性，未覆盖宽频带信道估计需求。针对文献[13]未涉及的宽频带混合毫米波MIMO场景，文献[14]聚焦信道估计与预编码协同设计，将稀疏贝叶斯学习展开为DNN结构，通过3D卷积捕捉信道稀疏特征，借助多块扩展利用时间相关性，使归一化均方误差优于传统方法；但模型训练需要大量信道数据支撑，对非稀疏信道的适应性有限，也未考虑硬件结构约束。

在射频链和天线间部署低功率开关网络的自适应连接结构，可实现混合预编码器的功耗与频谱效率平衡。文献[15]针对固定子阵列架构无法根据信道变化自适应调整、限制系统性能提升的问题，提出自适应部分连接（adaptive partially-connected，APC）结构，并给出基于特征值的算法以降低复杂度。文献[16]针对部分连接结构性能难以逼近全数字的问题，使用条件生成对抗网络优化模拟预编码和数字预编码矩阵，频谱效率提升12%~19%，达到全数字预编码的87%，但当天线数量增加时矩阵估计误差增大，且复杂度高于贪婪算法。文献[15-16]基于PC结构实现自适应连接，与PC结构相比，FC结构具有更大的性能潜力。为降低FC结构硬件复杂度，文献[17]提出基于开关网络的自适应全连接（adaptive fully-connected，AFC）结构，引入开关网络连接移相器和天线，大幅减少移相器数量，降低功耗，同时提出Two-stage算法解决AFC结构下的混合预编码问题。文献[18]进一步优化开关网络，提出基于CNN的自适应开关模块与相位调制阵列混合预编码方案，其能耗低于传统移相器方案，且鲁棒性强。文献[19]在AFC结构下，同样使用CNN优化相移预编码矩阵和开关预编码矩阵，并整合为联合优化网络CNN-JO，以获得较优模拟预编码矩阵，但该网络使用阈值处理二值约束的机制，导致开关状态的次优选择。

AFC结构存在以下问题：硬件开关的物理特性让开关预编码矩阵元素只能是0或1的离散值，虽降低功耗却大幅缩小了可行解的范围，难以精准匹配最优编码策略，影响信号传输质量；恒模约束限制了信号调制的灵活性，制约系统性能提升；毫米波信道具有高纬度、稀疏多径的特点，其结构信息难以被充分利用。

针对以上问题，本文设计了两个CNN子网——发射端开关网络（transmit switching network，TsNet）和发送端相移网络（transmit phase-shift network，TpsNet），分别用于优化开关预编码矩阵和相移预编码矩阵。TsNet采用Gumbel-Softmax方法处理二值约束，在提高开关预编码矩阵优化精度的同时保持网络可微性，训练时输出连续松弛概率以支持反向传播，测试时通过阈值操作得到严格二值矩阵；TpsNet通过相位层将输出限定在移相器相位范围内，并在C2层实现恒模约束。在此基础上构建预编码联合网络（precoding coordinated network，PCNet），利用残差网络（residual network，ResNet）提取毫米波信道特征，将特征向量作为TsNet和TpsNet的输入。两子网基于相同的特征向量并行训练，协同优化开关与相移矩阵，从而得到较优的模拟预编码矩阵。以最优预编码矩阵作为训练标签，使联合网络生成的混合预编码矩阵尽可能接近最优，提高系统的频谱效率和能量效率并满足硬件限制。

1　系统模型

1.1　系统架构与信号模型

AFC结构毫米波大规模MIMO下行链路发射端如图1所示，毫米波MIMO系统中发射端和接收端分别配备N_t根发射天线、N_r根接收天线和N_{rf}条射频链路，两端传输N_s个独立数据流，且满足$N_s \leqslant N_{rf} \leqslant \min\{N_t, N_r\}$。AFC结构引入$N_c$个加法器（$N_c < N_t$）以减少移相器数量，每个加法器与$N_{rf}$条射频链对应的移相器相连，移相器数量

为$N_c N_{rf}$。加法器接收移相器信号并聚合，聚合后的信号通过开关控制网络连接到天线。

接收信号$\boldsymbol{x}$可表示为：

$$\boldsymbol{x}=\boldsymbol{F}_S\boldsymbol{F}_{PS}\boldsymbol{F}_{BB}\boldsymbol{s}=\boldsymbol{F}_{RF}\boldsymbol{F}_{BB}\boldsymbol{s} \tag{1}$$

其中，$\boldsymbol{F}_S\in\mathbf{C}^{N_t\times N_c}$为开关预编码矩阵，$\boldsymbol{F}_{PS}\in\mathbf{C}^{N_c\times N_{rf}}$为相移预编码矩阵，$\boldsymbol{F}_{BB}\in\mathbf{C}^{N_{rf}\times N_s}$为数字预编码矩阵，$\boldsymbol{s}\in\mathbf{C}^{N_s\times 1}$为$E\left[\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\right]=\frac{1}{N_s}\boldsymbol{I}_{N_s}$的信号矢量，$\boldsymbol{F}_{RF}$为模拟预编码矩阵（$\boldsymbol{F}_{RF}=\boldsymbol{F}_S\boldsymbol{F}_{PS}$）。其中$\boldsymbol{I}_{N_s}$为一个维度为$N_s\times N_s$的单位矩阵，用于表征$N_s$个独立且功率相等的数据流的自相关特性。

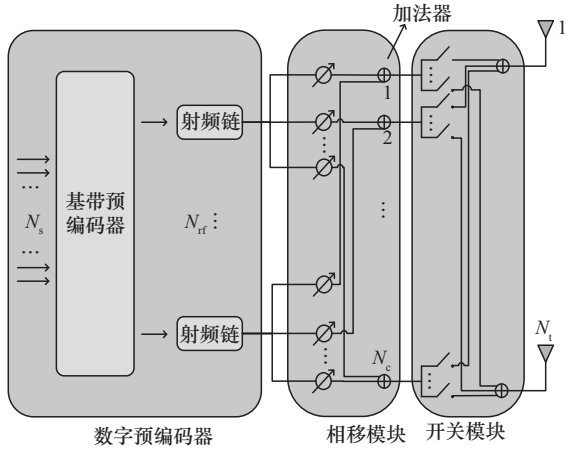

图1 AFC结构毫米波大规模MIMO下行链路发射端

模拟移相器特性决定$\boldsymbol{F}_{PS}$的元素模值相同，即$\left|\left[\boldsymbol{F}_{PS}\right]_{i,j}\right|=\frac{1}{\sqrt{N_t}}$。$\boldsymbol{F}_S$的元素取值为$\{0,1\}$，对应各开关通断状态。归一化总功率约束满足$\left\|\boldsymbol{F}_S\boldsymbol{F}_{PS}\boldsymbol{F}_{BB}\right\|_F^2=N_s$。采用窄带分块衰落信道模型，接收端为全连接结构时，合并后的接收信号$\boldsymbol{y}\in\mathbf{C}^{N_s\times 1}$可表示为：

$$\boldsymbol{y}=\sqrt{\rho}\,\boldsymbol{W}_{BB}^{\mathrm{H}}\boldsymbol{W}_{RF}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{x}+\boldsymbol{W}_{BB}^{\mathrm{H}}\boldsymbol{W}_{RF}^{\mathrm{H}}\boldsymbol{n} \tag{2}$$

其中，ρ为平均接收功率，$\boldsymbol{n}\in\mathbf{C}^{N_r\times 1}$为加性复高斯白噪声，$\boldsymbol{H}\in\mathbf{C}^{N_r\times N_t}$为信道矩阵，$\boldsymbol{W}_{BB}\in\mathbf{C}^{N_{rf}\times N_s}$为数字组合器，$\boldsymbol{W}_{RF}\in\mathbf{C}^{N_r\times N_{rf}}$为模拟组合器，$\boldsymbol{W}_{RF}$元素满足恒模约束的限制，即$\left|\left[\boldsymbol{W}_{RF}\right]_{i,j}\right|=\frac{1}{\sqrt{N_r}}$。

1.2 毫米波信道模型

下行信道基于Saleh-Valenzuela模型[20]，信道矩阵$\boldsymbol{H}$定义为N_{cl}个簇的和，每个簇包含N_{ray}条传播路径。信道矩阵$\boldsymbol{H}$表达式为：

$$\boldsymbol{H}=\sqrt{\frac{N_tN_r}{N_{cl}N_{ray}}}\sum_{i=1}^{N_{cl}}\sum_{k=1}^{N_{ray}}\alpha_{i,k}\boldsymbol{a}_r\left(\varnothing_{i,k}^{r},\theta_{i,k}^{r}\right)\boldsymbol{a}_t^{\mathrm{H}}\left(\varnothing_{i,k}^{t},\theta_{i,k}^{t}\right) \tag{3}$$

其中，$\alpha_{i,k}$为第i簇中第k条射线的复通道增益，$\varnothing_{i,k}^{t}\left(\theta_{i,k}^{t}\right)$和$\varnothing_{i,k}^{r}\left(\theta_{i,k}^{r}\right)$分别表示与第$i$簇中第$k$条射线相关的方位出发角（angle of departure，AOD）和方位到达角（angle of arrival，AOA），$\boldsymbol{a}_t\left(\varnothing_{i,k}^{t},\theta_{i,k}^{t}\right)$和$\boldsymbol{a}_r\left(\varnothing_{i,k}^{r},\theta_{i,k}^{r}\right)$分别表示发射端和接收端的阵列响应向量。

当采用yz平面上具有半波长间距的$M\times N$均匀平面阵列时，阵列响应向量为：

$$\boldsymbol{a}_{UPA}(\varnothing,\theta)=\frac{1}{\sqrt{MN}}\left[1\ \mathrm{e}^{\mathrm{j}\pi(\sin\varnothing\sin\theta+\cos\theta)}\cdots\mathrm{e}^{\mathrm{j}\pi(m\sin\varnothing\sin\theta+n\cos\theta)}\cdots\mathrm{e}^{\mathrm{j}\pi((M-1)\sin\varnothing\sin\theta+(N-1)\cos\theta)}\right]^{\mathrm{T}} \tag{4}$$

其中，$0\leqslant m\leqslant M-1$和$0\leqslant n\leqslant N-1$分别为数组元素的$y$轴和$z$轴下标。

2 问题建模

2.1 混合预编码优化目标

根据文献[21]，高斯信号经毫米波信道传输，可达到的频谱效率R为：

$$R=\mathrm{lb}\left(\left|\boldsymbol{I}_{N_s}+\frac{\rho}{N_s}\boldsymbol{R}_n^{-1}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{W}\right|\right) \tag{5}$$

其中，$\boldsymbol{R}_n=\sigma_n^2\boldsymbol{W}_{BB}^{\mathrm{H}}\boldsymbol{W}_{RF}^{\mathrm{H}}\boldsymbol{W}_{RF}\boldsymbol{W}_{BB}$表示噪声协方差矩阵，$\boldsymbol{W}=\boldsymbol{W}_{RF}\boldsymbol{W}_{BB}$表示接收端的混合组合器，$\boldsymbol{F}=\boldsymbol{F}_S\boldsymbol{F}_{PS}\boldsymbol{F}_{BB}$表示发送端的混合预编码器。

根据文献[22]，混合预编码器的优化问题可近似为如下子问题：

$$\begin{cases} \min\limits_{\boldsymbol{F}_{\mathrm{S}},\boldsymbol{F}_{\mathrm{PS}},\boldsymbol{F}_{\mathrm{BB}}} \left\| \boldsymbol{F}_{\mathrm{OPT}}-\boldsymbol{F}_{\mathrm{S}}\boldsymbol{F}_{\mathrm{PS}}\boldsymbol{F}_{\mathrm{BB}} \right\|_F^2 \\ \quad \mathrm{C1}:\left[\boldsymbol{F}_{\mathrm{S}}\right]_{m,n}\in\{0,1\},\ {\scriptstyle m=1,2,\cdots,N_{\mathrm{t}} \atop \scriptstyle n=1,2,\cdots,N_{\mathrm{c}}} \\ \text{s.t.}\quad \mathrm{C2}:\left|\left[\boldsymbol{F}_{\mathrm{PS}}\right]_{i,j}\right|=\dfrac{1}{\sqrt{N_{\mathrm{t}}}},\ {\scriptstyle i=1,2,\cdots,N_{\mathrm{c}} \atop \scriptstyle j=1,2,\cdots,N_{\mathrm{rf}}} \\ \quad \mathrm{C3}:\left\| \boldsymbol{F}_{\mathrm{S}}\boldsymbol{F}_{\mathrm{PS}}\boldsymbol{F}_{\mathrm{BB}} \right\|_F^2=N_{\mathrm{s}} \end{cases} \tag{6}$$

其中，$\boldsymbol{F}_{\mathrm{OPT}}$为全数字最优预编码矩阵，由信道矩阵$\boldsymbol{H}$的奇异值分解得到，C1为开关预编码矩阵中的二值约束，C2为相移矩阵的恒模约束，C3为功率限制。

2.2　硬件功耗分析

根据文献[23]，系统总功耗P表示为：

$$P=P_{\mathrm{t}}+P_{\mathrm{circuit}} \tag{7}$$

其中，P_{t}表示信号发射过程的能量消耗，P_{circuit}表示硬件架构固有功耗，即射频链、移相器和电路开关的功耗。

根据不同的混合预编码结构，P_{circuit}可以具体表示为：

$$P_{\mathrm{circuit}}=\begin{cases} P_{\mathrm{rf}}N_{\mathrm{rf}}+P_{\mathrm{ps}}N_{\mathrm{rf}}N_{\mathrm{t}}\text{，FC 结构} \\ P_{\mathrm{rf}}N_{\mathrm{rf}}+P_{\mathrm{ps}}N_{\mathrm{t}}\text{，PC 结构} \\ P_{\mathrm{rf}}N_{\mathrm{rf}}+P_{\mathrm{ps}}N_{\mathrm{t}}+P_{\mathrm{s}}N_{\mathrm{t}}\text{，APC 结构} \\ P_{\mathrm{RF}}N_{\mathrm{rf}}+P_{\mathrm{ps}}N_{\mathrm{c}}N_{\mathrm{rf}}+P_{\mathrm{s}}N_{\mathrm{t}}N_{\mathrm{c}}\text{，AFC 结构} \end{cases} \tag{8}$$

其中，P_{rf}、P_{ps}和P_{s}分别表示单条射频链、单个移相器和单个开关的功耗。

能量效率E表示为：

$$E=\frac{R}{P} \tag{9}$$

其中，R为频谱效率，单位为bit/(s·Hz)；P为系统总功耗，单位为W。

3　PCNet设计

毫米波大规模MIMO信道具有高维度、稀疏多径等特点，为混合预编码设计带来巨大挑战。为充分利用信道结构信息并满足AFC架构的二值与恒模约束，本节设计了联合优化网络PCNet。其核心思路在于：首先利用ResNet强大的特征提取能力，捕捉高维稀疏信道中的关键多径簇特征；进而针对不同的硬件约束，设计专用子网络并采用差异化策略。特别地，引入Gumbel-Softmax方法[24]处理开关预编码矩阵的二值约束，该方法因其独特机制非常适合毫米波场景，其优势主要体现在以下3个方面。

（1）Gumbel-Softmax能够有效适配毫米波信道的稀疏特性。毫米波信道在角度域表现出固有的稀疏性，即有效路径数量有限。Gumbel-Softmax通过可微的松弛采样，使网络能够以概率形式学习并精准地开启或关闭对应不同传播路径的开关，从而高效利用信道的稀疏结构，将能量集中在关键路径上，避免资源浪费。

（2）该方法擅长处理高维离散决策问题。开关预编码矩阵的维度通常很高，是一个典型的高维离散优化问题。传统迭代算法在此问题上复杂度高且易陷入局部最优。Gumbel-Softmax通过连续的松弛变量和温度退火策略，将这一高维离散选择问题嵌入可导的神经网络框架中，使得端到端的训练成为可能，并能够高效地搜索巨大的解空间。

（3）Gumbel-Softmax具有增强模型鲁棒性的优势。其采样过程中引入的随机性（Gumbel噪声）作为一种隐式的正则化手段，可以提升模型对训练数据中噪声和不确定性的适应能力。这对于存在信道估计误差的非理想场景尤为重要，使得所设计的预编码器对非理想的信道状态信息更具鲁棒性。

3.1　ResNet特征提取设计

ResNet凭借深层架构与残差连接机制，非常适合处理毫米波大规模MIMO信道。其核心结构包含主路径与残差路径：主路径内的卷积层、批量归一化层及ReLU激活函数可学习信道矩阵输入与输出特征间的关系映射，提取多径簇的幅度衰减与相位偏移特征；残差路径通过恒等映射或卷积实现维度匹配，确保原始信道信息直接传递至输出，避免深层网络中微弱多径分量特征的丢

失，尤其适用于毫米波信道的稀疏结构建模。

ResNet结构如图2所示。ResNet参数设计充分考虑了信道特性，卷积层使用3×3卷积核以保持感受野，捕捉相邻天线间的信道相关性，步幅设为1，填充为1，确保特征图尺寸不变，避免下采样丢失多径细节，有助于精准建模毫米波信道的稀疏多径。全连接层将高维信道矩阵降维为512维特征向量，保留信道主径方向和能量分布信息，降低计算复杂度，便于后续特征共享。ResNet包含4个卷积层和1个全连接层，可逐层提取信道空间相关性及多径簇特征，避免网络过浅导致特征提取不足或过深导致梯度消失问题。其在无线通信领域中的有效性已在文献[25-27]中得到验证。

3.2 TsNet开关预编码矩阵生成设计

训练阶段TsNet结构如图3所示，TsNet的输入为三通道实值张量$\boldsymbol{X}$，由信道矩阵实部、虚部和绝对值堆叠构成（数据转换详见第4.5节）。张量$\boldsymbol{X}$经ResNet提取特征后，生成特征向量$\boldsymbol{v}$；$\boldsymbol{v}$输入全连接层，映射至开关预编码矩阵，输出维度为$N_{\mathrm{t}}N_{\mathrm{c}}\times 1$，对应开关预编码矩阵的矢量形式。全连接层采用$\sigma_{\mathrm{Sig}}(x)=1/(1+\mathrm{e}^{-x})$非线性激活函数。为便于后续计算，全连接层的输出定义为$\boldsymbol{\Phi}$，其元素定义为$\boldsymbol{\Phi}_{1,1},\boldsymbol{\Phi}_{1,2},\cdots,\boldsymbol{\Phi}_{N_{\mathrm{c}},N_{\mathrm{t}}}$，其中$\boldsymbol{\Phi}_{i,j}\in(0,1)$。

在AFC结构中，开关预编码矩阵须满足二值约束C1，其元素取值必须为0或1。将该约束嵌入TsNet架构时，由于离散函数的不可微性，梯度下降算法难以直接优化。Gumbel-Softmax方法通过Gumbel噪声和温度参数，实现离散采样的连续松弛，将离散操作转化为可微的连续操作，既近似离散分布，又支持神经网络端到端训练，是处理离散分布采样的关键方法。因此，本文使用该方法优化开关预编码矩阵。

开关预编码矩阵$\boldsymbol{F}_{\mathrm{S}}$的元素取0-1，分别对应“开关断开”与“开关导通”状态，将每个开关的状态建模为二分类问题：“导通”状态（类别1）概率为$\boldsymbol{\Phi}_{i,j}$，“断开”状态（类别0）概率为$1-\boldsymbol{\Phi}_{i,j}$，其概率分布$\left[\boldsymbol{\Phi}_{i,j},1-\boldsymbol{\Phi}_{i,j}\right]$为Gumbel-Softmax采样提供输入。

为实现离散状态的可微分采样，为每个开关的两类状态独立生成Gumbel噪声。噪声采样自标准Gumbel分布，生成表达式为：

$$g_k=-\log\left(-\log\left(u_k\right)\right) \tag{10}$$

其中，$u_k\sim\mathrm{Uniform}(0,1)$，表示Gumbel噪声，$k=1$、2分别对应“导通”与“断开”状态。

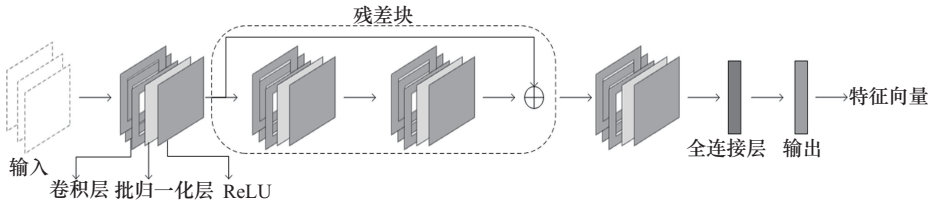

图2 ResNet结构

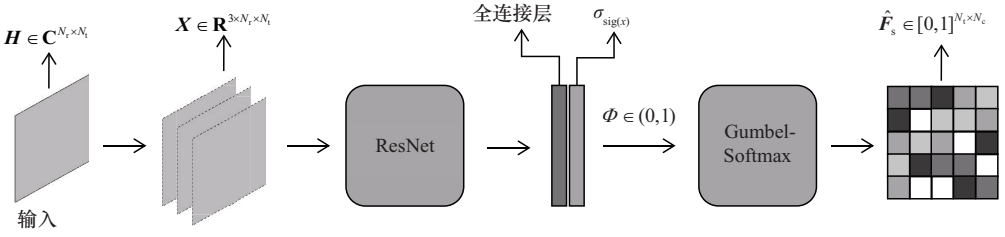

图3 训练阶段TsNet结构

将噪声注入两类状态的对数概率中，得到带噪声的对数概率分别为：

$$L_1^{(i,j)}=\log \Phi_{i,j}+g_1 \tag{11}$$

$$L_2^{(i,j)}=\log\left(1-\Phi_{i,j}\right)+g_2 \tag{12}$$

引入温度参数$\tau>0$，对带噪声的对数概率进行 Softmax 操作，得到连续松弛的概率分布，以支持神经网络反向传播：

$$\hat{y}_{(i,j)}=\mathrm{Softmax}\left(\frac{\left[L_1^{(i,j)},L_2^{(i,j)}\right]}{\tau}\right) \tag{13}$$

其中，$\hat{y}_{(i,j)}=\left[y_1^{(i,j)},y_2^{(i,j)}\right]$，$y_1^{(i,j)}$为“导通”状态的松弛概率，$y_2^{(i,j)}$为“断开”状态的松弛概率，且满足$y_1^{(i,j)}+y_2^{(i,j)}=1$。

如图 3 所示，在训练阶段，$\boldsymbol{F}_{\mathrm{S}}$的元素采用松弛概率$y_1^{(i,j)}$作为开关状态的“软输出”，参与网络反向传播。此时$y_1^{(i,j)}$为$(0,1)$中的连续值，通过温度控制逼近离散状态，确保损失函数可优化。

测试阶段 TsNet 结构如图 4 所示，在测试阶段，当温度参数τ趋近于 0 时，对$\hat{y}_{(i,j)}$取 arg max 操作，若$y_1^{(i,j)}>y_2^{(i,j)}$，则$\left[\boldsymbol{F}_{\mathrm{S}}\right]_{i,j}=1$（导通），否则$\left[\boldsymbol{F}_{\mathrm{S}}\right]_{i,j}=0$（断开），最终得到严格满足二值约束的矩阵。

3.3　TpsNet 相移预编码矩阵生成设计

TpsNet 结构如图 5 所示，TpsNet 输入为实值张量$\boldsymbol{X}$，经 ResNet 处理生成特征向量$\boldsymbol{v}$。$\boldsymbol{v}$先通过一个包含 1 024 个神经元的全连接层，再经过批归一化层和 ReLU 激活函数后进入相位层。相位层为全连接层，其激活函数采用缩放的 Sigmoid 函数$\sigma'_{\mathrm{Sig}}(x)=2\pi\sigma_{\mathrm{Sig}}(x)$，以保证输出$\beta$的范围为$(0,2\pi)$，$\beta$为相移预编码矩阵的相位值。特征最终输入 C2 层，为满足恒模约束，C2 层设计如下：

$$\boldsymbol{f}_{\mathrm{ps}}=\mathrm{vec}\{\boldsymbol{F}_{\mathrm{PS}}\}=\frac{1}{\sqrt{N_{\mathrm{t}}}}\exp(\mathrm{j}\boldsymbol{\beta}) \tag{14}$$

其中，$\boldsymbol{f}_{\mathrm{ps}}$为 C2 层的输出，表示相移预编码矩阵$\boldsymbol{F}_{\mathrm{PS}}$的矢量化形式，其元素符合式（6）中的恒模约束 C2；$\boldsymbol{\beta}=\left[\beta_1,\beta_2,\cdots,\beta_i,\cdots,\beta_{N_{\mathrm{c}}N_{\mathrm{rf}}}\right]^{\mathrm{T}}$为 C2 层的输入，$\beta_i\in(0,2\pi)$。

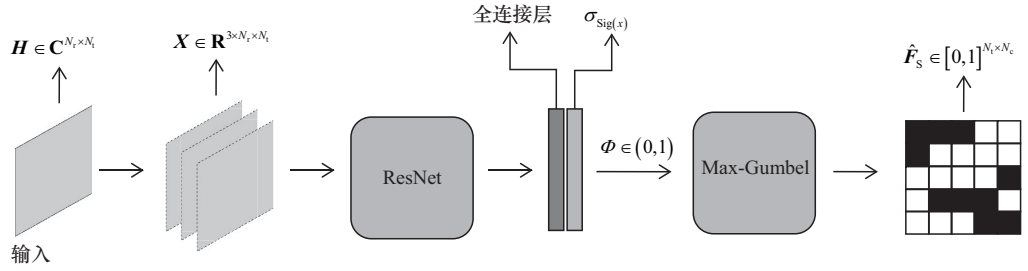

图 4　测试阶段 TsNet 结构

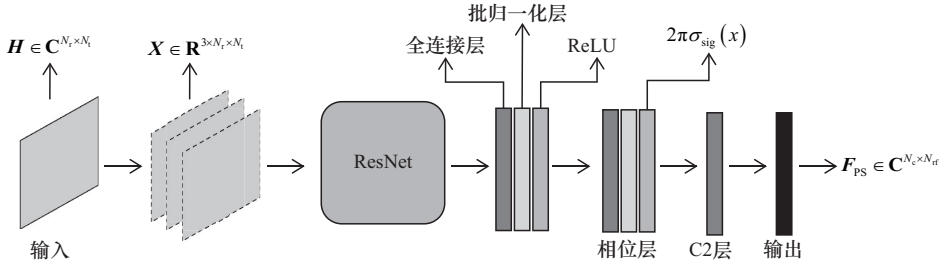

图 5　TpsNet 结构

3.4 PCNet联合优化网络设计

为实现开关预编码矩阵与相移预编码矩阵协同优化，本文将TsNet与TpsNet以并联方式整合为联合优化网络PCNet。核心设计是通过参数共享与并行训练，提升模拟预编码矩阵的逼近精度，并高效求解数字预编码矩阵。测试阶段联合网络PCNet结构如图6所示。

PCNet的输入为融合信道矩阵实部、虚部及绝对值的三通道实值张量，经ResNet模块提取特征。因两子网针对的预编码矩阵关联性强，故共享ResNet特征提取参数，确保关键特征的一致性，为协同优化提供统一基础。

ResNet输出的特征向量同步输入两子网步骤如下：TsNet通过全连接层与Gumbel-Softmax变换，生成满足二值约束的开关预编码矩阵$\boldsymbol{F}_{\mathrm{S}}$；TpsNet经全连接层、相位层及C2层处理，输出符合恒模约束的相移预编码矩阵$\boldsymbol{F}_{\mathrm{PS}}$。为构建完整混合预编码链路，两子网的输出需要被整合。PCNet在两子网输出端引入Lambda层，其输入为$\boldsymbol{F}_{\mathrm{S}}$、$\boldsymbol{F}_{\mathrm{PS}}$及由理想信道通过SVD得到的全数字最优预编码矩阵$\boldsymbol{F}_{\mathrm{OPT}}$。该层首先根据AFC结构的工作原理，将开关矩阵$\boldsymbol{F}_{\mathrm{S}}$与相移矩阵$\boldsymbol{F}_{\mathrm{PS}}$进行矩阵相乘，计算模拟预编码矩阵$\boldsymbol{F}_{\mathrm{RF}}=\boldsymbol{F}_{\mathrm{S}}\boldsymbol{F}_{\mathrm{PS}}$。这一运算精确地建模了硬件的信号流：相移矩阵$\boldsymbol{F}_{\mathrm{PS}}$的每一列代表一条射频链产生的移相信号，开关矩阵$\boldsymbol{F}_{\mathrm{S}}$的每一行则定义了某根天线与所有加法器输出之间的连接配置。矩阵乘法的过程模拟了将加法器聚合后的移相信号，通过开关网络路由到指定天线的物理机制。

在前向传播过程中，TsNet与TpsNet作为两个并行分支，其计算过程可由深度学习框架调度实现并行执行，以充分利用硬件计算资源，提升效率。然而，Lambda层的运算$\boldsymbol{F}_{\mathrm{RF}}=\boldsymbol{F}_{\mathrm{S}}\boldsymbol{F}_{\mathrm{PS}}$严格依赖于两个子网的输出张量$\boldsymbol{F}_{\mathrm{S}}$和$\boldsymbol{F}_{\mathrm{PS}}$，在计算图上构成了一个明确的数据依赖点。因此，框架会自动插入一个同步点，确保仅在$\boldsymbol{F}_{\mathrm{S}}$和$\boldsymbol{F}_{\mathrm{PS}}$均已计算完成并就绪后，才会执行Lambda层的矩阵乘法操作。这种基于数据依赖的隐式同步机制由底层计算图调度器自动管理，无须研究者手动干预，从而在保障计算正确性的同时，最大限度地维持了模型的并行计算效率。该层基于最小二乘原理与总功率约束，求解数字预编码矩阵$\boldsymbol{F}_{\mathrm{BB}}$，计算$\boldsymbol{F}_{\mathrm{BB}}$的过程如下。

首先，计算非归一化数字预编码矩阵$\bar{\boldsymbol{F}}_{\mathrm{BB}}$：

$$\bar{\boldsymbol{F}}_{\mathrm{BB}}=\left(\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{RF}}\right)^{-1}\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{OPT}} \tag{15}$$

然后，结合总功率约束进行归一化处理：

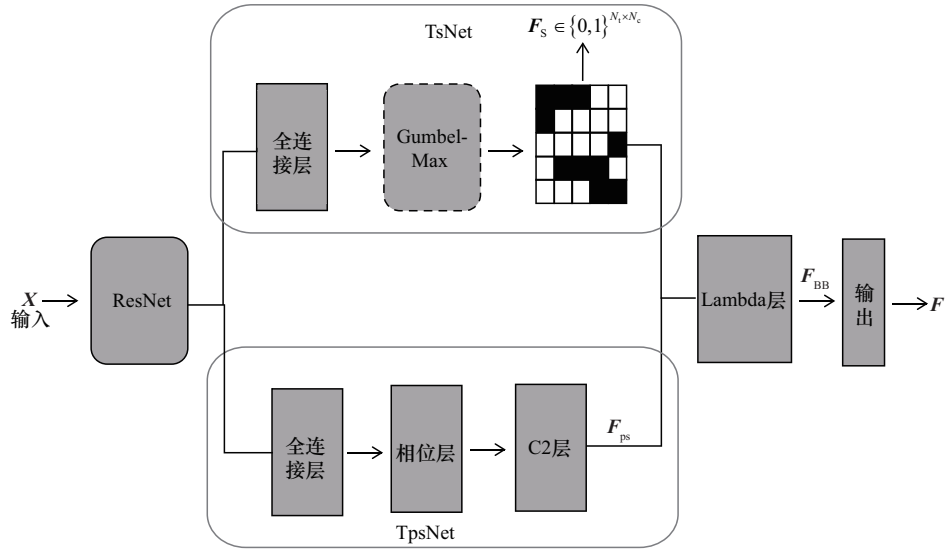

图6 测试阶段联合网络PCNet结构

$$\boldsymbol{F}_{\mathrm{BB}}=\frac{\sqrt{N_{\mathrm{s}}}}{\left\|\boldsymbol{F}_{\mathrm{RF}}\bar{\boldsymbol{F}}_{\mathrm{BB}}\right\|_F}\bar{\boldsymbol{F}}_{\mathrm{BB}} \tag{16}$$

至此，PCNet通过整合开关预编码矩阵、相移预编码矩阵及数字预编码矩阵，输出完整的混合预编码矩阵$\boldsymbol{F}=\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}$，实现从信道特征到预编码矩阵的端到端映射。

3.5　网络输入数据生成

为增强网络模型对不同信道估计精度的泛化能力，本文采用非理想信道矩阵作为网络输入。根据文献[28]，估计信道矩阵$\hat{\boldsymbol{H}}$的表达式为：

$$\hat{\boldsymbol{H}}=\eta\boldsymbol{H}+\sqrt{1-\eta^2}\,\boldsymbol{E} \tag{17}$$

其中，$\boldsymbol{H}$为由Saleh-Valenzuela模型生成的理想信道矩阵；$\eta\in[0,1.0]$为信道估计精度参数，η越接近1.0表示估计精度越高，$\eta=1.0$时$\hat{\boldsymbol{H}}$为理想信道矩阵；$\boldsymbol{E}$为误差矩阵，其元素服从独立同分布的复高斯分布。

针对每个理想信道矩阵$\boldsymbol{H}$，在$\eta\in[0.6,1.0]$时均匀选取L（$L=5$）个值（即$\eta=0.6,0.7,0.8,0.9,1.0$），生成5个不同精度的估计信道矩阵。选取该区间是因为$\eta=0.6$对应实际场景中可能出现的较低估计精度（误差占比40%），$\eta=1.0$对应理想无误差场景，该范围可覆盖毫米波通信中由噪声、干扰导致的信道估计质量变化区间。$L=5$的设置可在保证覆盖度的同时平衡训练数据量与计算复杂度，避免取值过多降低训练效率。

毫米波信道矩阵为复数矩阵，元素包含幅度和相位信息。通过提取矩阵元素的绝对值（体现幅度特性）、实部和虚部（共同体现相位特性），构建三通道实值张量，可完整保留复数信道的全部特征信息，同时适配实值神经网络的输入格式。转换关系为：

$$\begin{aligned}[\boldsymbol{X}]_{1,i,j}&=\left|\left[\hat{\boldsymbol{H}}\right]_{i,j}\right|\\ [\boldsymbol{X}]_{2,i,j}&=\Re\left\{\left[\hat{\boldsymbol{H}}\right]_{i,j}\right\}\\ [\boldsymbol{X}]_{3,i,j}&=\Im\left\{\left[\hat{\boldsymbol{H}}\right]_{i,j}\right\}\end{aligned} \tag{18}$$

其中，$\left|\left[\hat{\boldsymbol{H}}\right]_{i,j}\right|$、$\Re\left\{\left[\hat{\boldsymbol{H}}\right]_{i,j}\right\}$、$\Im\left\{\left[\hat{\boldsymbol{H}}\right]_{i,j}\right\}$分别表示信道矩阵元素的绝对值、实部和虚部。

3.6　联合网络训练策略

两阶段训练策略流程如图7所示，PCNet采用两阶段训练策略，核心是通过离线训练阶段的特征学习与在线部署阶段的快速推理，平衡预编

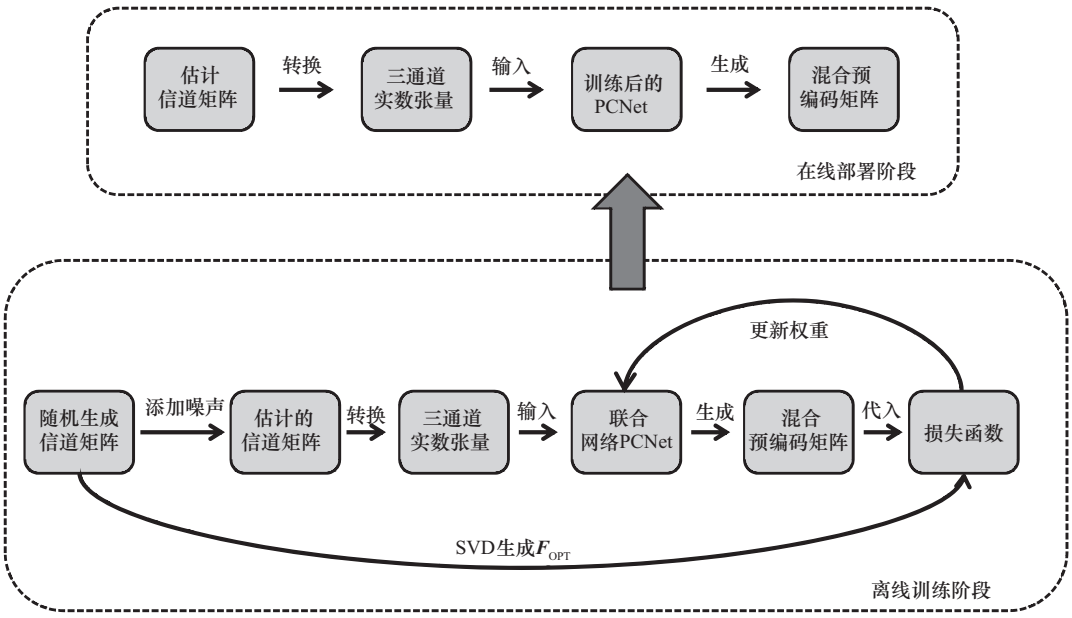

图7　两阶段训练策略流程

码性能与实时性需求。

在离线训练阶段，依据Saleh-Valenzuela信道模型随机生成理想信道矩阵$\boldsymbol{H}$，通过式(17)构造含不同估计误差的信道矩阵$\hat{\boldsymbol{H}}$。数据处理器将$\hat{\boldsymbol{H}}$转换为包含元素绝对值、实部和虚部的三通道实值张量，作为PCNet的输入特征。对理想信道矩阵$\boldsymbol{H}$进行奇异值分解得到全数字最优预编码矩阵$\boldsymbol{F}_{\mathrm{OPT}}$，作为监督学习的标签。

损失函数定义为：

$$L(\theta)=\frac{1}{N_{\mathrm{t}}N_{\mathrm{s}}}\left\|\boldsymbol{F}_{\mathrm{OPT}}-\boldsymbol{F}\right\|_2^2 \tag{19}$$

其中，θ为网络参数；$\boldsymbol{F}=\boldsymbol{F}_{\mathrm{S}}\boldsymbol{F}_{\mathrm{PS}}\boldsymbol{F}_{\mathrm{BB}}$为PCNet输出的混合预编码矩阵；$N_{\mathrm{t}}$、$N_{\mathrm{s}}$分别为发射天线数和数据流数。

训练以最小化损失函数为目标，采用Adam优化器迭代更新网络参数。在训练过程中，两子网并行前向计算，其输出在Lambda层整合并参与损失计算。梯度反向传播时，损失梯度首先反馈至Lambda层，随后分别流入TsNet和TpsNet分支以更新其特有参数。由于两子网共享来自同一ResNet的特征向量，来自两个分支的梯度在反向传播至此处时将自动汇合（相加），共同用于更新共享的ResNet特征提取参数。这一机制确保了ResNet学习到的特征能同时最优地服务于开关选择和相位调整两个子任务。尽管前向传播中两子网可并行执行以提升效率，但在反向传播更新共享的ResNet参数时，须等待来自TsNet和TpsNet两个分支的梯度均计算完毕后方可进行，该同步机制由深度学习框架的自动微分系统自动管理，保障了训练过程的正确性与稳定性。

本文设初始学习率为0.001，以平衡初始探索与收敛速度；批量大小设为128，以保证每次迭代统计的稳定性，并平衡内存占用与训练效率；总迭代次数为500次，与Gumbel-Softmax温度退火策略迭代步长匹配，确保模型在温度收敛至稳定值时同步达到最优参数。

为兼顾TsNet中开关预编码矩阵的离散约束建模与网络可微性，Gumbel-Softmax温度参数采用退火策略。初始温度$\tau=1.0$，借随机性促进状态空间探索，避免模型陷入局部最优。温度每100次迭代降低0.2，500次迭代后温度趋近于0，使松弛概率分布逼近离散二值分布，确保输出矩阵满足开关的二值约束。

此阶段引入多精度信道样本，使模型充分学习信道估计误差的统计特性。以理想预编码矩阵为优化目标，引导网络在误差存在时仍能逼近最优解，为非理想信道场景下的预编码优化构建从特征学习到误差补偿的完整机制。

在线部署阶段，将实时估计的信道矩阵$\hat{\boldsymbol{H}}$转换为三通道实值张量，输入训练完成的PCNet。网络经单次前向传播即可输出满足硬件约束的预编码矩阵，无须在线迭代或参数重训练。依托离线阶段学习的误差适应能力，可快速响应实际通信中信道信息的动态变化，保证预编码精度，满足毫米波系统对低时延的需求，有效支持非理想信道场景下的实时通信。

算法整体流程如下。

算法1 融合Gumbel-Softmax与CNN的混合预编码算法

（1）生成训练数据

for $n=1$ to N：

根据Saleh-Valenzuela信道模型生成理想信道$\boldsymbol{H}^{(n)}$

for $l=1$ to L：

选取信道精度参数$\eta_l\in[0.6,1.0]$；

生成估计信道矩阵$\hat{\boldsymbol{H}}^{(n,l)}=\eta_l\boldsymbol{H}^{(n)}+\sqrt{1-\eta_l^2}\,\boldsymbol{E}$；

将$\hat{\boldsymbol{H}}^{(n,l)}$转换为三通道实值张量$\boldsymbol{X}^{(n,l)}$；

对$\boldsymbol{H}^{(n)}$进行SVD分解，得到$\boldsymbol{F}_{\mathrm{OPT}}^{(n)}$，一个$\boldsymbol{F}_{\mathrm{OPT}}^{(n)}$对应$L$个$\boldsymbol{X}^{(n,l)}$。

（2）训练PCNet

初始化参数：学习率0.001、批量大小128、

总迭代次数500、初始温度$\tau=1.0$。

将训练数据$D=\left\{\left(\boldsymbol{X}^{(n,l)},\boldsymbol{F}_{\mathrm{OPT}}^{(n)}\right)\right\}$按7:3划分为训练集和测试集。

for $k=1$ to 500：

从训练集采样批量数据$(\boldsymbol{X},\boldsymbol{F}_{\mathrm{OPT}})$；

将$\boldsymbol{X}$输入PCNet，$\boldsymbol{F}_{\mathrm{OPT}}$为标签。

温度退火：每100次迭代降低温度，即$\tau=\max(\tau-0.2,0.01)$。

（3）部署训练好的PCNet模型

在线获取$\hat{\boldsymbol{H}}$，转换为$\boldsymbol{X}$；

将$\boldsymbol{X}$输入PCNet，前向推理输出$\boldsymbol{F}_{\mathrm{S}}$、$\boldsymbol{F}_{\mathrm{PS}}$、$\boldsymbol{F}_{\mathrm{BB}}$。

4　复杂度分析

联合网络PCNet的复杂度由3个部分构成：特征提取模块ResNet、TsNet和TpsNet。本节分别计算各部分时间复杂度，汇总得到PCNet的总体复杂度，并与其他主流算法进行对比。

4.1　ResNet的时间复杂度

ResNet包含4个卷积层和1个全连接层。卷积层的时间复杂度计算式为$O\left(\Sigma_{l\in L_{\mathrm{Conv}}}k_l^{(1)}k_l^{(2)}m_l^{(1)}m_l^{(2)}n_l^{\mathrm{in}}n_l^{\mathrm{out}}\right)$，其中$L_{\mathrm{Conv}}$为卷积层数目，$k_l^{(1)}k_l^{(2)}$为卷积核大小，$m_l^{(1)}m_l^{(2)}$为特征图尺寸，$c_l^{\mathrm{in}}$、$c_l^{\mathrm{out}}$分别为输入、输出通道数。本文ResNet中，第一层卷积输入通道为3，输出通道为64；其余3层卷积输入和输出通道均为64，卷积核尺寸为3×3，步幅为1，填充为1，特征图尺寸为$N_{\mathrm{r}}N_{\mathrm{t}}$。4个卷积层总时间复杂度近似为$O\left(4\times3^2\times N_{\mathrm{r}}N_{\mathrm{t}}\times64^2\right)$。

全连接层的时间复杂度计算式为$O\left(\Sigma_{l\in L_{\mathrm{FC}}}n_l^{\mathrm{in}}n_l^{\mathrm{out}}\right)$，其中$n_l^{\mathrm{in}}$、$n_l^{\mathrm{out}}$分别为输入、输出特征向量维度。ResNet全连接层将卷积层输出的$64\times N_{\mathrm{r}}\times N_{\mathrm{t}}$维高维特征降为512维，时间复杂度为$O(64N_{\mathrm{r}}N_{\mathrm{t}}\times512)$。

ResNet的总时间复杂度近似为$O\left(4\times3^2\times N_{\mathrm{r}}N_{\mathrm{t}}\times64^2+64\times N_{\mathrm{r}}N_{\mathrm{t}}\times512\right)$，可简化为$O(aN_{\mathrm{r}}N_{\mathrm{t}})$，其中$a$为与网络参数相关的常数。

4.2　TsNet的时间复杂度

TsNet包含1个全连接层和Gumbel-Softmax变换。全连接层输入为ResNet的512维特征向量，输出维度为$N_{\mathrm{t}}N_{\mathrm{c}}$，时间复杂度为$O(512N_{\mathrm{t}}N_{\mathrm{c}})$。Gumbel-Softmax的时间复杂度与开关预编码矩阵维度直接相关，即$O(N_{\mathrm{t}}N_{\mathrm{c}})$。

TsNet总时间复杂度为$O(N_{\mathrm{t}}N_{\mathrm{c}}+512N_{\mathrm{t}}N_{\mathrm{c}})$，可简化为$O(bN_{\mathrm{t}}N_{\mathrm{c}})$，其中$b$为与网络参数相关的常数。

4.3　TpsNet的时间复杂度

TpsNet包含全连接层、批归一化层、相位层和C2层。全连接层含1 024个神经元，输入为512维特征向量，时间复杂度为$O(512\times1\,024)$。相位层为全连接层，输入维度为1 024，输出维度为$N_{\mathrm{c}}N_{\mathrm{rf}}$，时间复杂度为$O(1\,024N_{\mathrm{c}}N_{\mathrm{rf}})$。C2层用于实现恒模约束，时间复杂度由输入维度决定，即$O\left(N_{\mathrm{c}}^2N_{\mathrm{rf}}^2\right)$。

忽略批归一化层和ReLU激活函数的低阶复杂度，TpsNet总时间复杂度为$O\left(512\times1\,024+1\,024N_{\mathrm{c}}N_{\mathrm{rf}}+N_{\mathrm{c}}^2N_{\mathrm{rf}}^2\right)$，简化为$O\left(cN_c^2N_{\mathrm{rf}}^2\right)$，其中$c$为常数，与全连接层的输入、输出维度相关。

4.4　PCNet的总体复杂度

PCNet总体复杂度为3个部分复杂度的最大值，即$O\left(\max\left\{aN_{\mathrm{r}}N_{\mathrm{t}},bN_{\mathrm{t}}N_{\mathrm{c}},cN_{\mathrm{c}}^2N_{\mathrm{rf}}^2\right\}\right)$。

4.5　算法对比

PCNet与其他主流算法的时间复杂度对比见表1，a、b、c均为常数，N_{r}、N_{t}、N_{c}、N_{rf}、N_{s}、i分别表示接收端天线数、发射端天线数、加法器数量、射频链数量、数据流数量和迭代次数。

由表1可知，PCNet复杂度高于基于FC结构的PE-AltMin算法和基于AFC结构的CNN-JO算法，但显著低于基于APC结构的Eigenvalue-based

算法和基于AFC结构的Two-stage算法，后两者依赖迭代优化（如特征值分解、坐标下降），复杂度随迭代次数i线性增长。实际执行时间方面，PCNet单次推理时间略高于PE-AltMin和CNN-JO，但远低于Eigenvalue-based和Two-stage，表明PCNet在保证性能的同时兼顾了实时性。

表1　时间复杂度对比

算法名称	时间复杂度	执行时间/s
PCNet	$O\left(\max\left\{aN_rN_t, bN_tN_c, cN_c^2N_{rf}^2\right\}\right)$	0.052 0
PE-AltMin	$O\left(iN_tN_{rf}N_s\right)$	0.013 2
CNN-JO	$O\left(\max\left\{N_rN_t, N_{rf}^2N_c^2, N_t^2N_c^2\right\}\right)$	0.046 1
Eigenvalue-based	$O\left(N_t^4/N_{rf}^2\right)$	3.852 2
Two-stage	$O\left(iN_tN_c^2N_{rf}^2N_s\right)$	0.131 0

5　仿真验证与结果分析

5.1　仿真参数设置

仿真参数见表2，其中发射端天线数为128根，接收端天线16根，毫米波传播信道包含4个分散簇，每个簇5条射线。每个簇的平均方位角和平均仰角均匀分布在$(0,2\pi)$范围内，扩展角为5°，每条射线的复增益遵循复高斯分布。加法器数量N_c、射频链数量N_{rf}、信噪比SNR作为仿真变量，在后续仿真中具体分析。设置数据流数量N_s始终与射频链数量N_{rf}相等。

表2　仿真参数

参数名称	参数值
发射端天线数	128
接收端天线数	16
信道簇数	4
每簇射线数	5
方位角范围	(0,2π)
仰角范围	(0,2π)
扩展角/(°)	5
复增益分布	CN(0, 1)

仿真采用文献[29]的功耗参数，设置$P_t=1$ W，$P_{rf}=250$ mW，$P_{ps}=50$ mW，$P_s=5$ mW，不同结构的功耗计算基准保持一致。虽然不同结构单元组成有差异，如FC结构无开关、APC结构无加法器，但通过量化各自硬件单元的数量与功耗乘积，可实现能量效率（定义为频谱效率与功耗的比值）的横向可比性。对比结果体现在相同功耗参数下，结构优化对能量效率的提升效果，而非绝对功耗值比较。

5.2　仿真结果分析

本文将PCNet与5种其他算法做了对比分析，包括全数字最优算法、FC结构下的PE-AltMin算法[2]、AFC结构下的Two-stage算法[17]、CNN-JO算法[19]以及APC结构下的Eigenvalue-based算法[15]。发射端与接收端公式相似，且发射端模型较接收端增加了功率限制和二值约束，故仿真中忽略接收端处理。本文的仿真中，所有算法接收端均采用PE-AltMin算法处理。

图8展示了不同加法器数量下的频谱效率对比。由图8可以看出，基于AFC结构的3种算法的频谱效率随加法器数量增加呈单调提升趋势，其原因在于：在AFC架构中，移相器数量为N_cN_{rf}，与加法器数量呈线性正相关，更多移相器单元使系统能实现更精细的信号幅度与相位调控，从而提升频谱效率。最优全数字、PE-AltMin、Eigenvalue-based算法的频谱效率值未随加法器数量变化，因PE-AltMin和Eigenvalue-based所基于的结构未使用加法器元件，而最优全数字算法的性能仅由信道特性和全数字预编码的理论最优解决定，与加法器数量无关。

在加法器数量N_c为4~16时，PCNet算法的频谱效率始终优于CNN-JO算法，原因包括两点：一是PCNet使用ResNet提取信道特征；二是PCNet使用Gumbel-Softmax方法处理二值约束，比CNN-JO的阈值处理方法更有效。当$N_c=4$时，PCNet算法的频谱效率为28.7 bit/(s·Hz)，随着加法器数量增

加，频谱效率呈现快速增长态势；当$N_c=8$时，频谱效率开始超越Two-stage算法；N_c进一步增加时，频谱效率逐步超过FC结构下的PE-AltMin算法；至$N_c=16$时，频谱效率达到38.4 bit/(s·Hz)。除基准最优算法外，优于其他对比方法，证实了PCNet算法在频谱效率提升方面的优势。

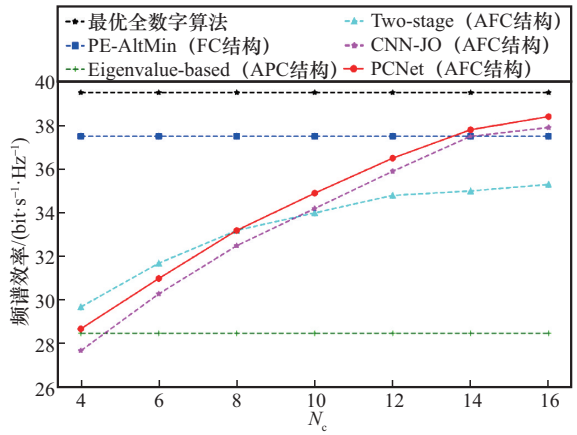

图8　不同加法器数量下的频谱效率对比（SNR=10 dB，$N_{rf}=4$）

图9展示了不同加法器数量下的能量效率对比。由图9可以看出，PCNet算法的能量效率随加法器数量N_c增加呈单调下降趋势，在AFC结构中移相器数量$N_{ps}=N_cN_{rf}$和开关数量$N_s=N_tN_c$均与N_c呈线性正相关，N_c增加时，移相器和开关的硬件规模扩大，固有功耗显著增加。此时频谱效率虽随加法器数量增加而提升（如图8所示），但功耗的增加速度超过频谱效率的提升速度，因此能量效率呈现下降趋势。

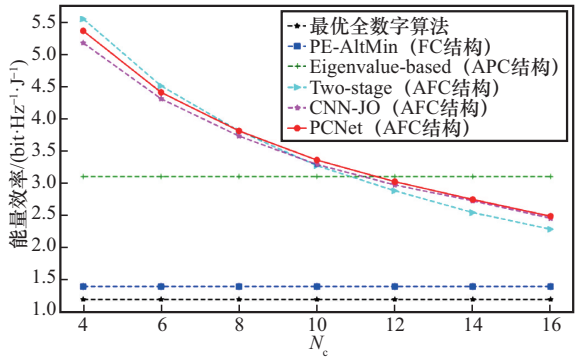

图9　不同加法器数量下的能量效率对比（SNR=10 dB，$N_{rf}=4$）

在AFC结构的3种算法中，Two-stage算法在$N_c=4$时能量效率最高（5.541 bit/(Hz·J)），但随着N_c增加，其能量效率下降最快，至$N_c=16$时降至2.286 bit/(Hz·J)，主要原因是Two-stage算法的交替迭代优化难以高效处理二值约束和恒模约束，易导致性能损失。CNN-JO和PCNet使用的深度学习方法对复杂非凸问题具有较强的拟合能力，能更精准地满足硬件约束，更接近全数字最优预编码器的性能。PCNet算法在$N_c=8$时，其能量效率与Two-stage算法持平（3.807 bit/(Hz·J)），此后逐步超越，至$N_c=16$时仍保持2.487 bit/(Hz·J)，较Two-stage算法与CNN-JO算法分别高出8.8%和1.3%，这得益于PCNet联合网络通过共享残差网络特征与并行训练机制，实现了开关预编码矩阵与相移预编码矩阵的协同优化，在硬件约束下最大化能量效率。

图10展示了不同射频链数量下的频谱效率对比。Two-stage算法与CNN-JO算法均采用加法器数量$N_c=14$的固定配置。由图10可知，所有算法的频谱效率随射频链数量增加而单调提升，原因是更多射频链提供更高的空间自由度，支持更多数据流传输，从而提高系统性能。

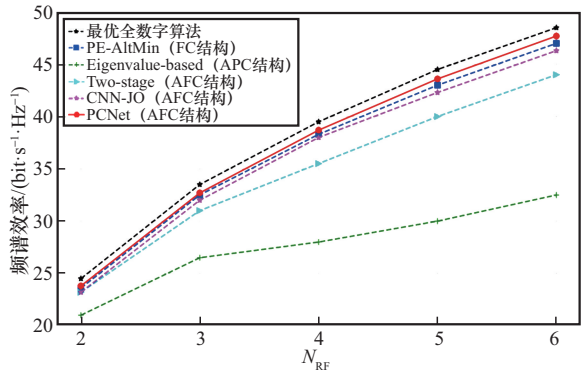

图10　不同射频链数量下的频谱效率对比（SNR=10 dB）

PCNet算法在两种加法器配置下均表现出明显优势：当$N_c=14$时，其频谱效率在全射频链数量范围内接近全数字最优预编码的性能极限，优于其他对比算法；即使在资源受限配置下（$N_c=6$），

PCNet的频谱效率也始终高于基于APC结构的Eigenvalue-based算法。

尽管PE-AltMin算法基于硬件资源更丰富的FC结构，其频谱效率仍略低于基于AFC结构的PCNet算法（$N_c=14$）。原因有两点：一是开关网络可动态配置连接方式，当N_c足够大（如$N_c=14$）时，AFC结构可逼近FC结构的连接自由度，实现与FC结构相似的预编码增益；二是PE-AltMin通过交替优化模拟和数字预编码矩阵求解非凸问题，易陷入局部最优，且未能充分利用毫米波信道结构信息，而PCNet借助两个子网络联合训练，直接输出预编码矩阵，且通过自适应优化开关状态，保留毫米波信道核心路径增益的同时规避冗余连接。

基于AFC结构的CNN-JO算法和Two-stage算法性能相近，均低于PCNet（$N_c=14$）。上述结果验证了PCNet算法在有效利用射频链路资源、最大化频谱效率方面的优越性。

图11展示了不同射频链数量下的能量效率对比。一个显著差异体现为算法对硬件架构的依赖性：基于FC结构的PE-AltMin算法的能量效率随射频链增加而明显下降，因为FC架构中每增加一条射频链需要为所有发射天线新增一组移相器，硬件功耗急剧上升超过频谱效率的收益；基于AFC和APC结构的算法，能量效率普遍随射频链增加而提升，因为这两类结构对移相器等硬件资源的利用更高效，射频链数量增加所带来的频谱效率提升可有效转化为能量效率增益。

在AFC结构算法中，Two-stage和CNN-JO算法的加法器数量均为14。PCNet在加法器数量为6时，在全射频链数量范围内能量效率最高，显著优于其他对比算法。当加法器数量为14时，其能量效率同样优于Two-stage和CNN-JO同类型AFC算法。这表明PCNet联合网络设计（共享特征提取、并行优化开关与相移矩阵、Gumbel-Softmax精准建模约束）的有效性，可在满足硬件限制的同时，最大化频谱效率对功耗的转化效率，实现优异的能量效率。

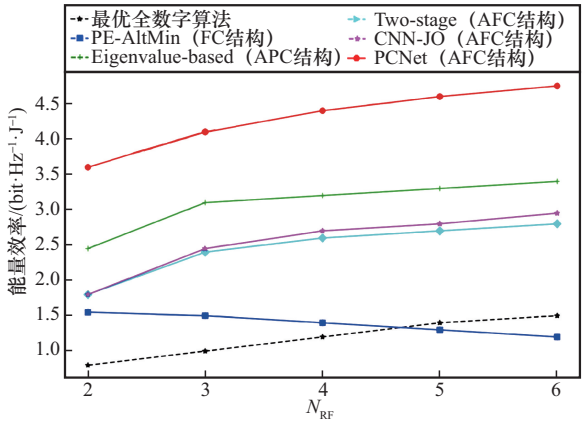

图11　不同射频链数量下的能量效率对比（SNR=10 dB）

图12展示了不同信噪比下频谱效率对比。整体来看，所有算法的频谱效率值均随信噪比提升呈单调递增趋势。信噪比升高使接收端信号质量改善，噪声干扰减弱，系统能更高效地传输数据，因此频谱效率提升。但不同算法的提升速度和最终趋近的性能水平存在明显差异。

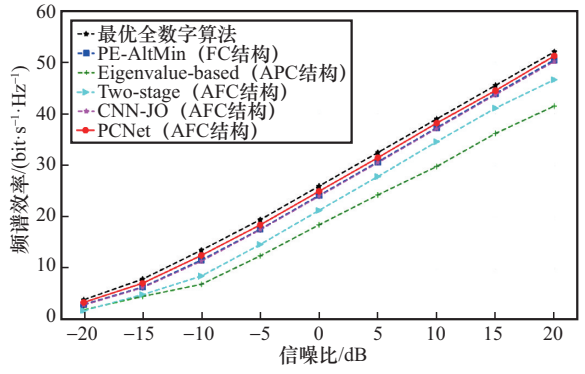

图12　不同信噪比下的频谱效率对比（$N_c=14$，$N_{rf}=4$）

具体而言，最优全数字算法作为理论性能上限，始终保持最高频谱效率，随信噪比提升增速最快，高信噪比时趋近52 bit/(s·Hz)。基于FC结构的PE-AltMin算法和AFC结构的CNN-JO算法表现相近，频谱效率差距均值仅为1.06%，但均低于PCNet算法。基于APC结构的Eigenvalue-based算法仅能利用部分天线增益，频谱效率始

终较低，随信噪比提升增速最慢。AFC 结构的 Two-stage 算法的频谱效率在各信噪比下均落后于 PCNet 算法约 20.11%。

PCNet 算法在全信噪比范围均展现出优势：低信噪比（SNR =−20 dB）时，其频谱效率已比 Two-stage 算法高约 45%；随着信噪比增加，其频谱效率值也快速提升，20 dB 时达到 51.2 bit/(s·Hz)，较全数字最优算法低 1.54%。该结果进一步证明了 PCNet 算法在不同信道条件下仍能高效逼近理论最优性能，且优于现有主流算法。

6　结束语

本文聚焦于毫米波大规模 MIMO 系统中混合预编码技术的优化，针对 AFC 结构存在的难题，提出了一种融合 Gumbel-Softmax 与 CNN 的混合预编码方法。主要研究结论如下。

（1）设计了联合优化网络 PCNet，通过 ResNet 提取毫米波信道的高维稀疏特征，捕捉多径簇的幅度衰减与相位偏移信息，为开关预编码矩阵和相移预编码矩阵的协同优化提供一致的特征基础。

（2）针对 AFC 结构的硬件约束，TsNet 引入 Gumbel-Softmax 方法，处理开关预编码矩阵的二值约束，解决了离散约束下网络训练难题；在 TpsNet 中通过相位层和 C2 层分别满足相移预编码矩阵的相位范围限制和恒模约束，确保硬件可行性。

（3）采用了两阶段训练策略，在离线学习阶段引入多精度信道样本，并结合 Gumbel-Softmax 温度退火机制，使网络在保证可训练性的同时输出严格满足约束的预编码矩阵；在线部署阶段仅须单次前向传播即可快速生成预编码矩阵，满足毫米波系统低时延需求。

仿真结果验证了所提方法的有效性。在不同加法器配置下，PCNet 的频谱效率和能量效率较对比算法平均提升 2.34%；在不同射频链配置下，频谱效率平均提升 10.98%，能量效率平均提升 27.06%，且时间复杂度低于依赖迭代优化的算法，兼顾了性能与实时性。

本文提出的基于 Gumbel-Softmax 和 CNN 的混合预编码方法，有效解决了 AFC 结构的硬件约束问题，实现了信道信息的充分利用和预编码矩阵的高效优化，为毫米波大规模 MIMO 系统的性能提升提供了可行方案。未来可进一步探索多用户场景下的扩展应用，以及更复杂信道环境中的鲁棒性优化。

参考文献：

[1] Venkateswaran V, Van Der Veen A J. Analog beamforming in MIMO communications with phase shift networks and online channel estimation[J]. IEEE Transactions on Signal Processing, 2010, 58(8): 4131-4143.

[2] Yu X H, Shen J C, Zhang J, et al. Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems[J]. IEEE Journal of Selected Topics in Signal Processing, 2016, 10(3): 485-500.

[3] Lin T, Cong J Q, Zhu Y, et al. Hybrid beamforming for millimeter wave systems using the MMSE criterion[J]. IEEE Transactions on Communications, 2019, 67(5): 3693-3708.

[4] Yuan M H, Wang H, Yin H, et al. Alternating optimization based hybrid transceiver designs for wideband millimeter-wave massive multiuser MIMO-OFDM systems[J]. IEEE Transactions on Wireless Communications, 2023, 22(12): 9201-9217.

[5] Shi B Z, Liu F L, Du R Y. A MCUR-TS hybrid precoding and combining algorithm for MIMO-OFDM systems[J]. IEEE Transactions on Wireless Communications, 2024, 23(6): 5488-5502.

[6] LóPez-Valcarce R, GonzáLez-Prelcic N. Hybrid beamforming designs for frequency-selective mmWave MIMO systems with per-RF chain or per-antenna power constraints[J]. IEEE Transactions on Wireless Communications, 2022, 21(8): 5770-5784.

[7] Zhang P, Pan L R, Laohapensaeng T, et al. Hybrid beamforming based on an unsupervised deep learning network for downlink channels with imperfect CSI[J]. IEEE Wireless Communications Letters, 2022, 11(7): 1543-1547.

[8] Liu Z Y, Yang Y W, Gao F F, et al. Deep unsupervised learning for joint antenna selection and hybrid beamforming[J]. IEEE Transactions on Communications, 2022,70(3): 1697-1710.

[9] Jin W J, Zhang J, Wen C K, et al. Model-driven deep learning for hybrid precoding in millimeter wave MU-MIMO system[J].

IEEE Transactions on Communications, 2023, 71(10): 5862-5876.

[10] Liu F L, Li X Y, Yang X H, et al. Deep learning based joint hybrid precoding and combining design for mmWave MIMO systems[J]. IEEE Systems Journal,2024,18(1): 560-567.

[11] Huang H J, Song Y W, Yang J, et al. Deep-learning-based millimeter-wave massive MIMO for hybrid precoding[J]. IEEE Transactions on Vehicular Technology, 2019, 68(3): 3027-3032.

[12] Peken T, Adiga S, Tandon R, et al. Deep learning for SVD and hybrid beamforming[J]. IEEE Transactions on Wireless Communications, 2020, 19(10): 6621-6642.

[13] Salh A, Alhartomi M A, Ali Hussain G, et al. Deep reinforcement learning-driven hybrid precoding for efficient mm-wave multi-user MIMO systems[J]. Journal of Sensor and Actuator Networks, 2025, 14(1): 20.

[14] Gao J B, Zhong C J, Li G Y, et al. Deep learning-based channel estimation for wideband hybrid MmWave massive MIMO[J]. IEEE Transactions on Communications, 2023, 71(6): 3679-3693.

[15] Park S, Alkhateeb A, Heath R W. Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems[J]. IEEE Transactions on Wireless Communications, 2017, 16(5): 2907-2920.

[16] Banerjee B, Elliott R C, Krzymień W A, et al. Hybrid beamforming for mmWave massive MIMO systems using conditional generative adversarial networks[J]. IEEE Transactions on Vehicular Technology, 2024, 73(10): 15803-15808.

[17] Liu F L, Kan X D, Bai X Y, et al. Two-stage hybrid precoding algorithm based on switch network for millimeter wave mimo systems[J]. Progress in Electromagnetics Research M, 2019, 77: 103-113.

[18] Hei Y Q, Liu C, Li W T, et al. CNN based hybrid precoding for MmWave MIMO systems with adaptive switching module and phase modulation array[J]. IEEE Transactions on Wireless Communications, 2022, 21(12): 10489-10501.

[19] Liu F L, Zhang L J, Yang X H, et al. DL-based energy-efficient hybrid precoding for mmWave massive MIMO systems[J]. IEEE Transactions on Vehicular Technology, 2023, 72(5): 6103-6112.

[20] Saleh A A M, Valenzuela R. A statistical model for indoor multipath propagation[J]. IEEE Journal on Selected Areas in Communications, 1987, 5(2): 128-137.

[21] Telatar E. Capacity of multi-antenna Gaussian channels[J]. European Transactions on Telecommunications, 1999, 10(6): 585-595.

[22] El Ayach O, Rajagopal S, Abu-Surra S, et al. Spatially sparse precoding in millimeter wave MIMO systems[J]. IEEE Transactions on Wireless Communications, 2014, 13(3): 1499-1513.

[23] Gao X Y, Dai L L, Sun Y, et al. Machine learning inspired energy-efficient hybrid precoding for mmWave massive MIMO systems[C]//Proceedings of the 2017 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2017: 1-6.

[24] Jang E, Gu S X, Poole B. Categorical reparameterization with gumbel-softmax[PP]. V5. arXiv (2017-08-05) [2025-08-24]. arXiv:arXiv. 1611.01144.

[25] Lin B, Gao F F, Zhang S, et al. Deep learning-based antenna selection and CSI extrapolation in massive MIMO systems[J]. IEEE Transactions on Wireless Communications, 2021, 20(11): 7669-7681.

[26] Cao Y S, Lv T J, Lin Z P, et al. Complex ResNet aided DoA estimation for near-field MIMO systems[J]. IEEE Transactions on Vehicular Technology, 2020, 69(10): 11139-11151.

[27] Tian Y, Pan G F, Alouini M S. Applying deep-learning-based computer vision to wireless communications: methodologies, opportunities, and challenges[J]. IEEE Open Journal of the Communications Society, 2021, 2: 132-143.

[28] Gao X Y, Dai L L, Han S F, et al. Energy-efficient hybrid analog and digital precoding for MmWave MIMO systems with large antenna arrays[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(4): 998-1009.

[29] Yu X H, Zhang J, Letaief K B. A hardware-efficient analog network structure for hybrid precoding in millimeter wave systems[J]. IEEE Journal of Selected Topics in Signal Processing, 2018, 12(2): 282-297.

[作者简介]

刘庆利（1981– ），男，博士，大连大学信息工程学院、通信与网络重点实验室教授，主要研究方向为无人机系统、网络通信等。

张兆庆（1998– ），男，大连大学信息工程学院硕士生，主要研究方向为无线通信、混合预编码。