Attention Mechanism Based Bidirectional LSTM Model for Broadband Power Amplifier Linearization

Rina Su [1]✉, Jiacheng Wang [2], Gaoming Xu [2] and Taijun Liu [2]✉

[1] the School of Cyber Science and Engineering (School of Cyber Science and Engineering), Ningbo University of Technology, Ningbo, Zhejiang, 315211, China;
[2] the Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, Zhejiang 315211, China;
✉ E-mail: srn@nbut.edu.cn
E-mail: liutaijun@nbu.edu.cn

In this letter, a novel model for broadband power amplifier (PA) linearization is proposed, namely Attention Mechanism based Bidirectional Long Short-term Memory network (AM-BiLSTM). In order to verify the linearization performance of the AM-BiLSTM model, a 100MHz bandwidth 5G new radio (5G NR) signal is employed to test the sub-6G PA operating at 2.6-GHz. The experimental results show that the adjacent channel power ratio (ACPR) of the PA with AM-BiLSTM can be improved by 24dB which is 6-dB better than the generalized memory polynomial (GMP) and 3-dB better than the Chebyshev polynomials LSTM (CP-LSTM) in ref [1]. Therefore, the proposed AM-BiLSTM is very effective for the linearization of broadband PA.

Introduction: As the development of the wireless communication system, the signal is tend to large bandwidth and high peak to average power ratio (PAPR) [2] which will exacerbate the PAs' memory effect and nonlinearity especially for the PAs operating in the low frequency band. To put it more elegantly, the relative bandwidth of a 100-MHz signal when paired with a 3.5-GHz PA is comparable to the relative bandwidth of an 800-MHz signal when matched with a 28-GHz PA. That is, PAs in the lower frequency range tend to exhibit stronger memory effects. This highlights the importance of further investigation into developing models that are capable of capturing and modeling the strong memory effects of these amplifiers.

The robust fitting abilities of neural network models make them a viable solution for modeling PAs' strong memory effects[3]. Researchers have introduced Convolutional Neural Networks (CNNs) [4], and LSTM network models [5] [6] of deep neural networks, which have shown promising results in various nonlinear model identification domains. Among this model, LSTM model has a more accurate characterization of PAs' memory effects due to its sensitivity to time series.

Though, the CP-LSTM has shown strong ability for linearizing the wideband 5G mmW PA. It still need a more efficient model for the PA in sub-6G frequency band. We porpose a novel model AM-BiLSTM which has more effective in describing memory effects. In AM-BiLSTM, the attention mechanism[7] allows the model to automatically determine which parts of the input are most important in making a prediction, rather than processing the entire input sequence equally. This allows the model to handle input sequences of varying lengths, attend to relevant information and make more accurate predictions. Therefore, we propose a pre-distorter model that leverages the attention mechanism to build a bidirectional LSTM, which accurately captures both the advanced and delayed memory effects of the PA, thereby providing a more comprehensive representation of the PAs' strong memory effect. The measured results demonstrate that the AM-BiLSTM exhibits superior linearization performance compared to other neural network models in cases where the PA exhibits strong memory effects.

AM-BiLSTM Model Architecture: The AM-BiLSTM model is composed of a normalization layer, BiLSTM layer (a forward LSTM layer and a backward LSTM layer) and an attention layer. The overall structure of the AM-BiLSTM-based power amplifier model is depicted in Figure 1.

In Figure 1, the I_{in} is the real value of the in-phase baseband signal and the Q_{in} is the quadrature-phase part. We use the maximum normalization to normalize the input IQ data which can be expressed as

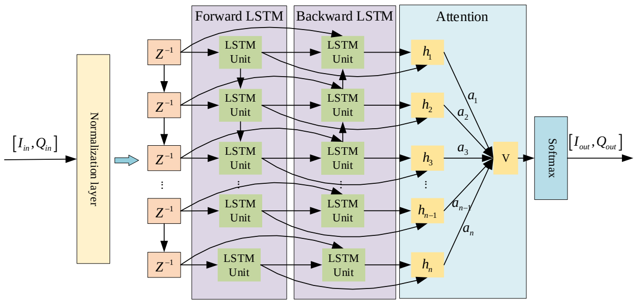

Fig. 1 *Proposed AM-BiLSTM model.*

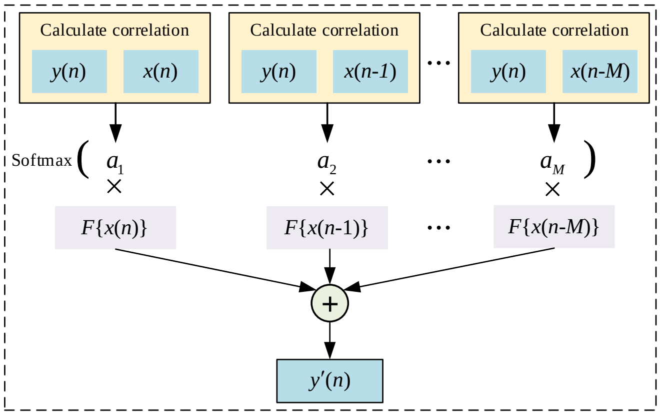

Fig. 2 *Attention mechanism calculation process.*

$$I_{in,N}(k)=\frac{I_{in}(k)}{\sum_{k=1}^{K}\sqrt{I_{in}^2(k)+Q_{in}^2(k)}},\quad k=1,2,\ldots,K \tag{1}$$

$$Q_{in,N}(k)=\frac{Q_{in}(k)}{\sum_{k=1}^{K}\sqrt{I_{in}^2(k)+Q_{in}^2(k)}},\quad k=1,2,\ldots,K \tag{2}$$

The structure of the LSTM layer has been thoroughly described in ref [1]. In BiLSTM layer, information data is fed into the model from both the forward and backward directions. The forward LSTM network learns the delayed historical information, and the backward LSTM network learns the future information ahead of time, thus achieving efficient global information training.

In an effort to minimize the computational complexity of the network, we investigate the utilization of attention mechanism to effectively optimize and choose memory items within the BiLSTM layer. The attention mechanism selectively preserves memory items with significant impact, while discarding those with minimal effect. The implementation of the attention mechanism is shown in Figure 2. The calculation process of the Attention mechanism designed in this paper is as follows:

First, the correlation between each memory item $x(n-m-1)$ and the output $y(n)$ is calculated. Then, the Softmax function is used to convert the correlation values obtained from the positive and negative directions of the memory items to the output into numerical values, so that the results meet the probability distribution with the sum of weights equal to 1. Finally, the calculated correlation coefficient a_m is multiplied by the PA model function $f(x(n-m))$, and the sum of these M products is calculated to obtain the Attention value $y'(n)$, which is the predicted value of the output $y(n)$. This way, a_m is used to select and retain memory items based on their contribution.

$$y'(n)=\sum_{m=0}^{M} a_m \times F\left\{x(n-m)\right\} \tag{3}$$

The calculation of the correlation, as mentioned in the previous text, is performed by multiplying the output matrix A of the BiLSTM layer with its transpose to obtain the correlation coefficient matrix B. The product of matrix B and A is then concatenated with A, and a fully connected layer is used to transform the data dimension. Finally, the model output is obtained.

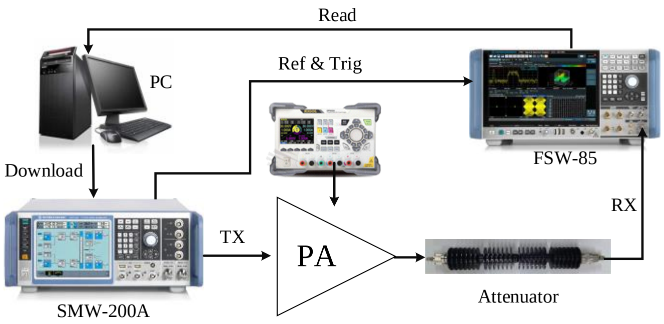

Fig. 3 *Digital predistortion hardware platform diagram.*

AM-BiLSTM Networks Model Training: The proposed model is implemented in the TensorFlow2.0 platform with the training steps as follows:

Step-1: Initialize the designed variables, set the epoch number and limited tolerance, load the training data. The Xavier parameter initialization method is employed to ensure that information flows more effectively within the network and that the variance of the output from each layer is as equal as possible.

Step-2: Build the AM-BiLSTM network. Set the memory term as 5, the order as 7, the time step as 2.

Step-3: Set the loss function $f_{loss}(k)$ as the mean squre error (MSE) between the predicted output $y'_{out}(k)$ and the measured output $y_{out}(k)$. The $f_{loss}(k)$ can be expressed as:

$$f_{Loss}(k)=\frac{1}{N}\sum_{n=0}^{N}\left\{\left[I'_{out}(k)-I_{out}(k)\right]^2+\left[Q'_{out}(k)-Q_{out}(k)\right]^2\right\} \quad (4)$$

Step-4: Feed the training data into the network and train the model at each epoch. When the epoch number or limited tolerance is reached, the behavioral model can be used to predict a new output.

Model Validation: In order to measure the linearization performance of the AM-BiLSTM model, a 5G sub-6G PA DPD experimental setup is built as Figure 3 exhibited. The experimental instruments used in this experiment contain a Rohde & Schwarz (R&S) vector signal generator (SMW-200A), an R&S power spectrum analyzer (FSW-85), a driver PA, an attenuator, some power suppliers, and RF cables. The test signal adopts a 5G NR signal with a bandwidth of 100 MHz and a peak-to-average-power ratio of 10.9-dB. The peak output power of the Doherty PA is 40 dBm and the center frequency is 2.6 GHz.

Modeling Result Analysis: In order to quantitatively describe the models' accuracy, the normalized mean squared error (NMSE) is employed to measure the error between the output of the model and the measured output of the PA. Table I list the NMSE of the generalized memory polynomial (GMP) [8], LSTM, CP-LSTM and the AM-BiLSTM with the 5G NR signal. The coefficient settings of the LSTM, CP-LSTM and AM-BiLSTM are all the same . The lagging /leading envelope memory depth is set to 2, memory depth is set to 5, and the order is set to 7 for the GMP to operate at the best linearization performance. From Table 1, it can be seen that the AM-BiLSTM model exhibits better accuracy and only causes a little complexity.

Figure 4 gives the comparisons of the amplitude modulation / amplitude modulation (AM/AM) and amplitude modulation/phase modulation (AM/PM) of the PA between simulated and measured, which describes the high accuracy of the AM-BiLSTM in another way.

Linearization Result Analysis: In order to measure the linearization performance of the AM-BiLSTM, a linearization performance test is taken in this letter. Figure 5 exhibits the PA output spectrums with DPD built by the GMP, LSTM, CP-LSTM, AM-BiLSTM and without any linearizer.

Table 1: Comparison of the models

Model	Numbers of (Multiplies, Additions)	NMSE
GMP	(2520,1470)	-35.79
LSTM	(1932,1596)	-36.48
CP-LSTM	(1968,1658)	-37.96
AM-BiLSTM	(2496,1748)	-40.77

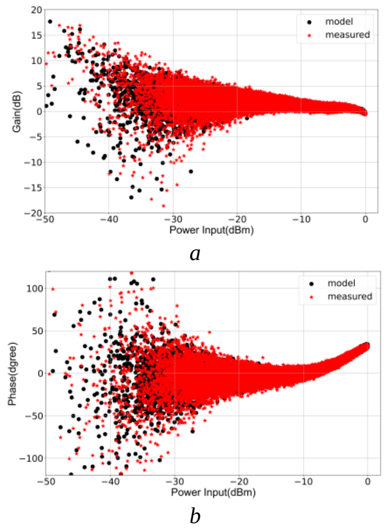

Fig. 4 *Comparison between measured and AM-BiLSTM modeled characteristics of the PA.*
a AM/AM
b AM/PM

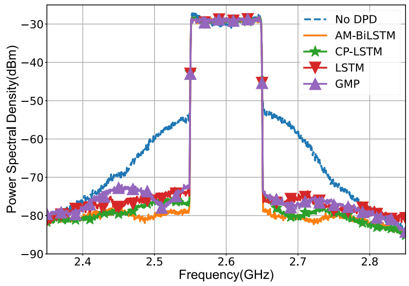

Fig. 5 *Comparison of the linearization capability for the four predistorters.*

Table 2: ACPR of the PA under different models

Model	ACPR +/-50MHz	ACPR +/-100MHz
W/O DPD	-27.26/-25.39-dB	-33.56/-31.34-dB
GMP	-44.71/-45.98-dB	-47.79/-47.86-dB
LSTM	-45.30/-45.51-dB	-47.51/-47.14-dB
CP-LSTM	-47.82/-45.60-dB	-48.12/-50.87-dB
AM-BiLSTM	-50.93/-50.02-dB	-51.86/-50.91-dB

Figure 5 clearly illustrates that the traditional linearization model GMP do not provide satisfactory linearization results when dealing with the PA with strong memory effects. In contrast, neural network models show significantly better performance than the traditional models. In the neural network models, while the mmW model CP-LSTM model has demonstrated a clear improvement over the LSTM model, its efficacy in the sub-6GHz frequency band remains limited. The AM-BiLSTM model we proposed has shown remarkable performance in broadband linearization at sub-6GHz frequency band.

Conclusion: In this paper, we propose a 5G amplifier non-linear behavior model namely AM-BiLSTM. We have presented a detailed description of the structure of the AM-BiLSTM model as well as the thought process behind constructing the model. In order to test the effectiveness of the model, a pre-distorter system experiment platform in sub-6G frequency band is build. By observing the power spectral density it is found that the linear improvement of the ACPR of AM-BiLSTM model can be improved by about 24dB, which is 6-dB better than the GMP and 3-dB better than the CP-LSTM. In summary, AM-BiLSTM model has good practical effects for wideband PAs with strong memory effects.

Acknowledgments: This work was supported by the Natural Science Foundation of China under Grant 62071264; in part by the Natural

Science Foundation of Zhejiang Province under Grant LY21F010004; and by the Research Project of Ningbo University of Technology 2022KQ54.

data availability: from the corresponding author upon reasonable request.

conflict of interest: all authors have not any conflicts of interest.

References

1. G. Xu, H. Yu, C. Hua and T. Liu, ‘Chebyshev Polynomial-LSTM Model for 5G Millimeter-Wave Power Amplifier Linearization’, IEEE Microw. Wireless Compon. Lett., vol. 32, no. 6, pp. 611-614, June 2022, doi: 10.1109/LMWC.2022.3142423.
2. Z. Popovic, ‘Amping Up the PA for 5G: Efficient GaN Power Amplifiers with Dynamic Supplies’, IEEE Microw. Mag., vol. 18, no. 3, pp. 137-149, May 2017, doi: 10.1109/MMM.2017.2664018.
3. S. Zhang et al., ‘Deep neural network behavioral modeling based on transfer learning for broadband wireless power amplifier’, IEEE Microw. Wireless Compon. Lett., vol. 31, no. 7, pp. 917-920, July 2021, doi: 10.1109/LMWC.2021.3078459.
4. Z. Liu, X. Hu, L. Xu, W. Wang and F. M. Ghannouchi, ‘Low Computational Complexity Digital Predistortion Based on Convolutional Neural Network for Wideband Power Amplifiers’, IEEE Trans. Circuits Syst. II, Exp. Briefs,vol. 69, no. 3, pp. 1702-1706, March 2022, doi: 10.1109/TCSII.2021.3109973.
5. H. Li, Y. Zhang, G. Li and F. Liu, ‘Vector Decomposed Long Short-Term Memory Model for Behavioral Modeling and Digital Predistortion for Wideband RF Power Amplifiers’, IEEE Access, vol. 8, pp. 63780-63789, 2020, doi: 10.1109/ACCESS.2020. 2984682.
6. J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, ‘Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems’, IEEE Trans. Veh. Technol., vol. 68, no. 11, pp. 10348-10356, Nov. 2019, doi: 10.1109/TVT.2019. 2925562..
7. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, ‘‘Attention is all you need,’’ in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 5998–6008.
8. D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, “A generalized memory polynomial model for digital predistortion of RF power amplifiers,” IEEE Trans. Signal Process., vol. 54, no. 10, pp. 3852-3860, Oct. 2006, doi: 10.1109/TSP.2006.879264.