

# 메모리 효과를 효과적으로 보상하는 트랜스포머 기반 디지털 사전왜곡 모델 구현

최민호, 정기석\*  
한양대학교

mkiuyh@hanyang.ac.kr, \*kchung@hanyang.ac.kr

## Implementation of a Transformer-based Digital Predistortion Model for Effective Compensation of Memory Effects

Minho Choi, Ki-Seok Chung\*  
Hanyang University, Seoul, Korea

### 요약

무선 통신 시스템의 송신단 핵심 소자인 전력 증폭기(Power Amplifier, PA)는 신호를 증폭하여 원거리까지 전송하는 역할을 수행한다. 그러나 PA의 동작 과정에서 발생하는 비선형 특성은 신호 품질 저하와 에너지 효율 감소의 원인이 된다. 이러한 문제를 해결하고자 비선형성 보정을 위한 디지털 사전왜곡(Digital Predistortion, DPD) 연구가 활발히 진행되어 왔으며, 그 기법이 메모리 다항식(Memory Polynomial)부터 인공신경망(Neural Network) 모델로 고도화되었다. PA의 메모리 효과는 현재 입력 신호뿐만 아니라 과거 입력 신호의 히스토리 때문에 증폭기 특성이 변하거나 비선형 왜곡이 발생하는 현상으로, 본 논문에서는 긴 시퀀스의 과거 데이터를 입력으로 받아 효과적으로 보상할 수 있도록 트랜스포머(Transformer) 기반 DPD 모델을 제안한다. 이를 통해 장기 메모리 효과(Long-Term Memory Effects)를 효과적으로 보상하며 비선형 보정 성능을 크게 개선하였다. 실험 결과, 제안하는 모델은 기존 SOTA(State-of-the-art) 모델인 DeltaGRU-tcnskip 대비 EVM은 3.34 배, ACLR은 2.48 배 향상된 성능을 보였다.

### I. 서론

6G 및 5G Advanced 환경 등 무선 통신 시스템에서 전력 증폭기(Power Amplifier, PA)는 전송 거리 확보, 신호 대 잡음비(Signal-to-Noise Ratio, SNR) 유지, 효율적인 에너지 사용 등의 통신 품질을 결정하는 가장 중요한 부품 중 하나로 꼽힌다. PA는 에너지 효율 극대화를 위해 주로 포화 영역(Saturation Operation Point) 근처에서 동작 시키지만, 비선형(Non-linear) 특성이 나타나고 과형의 왜곡이 발생한다. 이러한 PA의 비선형성에 의해 신호가 변질되어 오차 벡터 크기(Error Vector Magnitude, EVM)를 악화시키고 인접 채널 간섭(Adjacent Channel Interference)을 일으킨다.[1]

PA의 비선형 왜곡 문제를 해결하기 위해 디지털 사전왜곡(Digital Predistortion, DPD) 모델을 사용한다. DPD는 PA의 비선형 특성과 정반대되는 역특성을 입력 신호에 미리 적용함으로써, 최종 출력에서 선형성을 확보하는 원리를 가진다. 최근 복잡한 환경에서의 PA의 메모리 효과(Memory Effect)를 모델링하기 위해 딥러닝 기법을 사용하는데, 특히 시계열 데이터 처리에 강점을 가진 순환 신경망(Recurrent NN) 기반의 DeltaGRU[2] 모델이 SOTA 성능을 입증하였다. 그러나 RNN 모델은 순차 처리에 의해 연산 병렬화에 한계가 있으며, 장기 의존성(Long-term dependency) 문제로 인해 정보 손실이 발생할 수 있다. 반면, Transformer[3] 기반 신경망은 셀프 어텐션(Self-Attention) 메커니즘을 통해 입·출력 데이터 간의 상관관계를 병렬적으로 분석함으로써 연산 효율성을 높이고, 더욱 복잡한 비선형 메모리 효과를 정밀하게 모델링하는 데 탁월한 장점을 보인다. 이에 본 논문에서는 기존 GRU 기반 모델의 한계를 극복하고 비선형 보정 성능을 극대화하기 위한 Transformer-based DPD 구조를 제안한다.

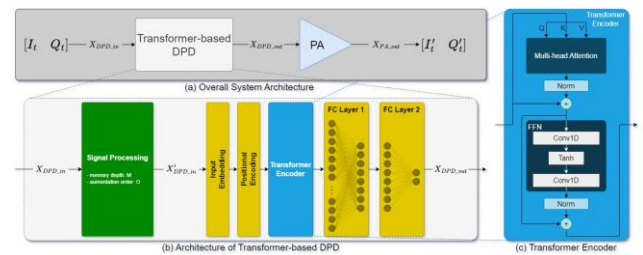


그림 1. 제안하는 Transformer-based DPD 아키텍처

### II. 본론

본 장에서는 Transformer-based DPD 아키텍처와 그 성능 검증 결과를 기술한다.

#### 1. Transformer-based DPD 모델

제안하는 모델의 상세 아키텍처는 그림 1(b)에 나타나 있다. 신호 처리(Signal Processing), 트랜스포머 인코더(Transformer Encoder), 그리고 최종 출력단인 완전 연결 계층(Fully Connected Layer)으로 구성된다.

1.1) 입력 신호 처리 및 임베딩: Signal Processing은  $[I_t, Q_t]$  쌍으로 표현되는 입력 신호로부터 특징을 추출한다. 식 (1)의 입력 데이터는 memory depth  $M$  만큼의 과거 시점 데이터와 이들의 진폭( $|x_t|$ ), 삼각함수 항( $\sin\theta_t, \cos\theta_t$ ) 등을 포함하는 식 (2)의 형태로 증강되어 DPD 모델 입력으로 사용된다. 이 증강된 시퀀스 데이터는 Input Embedding과 Positional Encoding layer를 거쳐 트랜스포머 인코더에 입력된다.[3]

표 1. 모델에 따른 DPD의 성능 비교

Metric	without DPD	DeltaGRU	Transformer
EVM (dB)	-14.419784	-40.446561	-50.915954
ACLR (dB)	-30.822504	-49.129678	-53.069202

$$X_{DPD\_in} = \begin{bmatrix} I_t \\ Q_t \end{bmatrix}^T \quad (1)$$

$$X'_{DPD\_in} = \begin{bmatrix} I_t & I_{t-1} & I_{t-2} & \cdots & I_{t-(M-1)} \\ Q_t & Q_{t-1} & Q_{t-2} & \cdots & Q_{t-(M-1)} \\ |x_t| & |x_{t-1}| & |x_{t-2}| & \cdots & |x_{t-(M-1)}| \\ |x_t|^3 & |x_{t-1}|^3 & |x_{t-2}|^3 & \cdots & |x_{t-(M-1)}|^3 \\ \sin\theta_t & \sin\theta_{t-1} & \sin\theta_{t-2} & \cdots & \sin\theta_{t-(M-1)} \\ \cos\theta_t & \cos\theta_{t-1} & \cos\theta_{t-2} & \cdots & \cos\theta_{t-(M-1)} \end{bmatrix}^T \quad (2)$$

1.2) 트랜스포머 인코더: 그림 1 (c)는 트랜스포머 인코더 블록의 상세 구조를 보여준다. 먼저 MHA는 입력 시퀀스 내의 다양한 위치에 있는 정보로부터 관계성을 학습한다.  $d_{model} = 6$ 을  $n_{head} = 2$ 개의 헤드로 분할하여 병렬 처리한다. 이는 PA 메모리 효과의 복잡한 비선형적 의존성을 포착하는 데 효과적이다. 다음으로 FFN에서 기존 연구[3]의 단순 선형 계층(Linear layer) 방식 대신, locally feature extraction 능력을 강화하기 위해 Conv1D를 사용한다. 이를 위해 두 Conv1D 연산에서  $d_{ff} = 10$ 의 차원으로 처리한다. 또한, 전체 모델의 복잡성을 줄이고 실시간 DPD 연산의 효율성을 확보하기 위해 단일 트랜스포머 레이어( $n_{layer} = 1$ )만을 사용하였다.

1.3) 완전 연결 계층: 트랜스포머 인코더를 거쳐 출력된 특징 맵은 최종적으로 원하는 출력 신호를 생성하기 위해 두 개의 FC layer를 통과한다. FC layer 1은 인코더의 최종 출력 텐서를 flattening한 후 단일 은닉층을 통해 데이터의 차원을  $d_{FC1} = 8$ 로 축소한다. FC layer 2는 그 데이터의 차원을  $d_{FC2} = 2$ 로 다시 한번 더 축소하여 최종적으로 PA의 비선형성이 보정된 2개의 출력 신호, 즉 동상 성분(I)과 직교 성분(Q)을 추출한다. 이를 통해 PA 입력 신호  $X_{DPD\_out}$ 이 결정된다.

### 3. 실험 결과 및 성능 분석

본 장에서는 DeltaGRU-, Transformer-based DPD 모델의 성능을 비교한 결과를 분석한다. 각 모델은 Python 및 PyTorch 기반으로 구현했다. 실험 방법은 다음과 같다. 먼저 실제 PA 데이터를 이용해 비선형 특성을 갖는 PA 모델을 학습한다. 동일 구조 모델을 이용한 DPD 모델과 결합하여 하나의 cascaded 모델을 구성한다. DPD 모델은 cascaded 모델이 최대한 선형적 특성을 보이도록 PA 모델의 입력을 사전왜곡하도록 훈련되었다. 이때 PA 모델의 파라미터는 고정(frozen)한 채로 훈련을 진행한다.

그림 2는 AM-AM 및 AM-PM 그래프이다. 그림 2(a)를 통해 알 수 있듯, DPD model을 사용하지 않은 경우 심각한 비선형 왜곡을 보였다. DeltaGRU-based DPD를 사용하여 이를 상당 부분 개선한 것을 그림 2(b)에서 볼 수 있다. 그림 2(c)는 Transformer-based DPD가 가장 우수한 선형성과 최소한의 위상 왜곡을 달성했음을 보여준다.

표 1의 EVM 및 ACLR 평가 결과에서도 Transformer-based DPD가 가장 낮은 값을 기록하여, 비선형 보상 및 스펙트럼 누설 억제 성능이 가장 우수함을 확인하였다. 기존 SOTA 모델인 DeltaGRU 대비 EVM은 3.34배, ACLR은 2.48배 향상된 성능을 보였다.

### III. 결론

본 논문에서는 6G 및 5G Advanced 환경에서 요구되는 고효율 PA의 비선형성과 메모리 효과를 극복하기 위해, Transformer-based DPD 아키텍처를 제안하였다. 제안된 모델은 특징 증강 과정을 통해 PA

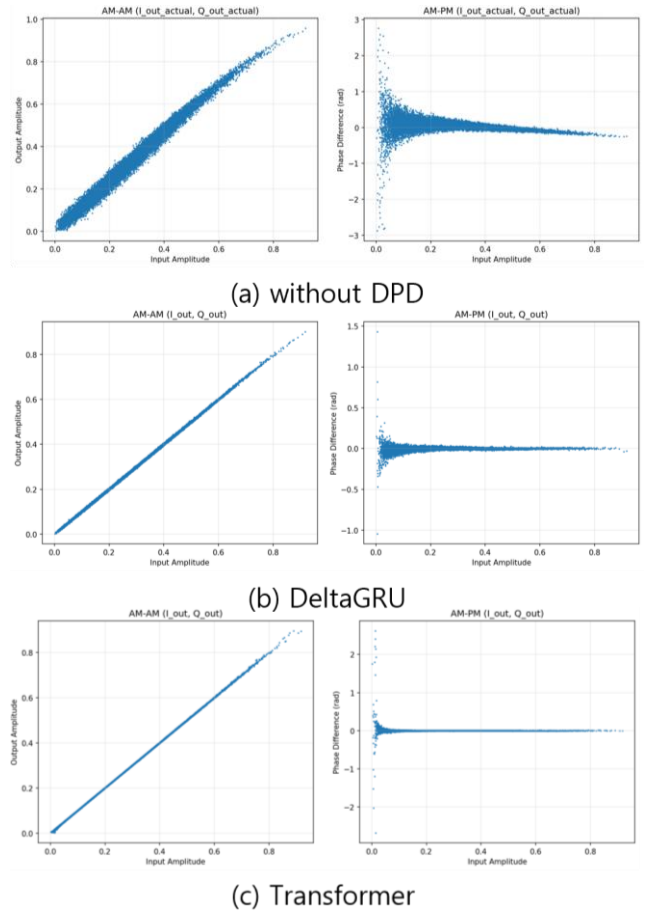


그림 2. AM-AM, AM-PM 그래프

특성을 효과적으로 모델링하였으며, Transformer의 셀프 어텐션 메커니즘을 활용하여 기존 RNN 기반 SOTA 모델 대비 월등한 선형화 성능을 입증하였다. 실험 결과, EVM과 ACLR 지표에서 각각 x3.34와 x2.48의 성능 향상을 달성하며 DPD 기술의 새로운 가능성을 제시하였다. 제안된 아키텍처는 향후 LLM(Large Language Model) 기반 모델로 확장함으로써, 보다 복잡한 환경에서도 PA 선형성을 안정적으로 보장할 수 있는 방향으로 발전시키는 것을 목표로 한다.

### ACKNOWLEDGMENT

이 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2024-00409492).

### 참고 문헌

- [1] Ku, Hyunchul, Michael D. McKinley, and J. Stevenson Kenney. "Quantifying memory effects in RF power amplifiers." IEEE Transactions on microwave theory and techniques 50.12 (2002): 2843-2849.
- [2] Wu, Yizhuo, Ang Li, and Chang Gao. "OpenDPDv2: A Unified Learning and Optimization Framework for Neural Network Digital Predistortion." arXiv preprint arXiv:2507.06849 (2025).
- [3] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).