

基于低复杂度Transformer的光纤信道快速精确建模技术



Fast and Accurate Optical Fiber Channel Modeling Technology Based on Low-Complexity Transformer

史明辉/SHI Minghui, 郑智雄/ZHENG Zhixiong,
牛泽坤/NIU Zekun, 义理林/YI Lili

(上海交通大学, 中国 上海 200240)
(Shanghai Jiao Tong University, Shanghai 200240, China)

DOI: 10.12142/ZTETJ.202506002

网络出版地址: <https://link.cnki.net/urlid/34.1228.TN.20251230.1452.002>

网络出版日期: 2025-12-30

收稿日期: 2025-10-28

摘要: 光纤信道建模对于表征光纤特性、开发先进数字信号处理算法十分重要。基于物理模型的分步傅里叶算法 (SSFM) 需要大量迭代运算, 复杂度较高, 限制了其应用前景。提出了一种基于低复杂度Transformer架构的光纤信道建模方法。该方法对传统Transformer架构做了两处关键改进——采用相对位置编码替代绝对位置编码, 并采用滑动窗口注意力机制替代全局注意力机制, 从而提升模型对光纤信道非线性效应的表征能力。结果表明, 所提方法的有效信噪比 (ESNR) 与SSFM相比仅相差0.15 dB, 计算时间较传统Transformer减少69.9%, 较SSFM降低96.9%, 从而在保持较高精度的同时, 显著降低了计算复杂度。

关键词: 光传输系统; 光纤信道建模; Transformer

Abstract: Optical fiber channel modeling is essential for characterizing fiber transmission characteristics and developing advanced digital signal processing algorithms. The split-step Fourier method (SSFM), as a physics-based numerical method, provides high accuracy but suffers from high computational complexity due to its iterative nature, which limits its practical deployment. A low-complexity Transformer-based architecture for modeling optical fiber channel waveforms is proposed. It incorporates two key modifications to the standard Transformer: replacing absolute positional encoding with relative positional encoding and substituting global attention with a sliding-window attention mechanism. These changes enhance the model's ability to capture nonlinear transmission effects. Results show that the proposed method achieves an Effective Signal-to-Noise Ratio (ESNR) within only 0.15 dB of that obtained by SSFM, while reducing the computation time by 69.9% compared to the traditional Transformer and by 96.9% compared to SSFM. This verifies that the method maintains high accuracy while significantly lowering computational complexity.

Keywords: optical transmission system; optical fiber channel modeling; Transformer

引用格式: 史明辉, 郑智雄, 牛泽坤, 等. 基于低复杂度Transformer的光纤信道快速精确建模技术 [J]. 中兴通讯技术, 2025, 31(6): 3-12. DOI: 10.12142/ZTETJ.202506002

Citation: SHI M H, ZHENG Z X, NIU Z K, et al. Fast and accurate optical fiber channel modeling technology based on low-complexity transformer [J]. ZTE technology journal, 2025, 31(6): 3-12. DOI: 10.12142/ZTETJ.202506002

光纤通信系统是现代信息传输基础设施的骨干。快速精确地建模光纤信道, 对于评估光纤特性、优化光网络配置^[1-5]、开发先进的数字信号处理 (DSP) 算法^[6-10]以及实现端到端 (E2E) 优化^[11-14]具有关键作用。光信号在光纤中的传播服从非线性薛定谔方程 (NLSE)^[15], 该方程在大多数情况下缺乏解析解。高斯噪声 (GN) 模型及其变体^[16-17]属于功率级建模方法, 将非线性效应等效为高斯噪声, 能够快速、准确地估算广义信噪比。然而, 此类GN模型无法提供详细的信号波形信息, 因而在DSP算法开发中的应用受

限, 尤其在对非线性补偿要求较高的场景中。分步傅里叶方法 (SSFM)^[18]是一种传统的波形级建模方法, 通过迭代求解NLSE。虽然SSFM精度较高, 但其计算效率低, 且复杂度与信号带宽呈四次方关系^[18], 这限制了其在下一代高速光传输系统中的应用。

近年来, 基于神经网络 (NN) 的数据驱动方法被应用于实现快速、准确的波形级光纤信道建模。这些方法通过利用带标签的数据与部分先验知识来学习信号与信道特性^[19]。基于NN的方法在保持与SSFM相近精度的同时, 能显著降

低计算复杂度。生成对抗网络 (GAN)^[20]、傅里叶神经算子 (FNO)^[21] 等架构已在单通道系统中展现出优异的建模效果。时序神经网络, 如双向长短期记忆网络 (BiLSTM)^[19] 和 Transformer^[22], 凭借其循环结构和自注意力机制, 更擅长捕捉具有符号关联性的光纤非线性特征, 表现出更高的建模精度。通过引入物理先验知识, 一种结合分布式特征解耦 (FDD) 与 BiLSTM 的方法已成功应用于多通道波分复用 (WDM) 系统。进一步地, 将 Transformer 引入 FDD 架构^[23], 提升了原有 BiLSTM 的非线性建模能力。然而, 传统 Transformer 所采用的绝对位置编码^[24] 难以有效处理符号间非线性串扰所需的相对位置关系。此外, 其自注意力机制的计算复杂度与输入符号长度呈平方关系, 在符号关联性较强的长序列高速系统中仍面临复杂度较高的挑战。

为解决上述问题, 本文提出一种结合旋转位置编码 (RoPE) 和滑动窗口注意力机制^[25] 的低复杂度 Transformer 架构 (Slide-Roformer)。该架构将绝对位置编码优化为 RoPE, 从而能更有效地感知符号间的相对位置信息, 使其更符合光纤非线性的物理特征, 以提升建模精度。同时, 本文将传统的全局自注意力机制改进为滑动窗口注意力机制, 采用局部感知方式, 使每个符号仅需计算在其非线性关联长度范围内的相关性, 从而显著降低了计算复杂度^[26]。实验结果表明, 在 5 通道、140 GBaud、发射功率为 8.5 dBm 的传输场景下, Slide-Roformer 相比传统 Transformer 在波形建模精度上提升

了 75%, 计算时间减少了 69.9%。此外, Slide-Roformer 与传统 SSFM 之间的有效信噪比 (ESNR) 误差仅为 0.15 dB, 且计算时间可降低 96.9%。这些结果表明, 本文提出的 Slide-Roformer 有望成为下一代高速光纤通信系统中信道建模的有效工具, 推动光通信技术的进一步发展。

1 原理

1.1 基于 SSFM 的光传输系统架构

在本章中, 我们主要介绍光传输系统的整体架构。该架构为 NN 的训练提供了数据来源, 同时也作为对比建模精度与计算复杂度的参考基准。一个典型的基于 SSFM 的光纤传输仿真系统如图 1 所示, 主要包括发射端、光纤信道和接收端。在发射端, 首先通过伪随机数种子生成比特序列, 并利用双偏振 16 阶正交幅度调制将其映射为符号序列。该符号序列随后以 4 倍过采样率进行上采样, 并通过滚降系数为 0.1 的根升余弦滤波器完成脉冲成形。接着, 各信道信号被调制到不同的载波频率上, 合成为全场的 WDM 信号, 并送入光纤信道进行传输。该 WDM 信号可表示为:

$$\mathbf{A}(z, t) = \sum_{k=1}^C \mathbf{A}_k(z, t) \exp(j\Delta\omega_k t) \quad (1),$$

其中, \mathbf{A} 表示在两种任意正交偏振模式下的光信号, 即 \mathbf{A}_x 和 \mathbf{A}_y 。 \mathbf{A}_k 表示第 k 个信道的光信号。变量 z 表示传输距离, t 表

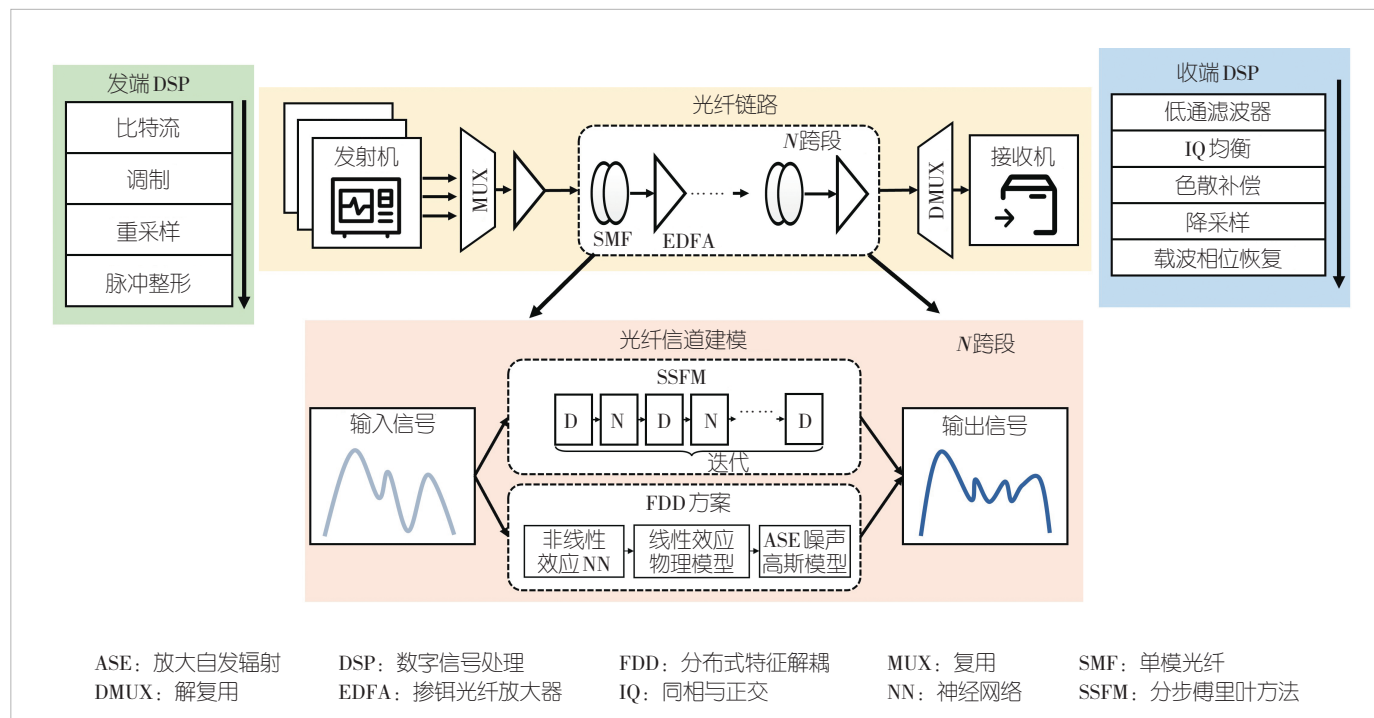


图1 光传输系统架构及信道建模方法

示时间坐标, C 为 WDM 通道总数, $\Delta\omega_k = \omega_k - \omega_0$ 为第 k 个信道中心频率与 WDM 信号中心频率之间的差值。

光信号在单模光纤中的传播由 NLSE 描述, 其形式为:

$$\frac{\partial A(z, t)}{\partial z} = (\hat{D} + \hat{N})A \quad (2),$$

其中, \hat{D} 为线性算子, 用于描述衰减、色散 (CD); \hat{N} 为非线性算子, 用于描述克尔非线性效应, 包括自相位调制、交叉相位调制以及四波混频。在处理双偏振信号时, NLSE^[15] 通常采用耦合形式, 即耦合 NLSE, 或其简化形式——马纳可夫 (Manakov) ^[27-28] 方程。本文采用 Manakov 方程对光纤信道建模, 其形式为:

$$i \frac{\partial A}{\partial z} - \frac{1}{2} \beta_2 \frac{\partial^2 A}{\partial t^2} + \frac{8}{9} \gamma |A|^2 A + \frac{\alpha}{2} i A = 0 \quad (3),$$

其中, β_2 为群速度色散参数, γ 为非线性参数, α 为损耗参数。CD 会引起符号间干扰 (ISI)。在 WDM 系统中, 受 ISI 影响的符号个数通过公式 (4) 表示:

$$N_{\text{ISI}} = \frac{\Delta T}{\Delta t} = \frac{L \beta_2 \Delta \omega}{1/S} = L \beta_2 \Delta \omega S \quad (4),$$

其中, N_{ISI} 表示受 ISI 影响的符号数, ΔT 是受 ISI 影响的时间宽度, Δt 是一个符号的持续时间, L 是传输距离, $\Delta \omega$ 是频谱宽度, S 是符号速率。非线性与线性效应在光纤传输过程中同时存在并相互耦合, 这使得非线性也具备一定的符号关联性。非线性与信号的能量相关, 会在光纤的传输过程不断减弱, 导致其符号关联性比 CD 造成的 ISI 要弱。非线性的符号关联性可以采用有效非线性长度来近似:

$$L_{\text{eff}} = \frac{1}{\alpha}, N_{\text{NL}} = L_{\text{eff}} \beta_2 \Delta \omega S \quad (5),$$

其中, L_{eff} 表示非线性有效长度, 通常小于实际传输长度。 N_{NL} 表示非线性符号间相关长度。对于当前时刻符号来说, 与之更加相近的符号对其造成的非线性串扰更强, 而更远端符号造成的影响则较弱。非线性强度与符号之间的相对位置关系有关, 而非绝对位置。

SSFM^[29-30] 是求解 NLSE 最常用的数值方法。该方法将长距离光纤传输路径划分为大量微小步长, 使得线性与非线性算子可独立处理。每一步的对称 SSFM 运算可表示为:

$$A(z + h, t) \approx \exp\left(\frac{h}{2} \hat{D}\right) \exp\left\{h \hat{N}\left[A\left(z + \frac{h}{2}, t\right)\right]\right\} \times \exp\left(\frac{h}{2} \hat{D}\right) \quad (6),$$

其中, h 为单步长度。步长配置会影响 SSFM 的精度和计算效率。在本文中步长采用适用于非线性效应主导系统的变步长方法——非线性相位旋转法^[31]。采样率设置为单通道符号率

的 4 倍。为便于计算, 线性算子在频域中计算, 而非线性算子则在时域中计算。在每段光纤末端使用掺铒光纤放大器 (EDFA) 补偿信号衰减, 其引入的放大自发辐射 (ASE) 噪声可建模为高斯噪声。本文采用的光纤信道参数如表 1 所示。

在光纤信道传输后, 首先对目标信道信号进行解复用。接收端数字信号处理 (Rx DSP) 流程如下: 先使用匹配的根升余弦 (RRC) 滤波器进行滤波, 随后执行降采样。接着进行色散补偿 (CDC) 以校正线性损伤, 继而完成载波相位恢复 (CPR) 与解调。最终, 通过计算 Q 因子或有效信噪比 (ESNR) 等指标对传输性能进行定量评估。

基于 SSFM 的仿真系统所产生的 WDM 信号波形, 为 NN 模型提供了充足的训练数据。在性能测试阶段, 发送端与接收端采用完全一致的处理流程与参数设置, 以便公平比较基于 SSFM 与基于 NN 的两类光纤信道模型之间的性能差异。

1.2 基于分布式特征解耦的光纤信道建模

基于 SSFM 的光纤信道建模方法虽然精度较高, 但其计算复杂度高, 尤其在高波特率场景下, 限制了在下一代高速光传输系统中的应用。为缓解此问题, 研究人员引入 NN 进行光纤信道建模, 并结合部分物理先验信息, 构建数据-物理混合驱动的建模架构, 以进一步提升建模精度。FDD^[32] 是一种数据-物理混合驱动方案, 它通过物理模型建模线性效应、NN 建模非线性效应, 充分发挥二者优势, 从而提高整体建模精度。此外, FDD 每次仅建模一个光纤跨段, 通过多模块级联实现长距离传输。该方案在多通道 WDM 系统中展现出较好的建模效果, 因此本文选取该方案作为基础架构, 并对其中的 NN 结构进行优化。

为使 NN 有效学习具备符号关联性的非线性特征, 需通过滑动窗口方法构建包含过去与未来时刻填充符号的输入窗口。但在高速系统中, 由于符号关联性较长, 每次推理过程中需反复计算填充符号信息, 这导致复杂度较高。为此, 我们在 NN 的训练与推理过程中引入了序列到序列 (Seq2Seq) 架构^[30]。该架构采用多符号输入、多符号输出的模式, 能够

表1 光纤信道参数

参数	数值	参数	数值
载波波长	1 550 nm	色散系数	17 ps/nm·km
通道数	5	非线性系数	1.3 /(W·km)
符号速率	140 GBaud	跨段长度	80 km
发射功率	8.5 dBm	EDFA的噪声系数	5 dB
衰减	0.2 dB/km	SSFM的最大非线性相位旋转	0.005

EDFA: 掺铒光纤放大器 SSFM: 分步傅里叶方法

单次同时预测多个符号，从而减少 NN 在重复循环预测过程中对填充符号的重复计算次数，降低推理时间。

此外，在训练过程中，参考“近端符号对当前时刻符号非线性串扰更强，远端符号更弱”的物理机制，我们采用迁移学习构建了两阶段训练模式。第一阶段训练中，先以较少的填充符号构建输入-输出数据对，以较低复杂度有效学习近端影响较大的非线性串扰；第二阶段训练中，进一步构建更长的输入-输出数据对，补充学习远端符号的非线性串扰。该两阶段训练模式有效降低了训练资源开销，并提升了模型的最终训练精度。

2 基于 Slide-Roformer 的光纤信道建模技术

BiLSTM 凭借其序列建模能力，已成为光纤信道非线性建模的有效架构之一。Transformer 基于其自注意力机制，相比 BiLSTM 具有更强的非线性拟合能力，在非线性更强的多通道、高功率 WDM 系统中表现出更优的建模性能。然而，传统 Transformer 所采用的绝对位置编码难以有效处理符号间非线性串扰所依赖的相对位置关系。此外，其自注意力机制的计算复杂度与输入符号长度呈平方关系，导致在符号关联性较长的高速系统中面临较高的复杂度挑战。在本文中，我们通过优化 Transformer 的位置编码与注意力机制，提出了 Slide-Roformer 架构，如图 2 所示，从而增强了 Transformer 在高速光通信系统中的信道建模能力。下文将分别介绍传统 Transformer 架构，以及优化后的旋转位置编码（RoPE）和滑动窗口注意力机制。

2.1 Transformer 架构

光纤信道建模通常仅采用 Transformer 的编码器架构，而不使用解码器。这是因为仅凭编码器中的自注意力机制，就能实现对时序非线性特征的有效建模。Transformer 编码器由多个堆叠的相同层组成，每层主要包括多头自注意力机制、前馈神经网络与残差连接 3 个核心模块。首先，多头自注意力机制通过并行使用多个不同的自注意力函数，从多个子空间中捕获不同类型的关系。随后，自注意力机制对输入信号进行 3 次不同的线性变换，分别生成查询（ Q ）、键（ K ）和值（ V ）向量，并通过公式（7）计算不同输入元素之间的关联程度。接着，多头注意力机制的输出经过前馈网络处理，以增强模型的表达能力。此外，每个子层之间均包含残差连接，这有助于缓解深层网络中的梯度消失问题，并提升信息流动效率。通过多层编码器的堆叠，模型能够逐步建立从局部到全局的多层次特征表示，从而实现有效的时序特征建模。

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (7)。$$

此外，Transformer 不具备类似 BiLSTM 的循环结构，无法有效获取不同元素间的位置关系。为了解决这一问题，Transformer 通常引入绝对位置编码，通过一组确定性的正弦和余弦函数生成，其不同频率对应不同的时间尺度，从而在连续空间中为每个位置的元素分配唯一的位置表示形式。对于输入序列中第 pos 个位置和隐藏维度中的第 i 个维度，位

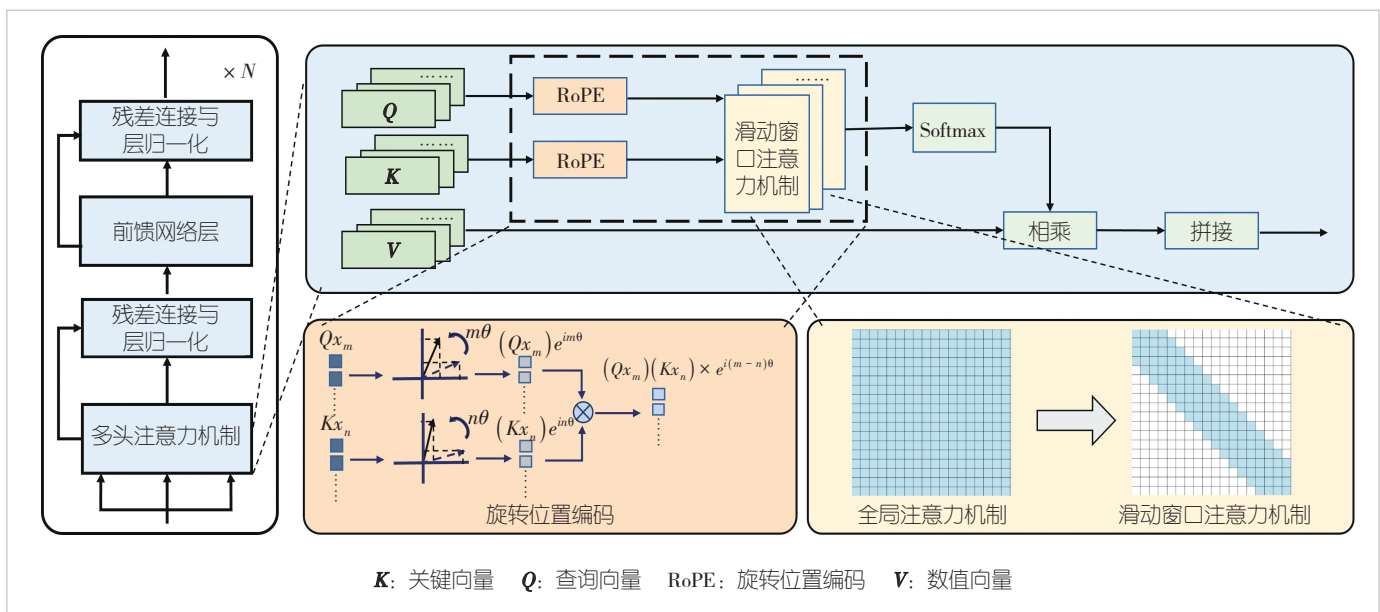


图2 Slide-Roformer 结构

置编码的计算公式为:

$$\text{PE}(\text{pos}, 2i) = \sin\left(\frac{\text{pos}}{\theta^{\frac{2i}{d}}}\right), \text{PE}(\text{pos}, 2i + 1) = \cos\left(\frac{\text{pos}}{\theta^{\frac{2i}{d}}}\right) \quad (8),$$

其中, d 表示模型隐藏层的维度, θ 代表旋转角度。通过将位置编码与输入嵌入相加, 模型在保持输入维度不变的情况下能够显式获得序列的顺序信息, 从而在注意力计算中兼顾内容与位置信息。

2.2 旋转位置编码

绝对位置编码可以有效为输入元素添加绝对位置信息, 但光纤信道的非线性效应主要取决于各符号之间的相对位置, 而非绝对位置, 这导致绝对位置编码方式效率低下, 影响建模精度。为此, 我们引入了一种相对位置编码方式—RoPE。该方法将位置信息以旋转变换的方式编码进各个位置的符号所对应的 Q 和 K 矩阵中, 并通过 Q 和 K 矩阵相乘, 以差分的形式表征各符号之间的相对位置关系, 而不再直接将绝对位置信息添加到各符号中。RoPE 的编码过程如图 2 金色部分所示, 在位置 m 和 n 时, Q 和 K 分别经过公式 (9) 和公式 (10) 表示的旋转变换。其中, W_q 代表向量经过 Q 矩阵映射, W_k 代表经过 K 矩阵映射, 并对 K 进行转置, 再与 Q 相乘取实部, 得到公式 (11)。在相乘的结果中, 通过 $m-n$ 差分地表示两个不同位置之间的相对位置关系。通过引入 RoPE, 模型可以更加高效地捕捉不同符号之间的相对非线性串扰强度, 更加适配非线性串扰的物理特征, 提高非线性建模精度。除此之外, 引入 RoPE 的模型还具备较好的外推性, 适配前文所提到的 Seq2Seq 方案。外推性是指模型在训练时和预测时输入长度不一致。这使得我们在模型的训练中采用输入长度相对较短序列, 提高训练效率; 在推理过程中, 直接外推到更长的输入和输出序列, 从而提高模型在预测长序列信号时的计算效率。

$$f_Q(\mathbf{x}_m, m) = (W_q \mathbf{x}_m) e^{im\theta} \quad (9),$$

$$f_K(\mathbf{x}_n, n) = (W_k \mathbf{x}_n) e^{in\theta} \quad (10),$$

$$g(\mathbf{x}_m, \mathbf{x}_n, m - n) = \text{Re}\left[(W_q \mathbf{x}_m)(W_k \mathbf{x}_n)^* e^{i(m-n)\theta}\right] \quad (11)。$$

2.3 滑动窗口注意力机制

在光纤信道建模中, 传统 Transformer 不仅存在位置编码上的不足, 还受限于较高的计算复杂度。这一限制主要源于自注意力机制需计算当前符号与序列中所有符号之间的相对

关联, 导致其复杂度与输入符号长度呈平方关系。此外, 光纤的非线性符号关联性可由公式 (5) 表示, 其关联长度被限制在一定范围内, 超出此范围的信息对于当前符号而言是冗余的。上述问题使得模型在处理较长符号序列时产生大量冗余计算, 增加了计算复杂度, 尤其对符号关联性更长的下一代高速光传输系统影响显著。

为克服这一限制的约束, 构建更加符合光纤信道非线性特征的注意力标准机制, 我们引入了滑动窗口注意力机制, 如图 2 黄色部分所示。在滑动窗口注意力中, 对于每个位置 i 处的符号, 仅计算其与当前位置前后 w 个符号的注意力, 从而使得整个注意力窗口的大小 $2w + 1$, 这显著降低了每个符号的计算复杂度。滑动窗口注意力机制首先使用一个滑动窗口操作在 K 、 Q 、 V 矩阵上提取局部窗口得到 K_{windowed} 、 Q_{windowed} 、 V_{windowed} 。我们以 K 的操作为例, 滑动窗口过程可以表示为:

$$K_{\text{windowed}} \in R^{n \times (2w + 1) \times d} \quad (12)。$$

每个 $K_{\text{windowed}}[i]$ 对应 K_{i-w}, \dots, K_{i+w} , 超出边界的可补零或掩码。随后使用张量广播方式, 将每个 $Q_{i-w:i+w}$ 与 $K_{i-w:i+w}$ 进行了点积, 最后得到使用滑动窗口注意力机制计算出的分数, 如公式 (13) 所示:

$$\text{score}_i = \text{Softmax}\left(\frac{1}{\sqrt{d}} Q_{i-w:i+w} \cdot K_{i-w:i+w}^T\right) V_{i-w:i+w} \in R^{2w+1} \quad (13)。$$

滑动窗口注意力机制相比全局注意力更高效, 能在保持模型性能的同时, 极大降低计算资源消耗, 并支持长序列的高效建模。

3 结果分析

3.1 训练数据集和超参数配置

为验证 Slide-Roformer 相比传统 Transformer (Vanilla-Transformer) 在光纤信道建模中精度与复杂度的优势, 我们构建了一套高速 WDM 相干光传输仿真系统。系统设置通道数为 5, 符号速率为 140 GBaud, 每通道发射功率为 8.5 dBm, 工作于高非线性区域, 以考察两模型在强非线性条件下的建模能力差异。

训练过程中, 为避免过拟合, 我们采用 10 组不同随机种子生成长度为 10 000 符号的发送信号, 并收集其经信道传输后各跨段的输入与输出波形作为训练数据集。Slide-Roformer 与传统 Transformer 的参数配置如表 2 所示, 二者主干网络参数基本一致。此外, 为保证神经网络准确学习非线性

表2 Slide-Roformer和Vanilla-Transformer的参数配置

参数	数值
输入尺寸	80
隐藏尺寸	240
注意力头数	6
前馈网络尺寸	960
层数	3

性效应, 根据公式 (5) 计算非线性效应影响长度, 将 Slide-Roformer 中滑动窗口注意力机制的长度 w 设为 300。旋转角度 θ 为工程默认值通常设置为 10 000, 并使得最大的波长远大于常见序列长度, 从而覆盖需要的频谱^[22, 25]。训练使用 Smooth L1 损失函数, 优化器选用 Adam, 初始学习率为 5×10^{-4} , 并采用余弦退火策略逐步降低学习率。在 Seq2Seq 架构的两阶段训练中, 第一阶段训练轮数设为 500, 输入窗口包含 60 个符号 (中心 20 个为有效符号, 前后各填充 20 个); 第二阶段训练轮数为 200, 输入窗口包含 580 个符号 (中心 300 个为有效符号, 前后各填充 140 个)。

在测试阶段, 我们采用与训练集不同的随机种子, 生成长度为 1×10^6 符号的发送序列, 分别通过不同模型进行传输, 以比较其建模效果与计算时间。

3.2 性能标准

基于 NN 的光纤信道波形建模性能评估包含精度评估与复杂度评估两方面。精度评估基于波形建模精度和传输性能预测精度, 并以 SSFM 的计算结果作为参考值。波形建模精度采用 (NMSE) 进行量化, 其定义为:

$$\text{NMSE} = \frac{\sum_{i=1}^{N_{\text{data}}} |\hat{y}_i - y_i|^2}{\sum_{i=1}^{N_{\text{data}}} |y_i|^2} \quad (14),$$

y_i 表示 SSFM 的输出样本, \hat{y}_i 表示 NN 的输出样本。传输性能的评估需要先对建模信号进行 Rx DSP 处理并进行解调, 计算传输的误码率 (BER) 和 Q 因子, 它们之间的关系为:

$$Q = 20 \log_{10} \left(\sqrt{2} \operatorname{erfc}^{-1}(2\text{BER}) \right) \quad (15)。$$

我们采用 SSFM 与 NN 所得的 Q 因子之间的误差, 作为传输性能预测精度的度量指标。

复杂度评估则通过比较 SSFM 与 NN 在传输相同信号时所需的计算时间来完成。NN 对于 SSFM 的加速比可表示为:

$$\text{Ratio} = \frac{\text{Time}_{\text{SSFM}} - \text{Time}_{\text{NN}}}{\text{Time}_{\text{SSFM}}} \quad (16)。$$

3.3 Slide-Roformer 的精度增益

为验证 Slide-Roformer 在建模精度上的提升, 我们将其与 Vanilla-Transformer 进行对比。为进一步评估两阶段训练带来的性能增益, Slide-Roformer 分别采用单阶段与两阶段训练进行实验。3 个模型在训练过程中的损失函数变化如图 3 所示。

结果显示, 采用两阶段训练的 Slide-Roformer 最终损失为 3.2×10^{-5} , 较单阶段训练的 5.9×10^{-5} 明显降低, 这证明了两阶段训练对提升模型精度的有效性。进一步观察发现, 在第一阶段训练结束时 (第 500 轮), Slide-Roformer 与 Vanilla-Transformer 的损失均约为 7.0×10^{-5} 。进入第二阶段后, Slide-Roformer 的损失迅速收敛至 4.0×10^{-5} , 而 Vanilla-Transformer 的损失不仅未下降, 反而上升至 2.5×10^{-2} , 凸显了 Slide-Roformer 在两阶段训练中的优势。

这一差异主要源于位置编码机制的不同: Slide-Roformer 采用相对位置编码, 即使两个阶段的输入符号长度不同, 仍能通过相对位置关系有效提取相同的非线性特征; 而 Vanilla-Transformer 采用绝对位置编码, 当输入序列长度增加时, 会引入第一阶段未见过的新位置信息, 导致模型初始阶段无法适应此类信息, 需重新训练收敛, 从而影响了最终精度。

为进一步评估 Slide-Roformer 相对于传统方法的精度优

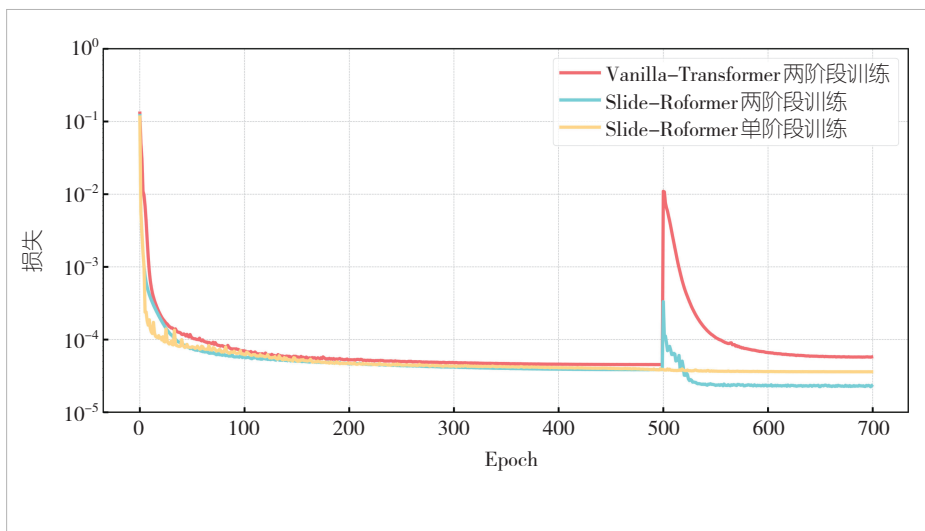


图3 Slide-Roformer与Transformer训练损失

势，我们将其与包括 Vanilla-Transformer、Seq2Seq-BiLSTM^[30] 和 FNO^[21] 在内的其他神经网络模型进行了对比。在 Seq2Seq-BiLSTM 中，BiLSTM 的隐藏层维度设为 240，层数为 3。该模型同样采用 Seq2Seq 架构，其两阶段训练的参数配置与 Slide-Roformer 保持一致：第 1 阶段输入窗口包含 60 个符号（20+20+20），第 2 阶段输入窗口包含 580 个符号（140+300+140）。FNO 模型的傅里叶模态数设为 24，宽度为 16，层数为 4，采用单阶段训练，输入窗口包含 300 个符号（140+20+140）。我们比较了各模型在 800 km 传输后的 NMSE，以评估其在长距离传输下的建模性能。如图 4 所示，Slide-Roformer 在 800 km 后的 NMSE 最低，仅为 2.0×10^{-3} ；而 Vanilla-Transformer、Seq2Seq-BiLSTM 和 FNO 的 NMSE 分别上升至 9.2×10^{-3} 、 4.2×10^{-3} 和 7.8×10^{-3} 。Slide-Roformer 相对于其他神经网络模型实现了约 50%~74% 的 NMSE 降低，这表明其具有更优的非线性效应建模能力。

其次，我们比较了深度学习模型与基准 SSFM 在传输性能预测上的差异，以评估它们在传输性能预测及 DSP 算法设计中的应用潜力。图 5 展示了基于 SSFM、Slide-Roformer 和 Vanilla-Transformer 建模的信号，在 800 km 传输并经过线性 DSP 处理后的星座图。可以看出，Slide-Roformer 与 SSFM 的星座图噪声分布高度吻合，而 Vanilla-Transformer 未能准确建模光纤信道的非线性噪声。从量化结果看，SSFM 的 ESNR 为 8.0 dB，Slide-Roformer 为 7.85 dB，误差仅为 0.15 dB；Vanilla-Transformer 的 ESNR 为 12.69 dB，与 SSFM 的误差达 4.69 dB。这些结果表明，Slide-Roformer 相比其他模型能更准确地反映光纤信道的特征。这些优势得益于 Slide-Roformer 采用的相对位置编码更符合光纤信道的非线性特性，且符号间的相对位置关系也使其能够适配更高效的两阶段训练模式，从而实现了更高的建模精度。

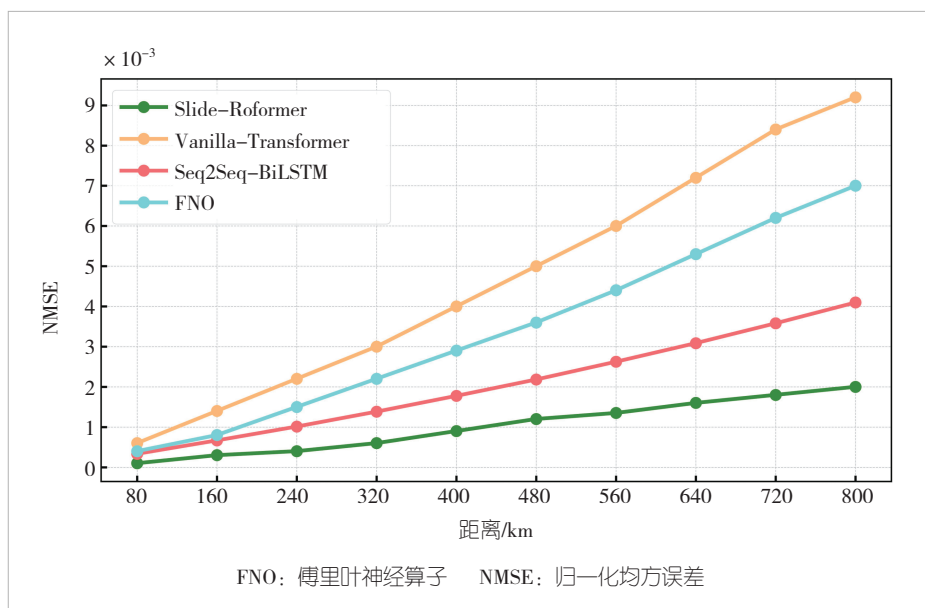


图 4 Slide-Roformer 与 Vanilla-Transformer 在 800 km 传输范围内 NMSE 的对比

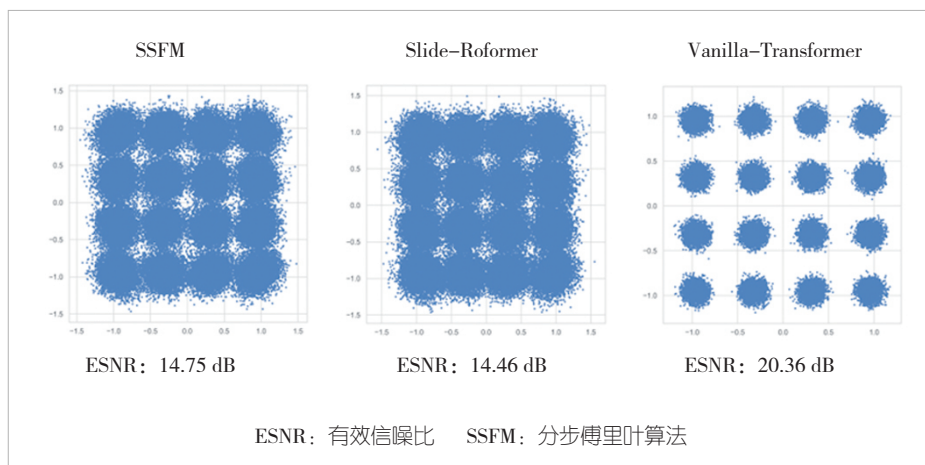


图 5 SSFM、Slide-Roformer、Vanilla-Transformer 在 800 km 处的星座图和对应的 ESNR

3.4 Slide-Roformer 复杂度优化

在模型推理过程中，采用多符号输出方式可有效减少两端填充符号的重复计算次数，从而缩短推理时间。为验证多符号输出机制的有效性，我们在建模相同总长度的符号序列时，分别设定模型单次推理预测 50 至 5 000 个符号，并记录完整信号建模所需时间。所有神经网络模型均部署于同一台配备 NVIDIA GeForce RTX 4090 24 GB GPU 的服务器上。

具体结果如图 6 所示，随着单次输出符号数量的增加，总计算时间逐渐减少，从输出 50 个符号时的 65 s 优化至输出 5 000 个符号时的 4.6 s，证明了该方法在降低计算时间方面的有效性。进一步地，为确保多符号输出模式下模型精度不发生劣化，我们测试了 Slide-Roformer 与 Vanilla-

Transformer在单次输出 300~5 000 个符号时的 NMSE，如图 7 所示。

在 800 km 传输距离下，Slide-Roformer 在不同输出符号长度下的 NMSE 整体稳定在约 2.8×10^{-3} 的水平，而 Vanilla-Transformer 的 NMSE 随输出符号长度的增加持续恶化，在单次输出 5 000 个符号时达到 1.9，比 Slide-Roformer 高出 3 个数量级。这表明 Slide-Roformer 能够有效适用于多符号预测模式，而 Vanilla-Transformer 在多符号预测中出现了显著的波形失真，无法保持建模精度。

这一优势得益于 Slide-Roformer 在位置编码与注意力机制上的有效设计：首先，RoPE 采用的相对位置编码保证了模型在不同输入长度下仍能有效感知符号间的相对位置信息；其次，滑动窗口注意力机制将符号的关联范围限制在特定长度内，使得无论输入序列多长，每个时刻的符号均可在固定范围内以相同模式学习非线性相关性，从而实现稳定有效的非线性建模。

为进一步评估计算效率，我们将 Slide-Roformer 与数值方法 SSFM 以及神经网络模型 Vanilla-Transformer、Seq2Seq-BiLSTM 和 FNO 进行了计算时间对比。Slide-Roformer 与 Seq2Seq-BiLSTM 均单次预测 5 000 个符号；而 Vanilla-Transformer 为了保持较高精度，单次预测 300 个符号；FNO 则与训练配置一致，单次预测 20 个符号。表 3 统计了上述方法在传输 80 km（单个跨段）时的计算时间。结果表明，Slide-Roformer 的计算时间仅为 4.6 s，较 Vanilla-Transformer（15.3 s）、Seq2Seq-BiLSTM（10.9 s）和 FNO（91.2 s）具有明显优势。FNO 因采用少符号预测模式，需多次循环并重复计算填充符号，导致复杂度较高。相比之下，Slide-Roformer

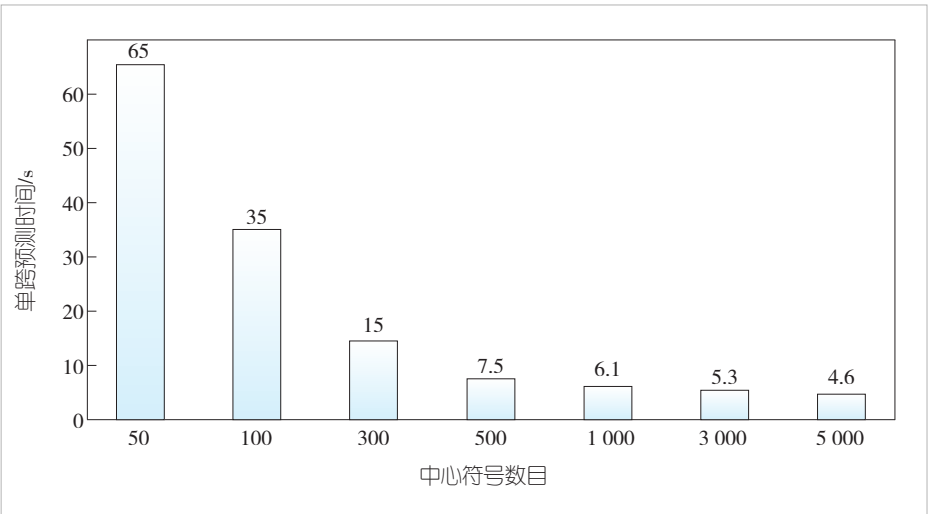


图6 不同输出中心符号长度下神经网络的单跨预测时间

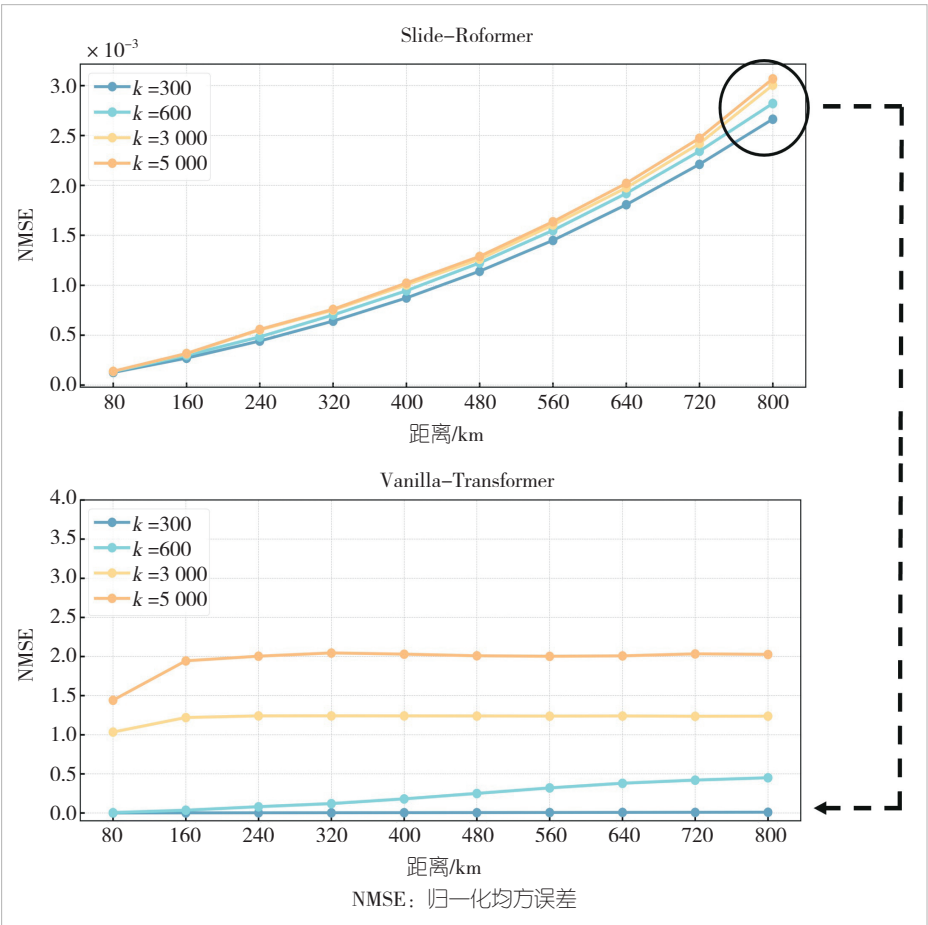


图7 Slide-Roformer和Transformer输出不同符号的NMSE对比

表3 不同模型的计算时间(传输距离 80 km)

模型	Slide-Roformer	Vanilla-Transformer	Seq2Seq-BiLSTM	FNO	SSFM
时间/s	4.6	15.3	10.9	91.2	151

FNO: 傅里叶神经算子 SSFM: 分步傅里叶算法

通过将全局注意力优化为滑动窗口注意力,在多符号预测模式下降低了注意力模块的计算复杂度,从而较传统 Trans-former 更为高效。与传统数值方法 SSFM (仿真时间 151 s) 相比, Slide-Roformer 实现了 96.9% 的加速。以上结果充分表明, Slide-Roformer 能够有效克服传统算法复杂度高的难题,有望成为下一代高速光传输系统中信道建模的高效工具。

4 结束语

本文针对传统分步傅里叶方法计算复杂度高的问题,提出了一种用于光纤信道建模的低复杂度 Slide-Roformer 架构。通过引入 RoPE 与滑动窗口注意力机制,该模型能够更有效地学习符号间的相对非线性串扰关系,并显著减少计算时间。实验结果表明,与传统 Transformer 相比, Slide-Roformer 的波形误差 NMSE 改善了 74%, 计算时间降低了 69.9%。与 SSFM 方法相比, ESNR 仅为 0.15 dB, 计算时间可降低 96.9%。这些结果验证了 Slide-Roformer 在光纤信道建模中兼具高精度与低复杂度的优势,为相关研究提供了一种高效的新方法,有望推动下一代高速光通信系统的进一步发展。

参考文献

- [1] ZHANG Y, PANG X H, SONG Y C, et al. Optical power control for GSNR optimization based on C+L-band digital twin systems [J]. Journal of lightwave technology, 2023, 42(1): 95–105. DOI: 10.1109/jlt.2023.3303783
- [2] CORREIA B, SADEGHI R, VIRGILLITO E, et al. Optical power control strategies for optimized C+L+S-bands network performance [C]//Proceedings of Optical Fiber Communication Conference (OFC) 2021. Optica Publishing Group, 2021: 1–3. DOI: 10.1364/ofc.2021.w1f.8
- [3] ROBERTS I, KAHN J M, HARLEY J, et al. Channel power optimization of WDM systems following Gaussian noise nonlinearity model in presence of stimulated Raman scattering [J]. Journal of lightwave technology, 2017, 35(23): 5237–5249. DOI: 10.1109/JLT.2017.2771719
- [4] SONG Y C, FAN Q R, LU C, et al. Efficient three-step amplifier configuration algorithm for dynamic C L-band links in presence of stimulated Raman scattering [J]. Journal of lightwave technology, 2023, 41(5): 1445–1453. DOI: 10.1109/JLT.2022.3223919
- [5] ZHANG C Y, WANG D S, JIA J W, et al. Potential failure cause identification for optical networks using deep learning with an attention mechanism [J]. Journal of optical communications and networking, 2022, 14(2): A122. DOI: 10.1364/jocn.438900
- [6] FAN Q R, ZHOU G, GUI T, et al. Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning [J]. Nature communications, 2020, 11: 3694. DOI: 10.1038/s41467-020-17516-7
- [7] DELIGIANNIDIS S, BOGRIS A, MESARITAKIS C, et al. Compensation of fiber nonlinearities in digital coherent systems leveraging long short-term memory neural networks [J]. Journal of lightwave technology, 2020, 38(21): 5991–5999. DOI: 10.1109/JLT.2020.3007919
- [8] LIN X, LUO S H, SOMAN S K O, et al. Perturbation theory-aided learned digital back-propagation scheme for optical fiber nonlinearity compensation [J]. Journal of lightwave technology, 2022, 40(7): 1981–1988. DOI: 10.1109/JLT.2021.3133475
- [9] NIU Z K, YANG H, LI L, et al. Learnable digital signal processing: a new benchmark of linearity compensation for optical fiber communications [J]. Light: science & applications, 2024, 13: 188. DOI: 10.1038/s41377-024-01556-5
- [10] IP E, KAHN J M. Compensation of dispersion and nonlinear impairments using digital backpropagation [J]. Journal of lightwave technology, 2008, 26(20): 3416–3425. DOI: 10.1109/JLT.2008.927791
- [11] KARANOV B, CHAGNON M, AREF V, et al. Concept and experimental demonstration of optical IM/DD end-to-end system optimization using a generative model [C]//Proceedings of Optical Fiber Communication Conference (OFC) 2020. Optica Publishing Group, 2020. DOI: 10.1364/ofc.2020.th2a.48
- [12] NIU Z K, YANG H, ZHAO H C, et al. End-to-end deep learning for long-haul fiber transmission using differentiable surrogate channel [J]. Journal of lightwave technology, 2022, 40(9): 2807–2822. DOI: 10.1109/JLT.2022.3148270
- [13] LI L, NIU Z K, XIAO J Z, et al. Joint pre- and post-learned perturbation nonlinearity compensation optimization for long-haul optical fiber transmission based on end-to-end deep learning [C]//Proceedings of Optical Fiber Communication Conference (OFC) 2025. Optica Publishing Group, 2025. DOI: 10.1364/ofc.2025.w1k.6
- [14] LI Z Y, WANG C X, JIA J L, et al. Model-driven deep-learning for end-to-end optimization in fiber-terahertz communication systems [J]. Journal of lightwave technology, 2025, 43(7): 3099–3117. DOI: 10.1109/JLT.2024.3519360
- [15] AGRAWAL G P. Chapter 2 – pulse propagation in fibers [EB/OL]. [2025-11-10]. <https://www.oreilly.com/library/view/nonlinear-fiber-optics/9780128170434/B9780128170427000099.xhtml>
- [16] POGGIOLINI P. The GN model of non-linear propagation in uncompensated coherent optical systems [J]. Journal of lightwave technology, 2012, 30(24): 3857–3879
- [17] POGGIOLINI P, BOSCO G, CARENA A, et al. The GN-model of fiber non-linear propagation and its applications [J]. Journal of lightwave technology, 2014, 32(4): 694–721. DOI: 10.1109/JLT.2013.2295208
- [18] SERENA P, LASAGNI C, MUSETTI S, et al. On numerical simulations of ultra-wideband long-haul optical communication systems [J]. Journal of lightwave technology, 2020, 38(5): 1019–1031. DOI: 10.1109/JLT.2019.2938580
- [19] WANG D S, SONG Y C, LI J, et al. Data-driven optical fiber channel modeling: a deep learning approach [J]. Journal of lightwave technology, 2020, 38(17): 4730–4743. DOI: 10.1109/JLT.2020.2993271
- [20] YANG H, NIU Z K, XIAO S L, et al. Fast and accurate optical fiber channel modeling using generative adversarial network [J]. Journal of lightwave technology, 2021, 39(5): 1322–1333. DOI: 10.1109/JLT.2020.3037905
- [21] HE X C, YAN L S, JIANG L, et al. Fourier neural operator for accurate optical fiber modeling with low complexity [J]. Journal of lightwave technology, 2023, 41(8): 2301–2311. DOI: 10.1109/jlt.2022.3229015

- [22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. [2025-11-02]. <https://www.tech-odyssey.cn/pdf/AI/Attention-is-all-your-need-ZH.pdf>
- [23] SHI M H, YANG H, NIU Z K, et al. Accurate and efficient optical fiber WDM transmission modeling using the encoder-only transformer with feature decoupling distributed method [C]// Proceedings of Asia Communications and Photonics Conference/ 2023 International Photonics and Optoelectronics Meetings (ACP/POEM). IEEE, 2023: 1-5. DOI: 10.1109/ACP/POEM59049.2023.10368835
- [24] ZHUANG B, LIU J, PAN Z, et al. A survey on efficient training of transformers [EB/OL]. [2025-11-10]. <https://arxiv.org/abs/2302.01107>
- [25] SU J L, AHMED M, LU Y, et al. RoFormer: enhanced transformer with rotary position embedding [J]. Neurocomputing, 2024, 568: 127063. DOI: 10.1016/j.neucom.2023.127063
- [26] BELTAGY I, PETERS M E, COHAN A. Longformer: the long-document transformer [EB/OL]. [2025-11-05]. https://milern.readthedocs.io/zh/latest/source/pretrain_model/longformer.html
- [27] MARCUSE D, MANYUK C R, WAI P K A. Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence [J]. Journal of lightwave technology, 1997, 15(9): 1735-1746. DOI: 10.1109/50.622902
- [28] EVANGELIDES S G, MOLLENAUER L F, GORDON J P, et al. Polarization multiplexing with solitons [J]. Journal of lightwave technology, 1992, 10(1): 28-35. DOI: 10.1109/50.108732
- [29] SHAO J, LIANG X J, KUMAR S. Comparison of split-step Fourier schemes for simulating fiber optic communication systems [J]. IEEE photonics journal, 2014, 6(4): 7200515
- [30] SHI M H, NIU Z K, YANG H, et al. Fast and accurate waveform modeling based on sequence-to-sequence framework for multi-channel and high-rate optical fiber transmission[J]. Optics Letters, 2025, 50(7): 2286-2289
- [31] SINKIN O V, HOLZLOHNER R, ZWECK J, et al. Optimization of the split-step Fourier method in modeling optical-fiber communications systems [J]. Journal of lightwave technology, 2003, 21(1): 61-68. DOI: 10.1109/JLT.2003.808628
- [32] YANG H, NIU Z K, ZHAO H C, et al. Fast and accurate waveform modeling of long-haul multi-channel optical fiber transmission using a hybrid model-data driven scheme [J]. Journal of lightwave technology, 2022, 40(14): 4571-4580. DOI: 10.1109/JLT.2022.3168698

作者简介



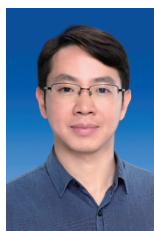
史明辉，上海交通大学集成电路学院（信息与电子工程学院）在读博士研究生；研究方向为基于深度学习的光通信系统建模；发表论文 5 篇。



郑智雄，上海交通大学集成电路学院（信息与电子工程学院）在读硕士研究生；主要研究方向为基于深度学习的光通信系统建模。



牛泽坤，上海交通大学博士后；研究并开发的 AI 光纤信道模型实现上千倍的速度提升，合作成果已应用于华为光产品线，主导开发的中国首个智能光传输开源仿真平台（ifibertrans.sjtu.edu.cn），被近百家研究机构使用，中国电信评价其为“光通信仿真技术国产化的里程碑产品”；发表论文 10 余篇。



义理林，上海交通大学教授、博士生导师，区域光纤通信网与新型光通信系统国家重点实验室副主任，OFC TPC 成员（2017—2019），《Optical Fiber Technology》编委；主要从事智能光通信和智能光纤激光器方面的研究；所负责的智能锁模激光器工作被美国光学学会杂志《Optics & Photonics News》评为 2019 年度全球光学进展，并入选 2019 中国光学十大进展，2020 年获国家杰出青年科学基金项目资助；发表论文 100 余篇。